Gresham's Law of Model Averaging

IN-KOO CHO AND KENNETH KASA

On-line Appendix

Proof of Proposition II.1

PRELIMINARIES

Here we collect some results on time-scales, and use them to establish a 3-tiered time-scale hierarchy among $(\beta_t(0), \beta_t(1), \pi_t)$.

A1. Dynamics of $\beta_t(0)$ and $\beta_t(1)$

To simplify notation, write the Kalman gains as

$$\lambda_t(i) = \frac{\Sigma_t(i)}{\sigma^2 + \Sigma_t(i) z_t^2}$$

for $i = 1, 2$. Define $\forall \tau > 0$

$$m_i(\tau) = \inf\{K \mid \sum_{k=1}^{K} \lambda_k(i) > \tau\}$$

as the first time that $\sum_{k=1}^{K} \lambda_k(i)$ exceeds τ. Since $\lambda_k(i) > 0$ and $\sum_{k=1}^{K} \lambda_k(i) \to \infty$ with probability 1, $m_i(\tau)$ is well defined with probability 1. Similarly, define

$$\tau_K(i) = \sum_{k=1}^{K} \lambda_k(i)$$

as the magnitude of the sum $\sum_{k=1}^{K} \lambda_k(i)$ after K iterations. Note that $m(\tau_K(i)+\tau) - K$ is the number of iterations necessary for $\sum \lambda_k(i)$ to move from τ_K to $\tau_K(i) + \tau$. Therefore $m(\tau_K(i) + \tau) - K$ is an inverse measure of the speed of evolution of the associated recursive formula: if the evolution speed is slow, then it takes many periods to move from τ_K to $\tau_K + \tau$. We are particularly interested in the speed of evolution when K is large.

To compare the speed of evolution, we calculate

$$\lim_{K \to \infty} \frac{m(\tau_K(1) + \tau) - K}{m(\tau_K(0) + \tau) - K}.$$

If this ratio converges to 0, we say that $\beta_t(0)$ evolves on a slower time-scale than $\beta_t(1)$.

Given $\sigma_v > 0$, note that

$$\lim_{K \to \infty} m(\tau_K(1) + \tau) - K$$

remains finite with probability 1. On the other hand,

$$\lim_{K\to\infty} m(\tau_K(0)+\tau) - K = \infty.$$

Thus, $\beta_t(0)$ evolves on a slower time-scale than $\beta_t(1)$. As a result, the right way to take limits is

$$\lim_{\sigma_v\to 0}\lim_{t\to\infty}$$

because in order to move τ distance for a large K, $\beta_t(0)$ needs infinitely many more observations than $\beta_t(1)$. One can therefore regard our exercise as calculating the long run dynamics of $(\pi_t, \beta_t(0), \beta_t(1))$ for an arbitrarily small $\sigma_v > 0$.

In order to move from $\tau_K(1)$ to $\tau_K(1)+\tau$, $\lambda_k(1)$ needs only a finite number of observations, $K_1(\tau)$. However,

$$\lim_{K\to\infty}\sum_{k=K}^{K+K_1(\tau)} \lambda_k(0) = 0.$$

As a result, $\forall\tau > 0$,

$$\lim_{K\to\infty} \beta_{K+K_1(\tau)}(0) - \beta_K(0) = 0$$

with probability 1. Therefore, when investigating the asymptotic dynamics of $\beta_t(1)$, we can treat $\beta_t(0)$ as fixed. By the same token, when investigating the asymptotic properties of $\beta_t(0)$, we can assume that $\beta_t(1)$ has converged to its own stationary distribution (which depends on $\beta_t(0)$).

