Chapter 3

GENERALIZED URN PROCESSES AND PROCESSES WITH NONATTRACTING POINTS

ABSTRACT

We generalize the Pólya urn process in two ways. First a simple time dependence is considered. For this process, we characterize the distribution of the random limiting fraction of red balls, giving conditions for its concentration on $\{0, 1\}$ and for its being non-atomic. Second, a multicolor process with nonlinear draws is considered. For this process, we give analogues of theorem 2.4, showing that the vector of fractions of balls of each color cannot converge to certain unstable equilibria (theorem 3.8). The method we use is to look at a certain functional of the urn process. This is then a sequence of real valued random variables, and we identify certain points as nonattracting points for the sequence. We prove a general lemma, also used in chapter five, stating that a sequence of real random variables cannot converge to a nonattracting point.

3.1 Simple time dependence

We begin by considering a Pólya urn with the single change that the number of extra balls added of the color drawn is a function of time. For simplicity we consider the two-color case, but the generalization to more than two colors is automatic by grouping together colors. Let $F : \mathbf{Z}^+ \to (0, \infty)$ be any function. Let $\{v_1, v_2, \ldots\}$ be the successive proportion of red balls in an urn that begins with one red ball and one black ball and evolves as follows: at discrete times $n = 1, 2, \ldots$ a ball is drawn and replaced in the urn along with $F(n)$ balls of the same color. Note that the proportions of red balls in the urn form a martingale, so it may be possible for the proportions of red balls to converge anywhere in $[0, 1]$ (see the discussion at the beginning of section 2.2). The usual Pólya urn scheme is the case where F is a constant. We show that v_n must converge for any F and determine those F for which the limit must be 0 or 1. The reasoning uses only elementary facts about bounded martingales. We give conditions implying that the law of the random limit is non-atomic except at 0 and 1 and conjecture that this is always true.

The following formal definition of the process is completely routine; it is included because it will be useful to refer to the underlying measure space in the proof of lemma 3.4. Let $\mathbf{\Omega}$ be $[0, 1]^{\mathbf{Z}^+}$ with the product uniform measure. All probabilities will be with respect to this space and all functions will be functions of ω where ω is a generic point in $\mathbf{\Omega}$, but the notation will supress the role of ω when no ambiguity arises. Let z_n be the n^{th} coordinate of ω so that $\{z_n : n = 0, 1, 2, \ldots\}$ is a set of independent uniformly distributed variables on $[0, 1]$, and let $\mathcal{F}_n$ be the σ-algebra generated by $\{z_i : i \leq n\}$. Let $S_1(0) = S_2(0) = 1$ and recursively define

$$\begin{aligned} S_1(n+1) &= S_1(n) + F(n)\mathbf{1}(z_n \leq S_1(n)/(S_1(n) + S_2(n))) \\ S_2(n+1) &= S_1(n) + F(n)\mathbf{1}(z_n > S_1(n)/(S_1(n) + S_2(n))) \end{aligned} \tag{1}$$

where $\mathbf{1}$ denotes the indicator function of a set. So $S_1(n)$ and $S_2(n)$ represent the numbers of red and black balls in the urn after n draws. For convenience we let

$$\delta_n = F(n)/(2 + \sum_{i=0}^{n-1} F(i))$$

denote the fractional additions. Let $v_n = S_1(n)/(S_1(n) + S_2(n))$ denote the proportion of red balls at time n. We will prove the following results.

Theorem 3.1 *For any function F as above, the random variables v_n converge almost surely to some random variable v.*

Theorem 3.2 *The limit v satisfies $prob(v = 0) = prob(v = 1/2) = 1$ if and only if $\sum_{i=1}^{\infty} {\delta_n}^2 = \infty$.*

This theorem applies, for example, when $F(n) = 2^n$. Roughly, the hypothesis means that F grows faster than polynomially, but one needs to look more closely if the growth is irregular since the function

$$F(n) = \begin{cases} n \text{ if } n \text{ is a power of } 2 \\ 2^{-n} \text{ otherwise} \end{cases}$$

satisfies the hypothesis.

Theorem 3.3 *The distribution of v has no atoms on* $(0, 1)$.

Remark: It is possible for the distribution of v to have atoms at 0 and 1 of weight less than 1/2 each; then the remainder of the time v is in $(0, 1)$ and this part of the distribution is nonatomic. An example where this occurs is if $F(n) = n$. In this case the probability that all draws are of the same color is $\frac{2}{3} \times \frac{6}{7} \times \frac{15}{16} \times \cdots > 0$, but according to theorem 3.2 the distribution is not entirely concentrated on $\{0, 1\}$. I do not know of a converse to theorem 3.2 giving a necessary condition for the probability that $v = 0$ or 1 to be nonzero.

Proof of theorem 3.1: $\{v_n : n = 1, 2, \ldots\}$ is a martingale. To see this, calculate

$$\begin{aligned}
\mathbf{E}(v_{n+1} \,|\, \mathcal{F}_n) &= v_n(S_1(n) + F(i))/(S_1(n) + F(n) + S_2(n)) \\
&+ (1 - v_n)S_1(n)/(S_1(n) + S_2(n) + F(n)) \\
&= S_1(n) + v_n F(n)/(S_1(n) + S_2(n) + F(n)) \\
&= v_n.
\end{aligned}$$

Now since $\{v_n\}$ is bounded, it converges almost surely to some v. □

Proof of theorem 3.2: We calculate the expected value of v^2. By symmetry this is most 1/2 with equality if and only if $prob(v = 0) = prob(v = 1) = 1/2$. Necessary and sufficient conditions for this will follow from the simple recurrence relation (2) on the values of $1/2 - \mathbf{E}({v_n}^2)$ which is denoted W_n.

Since v_n converges almost surely to v and the variables are bounded by 1, we know that $\mathbf{E}(v_n^2)$ converges to $\mathbf{E}(v^2)$. Let V_n denote $\mathbf{E}({v_n}^2)$. For a fixed F, v_n takes on only finitely many values and V_n