A2. Dynamics of π_t

To study the dynamics of π_t it is useful to rewrite

$$\text{(A1)} \qquad \frac{1}{\pi_{t+1}} - 1 = \frac{A_{t+1}(0)}{A_{t+1}(1)}\left(\frac{1}{\pi_t} - 1\right)$$

as follows

$$\text{(A2)} \qquad \pi_{t+1} = \pi_t + \pi_t(1-\pi_t)\left[\frac{A_{t+1}(1)/A_{t+1}(0) - 1}{1 + \pi_t(A_{t+1}(1)/A_{t+1}(0) - 1)}\right]$$

which has the familiar form of a discrete-time replicator equation, with a stochastic, state-dependent, fitness function determined by the likelihood ratio. Equation (A2) reveals a lot about the model averaging dynamics. First, it is clear that the boundary points $\pi = \{0, 1\}$ are trivially stable fixed points, since they are absorbing. Second, we can also see that there could be an interior fixed point, where $\mathsf{E}(A_{t+1}(1)/A_{t+1}(0)) = 1$. However, we shall also see there that this fixed point is unstable. So we know already that π_t will spend most of its time near the boundary points.

PROPOSITION A.1: *As long as the likelihoods of $\mathcal{M}_0$ and $\mathcal{M}_1$ have full support, the boundary points $\pi_t = \{0, 1\}$ are unattainable in finite time.*

PROOF:

With two full support probability distributions, you can never conclude that a history of any finite length couldn't have come from either of the distributions. Slightly more formally, if the distributions have full support, they are mutually absolutely continuous, so the likelihood ratio in eq. (A2) is strictly bounded between 0 and some upper bound B. To see why $\pi_t < 1$ for all t, notice that $\pi_{t+1} < \pi_t + \pi_t(1-\pi_t)M$ for some $M < 1$, since the likelihood ratio is bounded by B. Therefore, since $\pi + \pi(1-\pi) \in [0, 1]$ for $\pi \in [0, 1]$, we have

$$\pi_{t+1} \leq \pi_t + \pi_t(1-\pi_t)M < \pi_t + \pi_t(1-\pi_t) \leq 1$$

and so the result follows by induction. The argument for why $\pi_t > 0$ is completely symmetric.

Since the distributions here are Gaussian, they obviously have full support, so Proposition A.1 applies. Although the boundary points are unattainable in finite time, the replicator equation for π_t in (A2) makes it clear that π_t will spend most of its time near these boundary points, since the relationship between π_t and π_{t+1} has the familiar logit function shape, which flattens out near the boundaries. As a result, near its stable limits π_t evolves very slowly. In fact, we shall show that it evolves even more slowly than the t^{-1} time-scale of $\beta_t(0)$. This means that when studying the dynamics of the coefficient estimates near the boundaries, we can treat π_t as fixed.

Although π_t can evolve faster than $\beta_t(1)$ for small t, as $t \to \infty$, that π_t must stay in a small neighborhood of 1 or 0, slowly converging to the limit.

LEMMA A.2: *Let Π be the collection of all sample paths of $\{\pi_t\}$, and define the subset*

$$\Pi_0 = \{\{\pi_t\} \mid \textit{there is no subsequence converging to } 0 \textit{ or } 1\ \}$$

We then have

$$\mathsf{P}\left(\exists\{\pi_{t_k}\}_k,\ \textit{and}\ \exists \pi^* \in (0,1),\ \lim_{k\to\infty} \pi_{t_k} = \pi^*\right) = 0$$

and π_t evolves at a slower time scale than $\beta_t(0)$.

PROOF:

Fix a sequence $\{\pi_t\}$ in Π_0. Since the sequence is a subset of a compact set, it has a convergent subsequence. After renumbering the subsequence, let us assume that

$$\lim_{t\to\infty} \pi_t = \pi^* \in (0, 1)$$

since $\{\pi_t\} \in \Pi_0$. Depending upon the rate of convergence (or the time scale according to which π_t converges to π^*), we treat π_t as already having converged to π^*.[1]

[1] If π_t evolves at a slower time scale than $\beta_t(0)$, then we fix π_t while investigating the asymptotic properties of $\beta_t(0)$. As it turns out, we obtain the same conclusion for all cases.

We only prove the case in which $\pi_t \to \pi^*$ according to the fastest time scale, in particular, faster than the time scale of $\beta_t(1)$. Proofs for the remaining cases follow the same logic.

Since π_t evolves on the fastest time-scale, assume that

$$\pi_t = \pi^*.$$

Since $\beta_t(1)$ evolves on a faster time scale than $\beta_t(0)$, we first let $\beta_t(1)$ reach its own limit, and then let $\beta_t(0)$ go to its limit.

Fix $\sigma_v > 0$. Let $p_t^e(i)$ be $\mathcal{M}_i$'s period-t forecast of p_{t+1},

$$p_t^e(i) = \beta_t(i) z_t.$$

Since

$$p_t = \alpha\rho[(1-\pi_t)\beta_t(0) + \pi_t\beta_t(1)]z_t + \delta z_t + \sigma\epsilon_t,$$

model 1's forecast error is

$$p_t - p_t^e(1) = [\alpha\rho(1-\pi_t)\beta_t(0) + (\alpha\rho\pi_t - 1)\beta_t(1) + \delta]\, z_t + \sigma\epsilon_t.$$

We know,

$$\lim_{t\to\infty} \mathsf{E}\,[\alpha\rho(1-\pi_t)\beta_t(0) + (\alpha\rho\pi_t - 1)\beta_t(1) + \delta] = 0$$

in any limit point of the Bayesian learning dynamics.[2] Define

$$\overline{\beta}(1) = \lim_{t\to 0} \mathsf{E}\beta_t(1)$$

Note that its value is conditioned on π_t and $\beta_t(0)$. Since

$$\lim_{t\to 0} \mathsf{E}\left[\alpha\rho(1-\pi_t)\beta_t(0) + (\alpha\rho\pi_t - 1)\overline{\beta}(1) + \delta\right] + \mathsf{E}(\alpha\rho\pi_t - 1)(\beta_t(1) - \overline{\beta}(1)) = 0.$$

we have

$$\overline{\beta}(1) = \frac{\alpha\rho(1-\pi_t)\beta_t(0) + \delta}{1 - \alpha\rho\pi_t} \tag{A3}$$

for fixed π_t, $\beta_t(0)$. Define the deviation from the long-run mean as

$$\xi_t = \beta_t(1) - \overline{\beta}(1).$$

Model 1's mean-squared forecast error is then

$$\lim_{t\to 0} \mathsf{E}(p_t - p_t^e(1))^2 = \lim_{t\to 0} \mathsf{E} z_t^2 (\alpha\rho\pi_t - 1)^2 \sigma_\xi^2 + \sigma^2$$

[2]Existence is implied by the tightness of the underlying space.

Note that $\sigma_\xi^2 > 0$ if $\sigma_v > 0$, and

$$\lim_{\sigma_v^2 \to 0} \sigma_\xi^2 = 0.$$

To investigate the asymptotic properties of $\beta_t(0)$, let us write

$$\beta_t(1) = \frac{\alpha\rho(1-\pi_t)\beta_t(0) + \delta}{1 - \alpha\rho\pi_t} + \xi_t$$

We can then write Model 0's forecast error as

$$p_t - p_t^e(0) = z_t \left[-\frac{1-\alpha\rho}{1-\alpha\rho\pi_t} \left(\beta_t(0) - \frac{\delta}{1-\alpha\rho} \right) + \alpha\rho\pi_t\xi_t \right] + \sigma\epsilon_t.$$

Since $\beta_t(0)$ evolves according to

(A4) $$\beta_{t+1}(0) = \beta_t(0) + \left(\frac{\Sigma_t(0)}{\sigma^2 + \Sigma_t(0)z_t^2} \right) z_t(p_t - \beta_t(0)z_t),$$

$$\lim_{t\to\infty} \beta_t(0) = \frac{\delta}{1-\alpha\rho}$$

with probability 1. Thus, the mean-squared forecast error satisfies

$$\lim_{t\to\infty} \mathsf{E}(p_t - p_t^e(0))^2 = \lim_{t\to\infty} \mathsf{E} z_t^2 \sigma_\xi^2 (\alpha\rho\pi_t)^2 + \sigma^2$$

After substituting $\beta_t(0)$ into (A3), we have

$$\lim_{\sigma_v \to 0} \lim_{t \to 0} \beta_t(1) = \frac{\delta}{1-\alpha\rho}$$

weakly. Note that

(A5) $$\lim_{t\to\infty} \frac{\mathsf{E}(p_t - p_t^e(0))^2}{\mathsf{E}(p_t - p_t^e(1))^2} > 1$$

if and only if

$$\lim_{t\to\infty} \left(\frac{\alpha\rho\pi_t}{1-\alpha\rho\pi_t} \right)^2 > 1.$$

Finally, notice that

$$\frac{\alpha\rho\pi_t}{1-\alpha\rho\pi_t} < 1$$

if and only if

$$\alpha\rho\pi_t < \frac{1}{2}.$$

Note that the left hand side is an increasing function of π_t. Hence, if (A5) holds for some $t \geq 1$, then it holds again for $t+1$. Similarly, if (A5) fails for some $t \geq 1$, then the same condition continues to fail for $t+1$.

Thus, π_t continues to increase or decrease, if the inequality holds in either direction. Recall that $\pi^* = \lim_{t\to\infty} \pi_t$. Convergence to π^* can occur only if (A5) holds with equality for <u>all</u> $t \geq 1$, which is a zero probability event. We conclude that $\pi^* \in (0,1)$ occurs with probability 0.

A3. Log odds ratio

As usual, it is more convenient to consider the log odds ratio. Let us initialize the likelihood ratio at the prior odds ratio:

$$\frac{A_0(0)}{A_0(1)} = \frac{\pi_0(0)}{\pi_0(1)}.$$

By iteration we get

$$\frac{\pi_{t+1}(0)}{\pi_{t+1}(1)} = \frac{1}{\pi_{t+1}} - 1 = \prod_{k=0}^{t+1} \frac{A_k(0)}{A_k(1)},$$

Taking logs and dividing by $(t+1)$,

$$\frac{1}{t+1} \ln\left(\frac{1}{\pi_{t+1}} - 1\right) = \frac{1}{t+1} \sum_{k=0}^{t+1} \ln \frac{A_k(0)}{A_k(1)}.$$

Now define the average log odds ratio, ϕ_t, as follows

$$\phi_t = \frac{1}{t} \ln\left(\frac{1}{\pi_t} - 1\right) = \frac{1}{t} \ln\left(\frac{\pi_t(0)}{\pi_t(1)}\right)$$

which can be written recursively as the following stochastic approximation algorithm

$$\phi_t = \phi_{t-1} + \frac{1}{t}\left[\ln \frac{A_t(0)}{A_t(1)} - \phi_{t-1}\right].$$

Invoking well knowing results from stochastic approximation, we know that the asymptotic properties of ϕ_t are determined by the stability properties of the following ordinary differential equation (ODE)

$$\dot{\phi} = \mathsf{E}\left[\ln \frac{A_t(0)}{A_t(1)}\right] - \phi$$

which has a unique stable point

$$\phi^* = \mathsf{E} \ln \frac{A_t(0)}{A_t(1)}.$$

Note that if $\phi^* > 0$, $\pi_t \to 0$, while if $\phi^* < 0$, $\pi_t \to 1$. The focus of the ensuing analysis is to identify the conditions under which π_t converges to 1, or 0. Thus, the sign of ϕ^*, rather than its value, becomes the key object of investigation.

A4. Time scale of π_t

Given any $\alpha \geq 1$, a simple calculation shows

$$t^\alpha(\pi_t - \pi_{t-1}) = \frac{t^\alpha(e^{(t-1)\phi_{t-1}} - e^{t\phi_t})}{(1+e^{t\phi_t})(1+e^{(t-1)\phi_{t-1}})}.$$

As $t \to \infty$, we know $\phi_t \to \phi^*$ with probability 1. Hence, we have

$$\begin{aligned}\lim_{t\to\infty} t^\alpha(\pi_t - \pi_{t-1}) &= \lim_{t\to\infty} \frac{t^\alpha\left(e^{-\phi^*}-1\right)e^{t\phi^*}}{(1+e^{t\phi^*})(1+e^{(t-1)\phi^*})} \\ &= (e^{-\phi^*}-1)\lim_{t\to\infty}\frac{t^\alpha}{(1+e^{-t\phi^*})(1+e^{t\phi^*}e^{-\phi^*})}\end{aligned}$$

Finally, notice that for both $\phi^* > 0$ and $\phi^* < 0$ the denominator converges to ∞ faster than the numerator for any $\alpha \geq 1$. Note that $\pi_t \propto \frac{1}{t}$ if

$$0 < \liminf_{t\to\infty} |t^2(\pi_t - \pi_{t-1})| \leq \limsup_{t\to\infty} |t^2(\pi_t - \pi_{t-1})| < \infty.$$

In our case, the first strict inequality is violated, which implies that π_t evolves at a rate slower than $1/t$.

A5. Summary

It is helpful to summarize our findings on the time-scales of our three stochastic processes: π_t, $\beta_t(0)$ and $\beta_t(1)$. As indicated by (A2), π_t evolves quickly in the interior of $[0, 1]$. However, no sample path converges to $\pi^* \in (0, 1)$ with positive probability. Once π_t enters a small neighborhood of $\{0, 1\}$, the evolution of π_t slows down significantly. In the neighborhood of $\{0, 1\}$, we have a hierarchy of time-scales among three stochastic processes. $\beta_t(1)$ evolves according to a faster time scale than $\beta_t(0)$, which evolves at a faster time scale than π_t.

PROOF OF PROPOSITION II.1

Although the proof follows the same logic as the proof of Lemma A.2, we sketch it here as a reference. Along the way, we illustrate the domain of attraction of each locally stable point, and provide a description of a typical convergent path.

We use standard convergence results from the stochastic approximation literature (Kushner and Yin (1997)), and their large deviation properties (Dupuis and Kushner (1987)). The analysis requires that all stochastic processes are contained in compact convex sets. Since we assume Gaussian shocks, $\beta_t(i)$ has full support in $\mathbb{R}$. Following Kushner and

Yin (1997), we ensure compactness by imposing a projection facility, i.e., $\exists B > \frac{\delta}{1-\alpha\rho}$ such that $\beta_t(i) \in [-B, B]$ for $i = 0, 1$. When $\beta_t(i) \notin [-B, B]$, $\beta_t(i)$ is projected back into $[-B, B]$. Kushner and Yin (1997) show that the asymptotic properties of $(\pi_t, \beta_t(0), \beta_t(1))$ are not affected by the projection facility, as long as $[-B, B]$ contains the stable point of $\beta_t(i)$. Because the Gaussian distribution has thin tails, Dupuis and Kushner (1987) are able to show that the large deviation properties of $(\pi_t, \beta_t(0), \beta_t(1))$ are not affected by the projection facility. For the rest of the proof, we presume that $(\pi_t, \beta_t(0), \beta_t(1)) \in [0, 1] \times [-B, B] \times [-B, B]$. To simplify notation, however, we suppress the projection facility.

We first investigate the properties of $(\pi_t, \beta_t(0), \beta_t(1))$ as $t \to \infty$ for a fixed $\sigma_v > 0$. Since $\beta_t(1)$ evolves at the fastest time scale, we first investigate the asymptotic properties of $\beta_t(1)$ for fixed $(\pi, \beta(0))$. As we have shown in the proof of Lemma A.2, $\beta_t(1)$ has a stationary distribution, and its mean converges to

$$\overline{\beta}(1) = \frac{\alpha\rho(1-\pi_t)\beta_t(0) + \delta}{1 - \alpha\rho\pi_t}.$$

For later reference, define

$$\mathcal{S} = \left\{ (\pi, \beta(0), \beta(1)) \mid \beta(1) = \frac{\alpha\rho(1-\pi)\beta(0) + \delta}{1 - \alpha\rho\pi_t} \right\} \tag{B1}$$

which is a submanifold in $\mathbb{R}^3$.

Given the stationary distribution of $\beta_t(1)$, we investigate the asymptotic properties of $\beta_t(0)$, for a fixed value of π_t (in a small neighborhood of $\{0, 1\}$). Again, we have shown that

$$\lim_{t\to\infty} \beta_t(0) = \frac{\delta}{1 - \alpha\rho}$$

with probability 1, which implies that

$$\overline{\beta}(1) \to \frac{\delta}{1 - \alpha\rho}$$

$\forall \pi_t$. Then, observe that $\pi_t \to 1$ if and only if $\phi^* < 0$, and $\pi_t \to 0$ if and only if $\phi^* > 0$, where

$$\phi^* = \mathsf{E} \ln \frac{A_t(0)}{A_t(1)}$$

where the expectation is taken with respect to the stationary distribution. It is convenient to consider the deterministic dynamics on the time-scale of $\beta_t(0)$. The domain of attraction for $(\pi, \beta(0), \beta(1)) = (0, \delta/(1-\alpha\rho), \delta/(1-\alpha\rho))$ is

$$\mathcal{D}_0 = \left\{ (\pi, \beta(0), \beta(1)) \mid \mathsf{E} \log \frac{A_t(0)}{A_t(1)} > 0 \right\}$$

where $A_t(0)$ and $A_t(1)$ are the agent's perceived likelihood functions:

$$\log A_t(1) = -\frac{[(\alpha\rho\pi_t - 1)z_t\xi_t + \sigma\epsilon_t]^2}{2(\Sigma_t(1)z_t^2 + \sigma^2)} - \frac{1}{2}\log 2\left[\Sigma_t(1)z_t^2 + \sigma^2\right]$$

and

$$\log A_t(0) = -\frac{\left[-z_t\left[\frac{1-\alpha\rho}{1-\alpha\rho\pi_t}\left(\beta_t(0) - \frac{\delta}{1-\alpha\rho}\right)\right] + \alpha\rho\pi_t z_t\xi_t + \sigma\epsilon_t\right]^2}{2(\Sigma_t(0)z_t^2 + \sigma^2)} - \frac{1}{2}\log 2\left[\Sigma_t(0)z_t^2 + \sigma^2\right].$$

It is helpful to characterize $(\pi_t, \beta_t(0))$ along the boundary of $\mathcal{D}_0$, where $\phi^* = 0$. To simplify exposition, we treat z_t as deterministic, but the same analysis applies to the general case with minor modifications.

Since we are interested in the sign of ϕ^*, which is computed with respect to the asymptotic probability distribution as $t \to \infty$, we replace $\Sigma_t(1)$ by $\overline{\Sigma}$, and $\Sigma_t(0)$ by 0. After a tedious calculation (even with our simplifying assumption that z_t is deterministic), we find that

(B2)

$$\phi^* = -\frac{\alpha^2\rho^2\pi_t^2(\overline{\Sigma}z_t^2)^2 + 2\sigma^2\alpha\rho\pi_t\overline{\Sigma}z_t^2}{2\sigma^2(\sigma^2 + \overline{\Sigma}z_t^2)} - \frac{z_t^2}{2\sigma^2}\left(\frac{1-\alpha\rho}{1-\alpha\rho\pi_t}\right)\left(\beta_t(0) - \frac{\delta}{1-\alpha\rho}\right)^2 + \frac{1}{2}\log\left(1 + \frac{\overline{\Sigma}z_t^2}{\sigma^2}\right).$$

Note that the right-hand side is a decreasing function of π_t. This reflects the feedback in the system. A model's relative performance decreases when the weight on it decreases. Naturally, the right-hand side also decreases with $\left(\beta_t(0) - \frac{\delta}{1-\alpha\rho}\right)^2$, i.e., in response to deviations from the self-confirmin equilibrium. Thus, the contour of $(\pi_t, \beta_t(0))$ satisfying $\phi^* = 0$ is symmetric around $\beta(0) = \delta/(1-\alpha\rho)$, and as $\left(\beta_t(0) - \frac{\delta}{1-\alpha\rho}\right)^2$ decreases π_t must increase, in order to satisfy $\phi^* = 0$. In particular, if $\pi_t = 0$, then

$$d(\sigma_v) = \left|\beta_t(0) - \frac{\delta}{1-\alpha\rho}\right| = \frac{\sigma}{|z_t|\sqrt{1-\alpha\rho}}\sqrt{\log\left(1 + \frac{\overline{\Sigma}z_t^2}{\sigma^2}\right)} \tag{B3}$$

which is a strictly increasing function of $\overline{\Sigma}$, and therefore, a strictly increasing function of σ_v. In particular,

$$\lim_{\sigma_v \to 0} d(\sigma_v) = 0.$$

Among $(\pi_t, \beta_t(0))$ satisfying $\phi^* = 0$, π_t is maximized if $\beta_t(0) = \delta/(1-\alpha\rho)$. This π_t is the positive root of

$$\alpha^2\rho^2\overline{\Sigma}^2 z_t^2\pi_t^2 + 2\sigma^2\alpha\rho\pi_t - (\sigma^2 + \overline{\Sigma}z_t^2)\log\left(1 + \frac{\overline{\Sigma}z_t^2}{\sigma^2}\right) = 0$$

A simple calculation shows that if π_t is the positive root of the quadratic equation,

$$\lim_{\sigma_v \to 0} \pi_t = \frac{1}{2\alpha\rho}.$$

Thus, $\forall \epsilon > 0, \exists \sigma_v' > 0$ so that $\forall \sigma_v \in (0, \sigma_v')$,

$$\mathcal{D}_0 \subset \left\{ (\pi, \beta(0), \beta(1)) \mid \pi \leq \frac{1}{2\alpha\rho} + \epsilon \right\}.$$

Note that $\mathcal{D}_0$ looks like a pipe in $\mathbb{R}^3$, since it is independent of $\beta(1)$. As $\sigma_v \to 0$, the base of $\mathcal{D}_0$ on the surface spanned by $\beta(1)$ and $\beta(0)$ shrinks, making $\mathcal{D}_0$ "thinner."

It is instructive to visualize a typical sample path of $(\pi_t, \beta_t(0), \beta_t(1))$ to a locally stable point. Suppose that $\pi_1 \in (0, 1)$, and $(\pi_1, \beta_1(0), \beta_1(1))$ is outside of $\mathcal{D}_0$. Then, for a small value of t, π_t evolves rapidly toward a neighborhood of 1 or 0, with a speed of evolution that may be comparable to the speed of evolution of $\beta_t(1)$, while π_t remains away from the boundary points. Since $\beta_t(1)$ evolves on a faster time-scale than $\beta_t(0)$, $(\pi_t, \beta_t(0), \beta_t(1))$ evolves as if $\beta_t(0) = \beta_1(0)$, while π_t stays away from the boundary points. From the perspective of $\beta_t(0)$, $\beta_t(1)$ instantaneously moves to a neighborhood of submanifold $\mathcal{S}$. This is why $\mathcal{D}_0$ is independent of $\beta_t(1)$.

$(\pi_t, \beta_t(0), \beta_t(1))$ hits the neighborhood of submanifold $\mathcal{S}$ defined by (B1), as the distribution of $\beta_t(1)$ converges to its stationary distribution, while π_t converges to a neighborhood of either 0 or 1. Then, along the surface of $\mathcal{S}$, $(\pi_t, \beta_t(0), \beta_t(1))$ moves as $\beta_t(0)$ evolves, converging to $\frac{\delta}{1-\alpha\rho}$. After $\beta_t(0)$ reaches $\frac{\delta}{1-\alpha\rho}$ along the surface of $\mathcal{S}$ so that $\beta_t(1)$ also reaches $\frac{\delta}{1-\alpha\rho}$, π_t moves. If $(\pi_t, \beta_t(0), \beta_t(1)) \in \mathcal{S} \cap \mathcal{D}_0$, then it will converge to the limit point where $\pi_t = 0$. Otherwise, it converges to another limit point where $\pi_t = 1$.

*

REFERENCES

Dupuis, Paul, and Harold J. Kushner. 1987. "Asymptotic Behavior of Constrained Stochastic Approximations vis the Theory of Large Deviations." *Probability Theory and Related Fields*, 75: 223–44.

Kushner, Harold J., and G. George Yin. 1997. *Stochastic Approximation Algorithms and Applications.* Springer-Verlag.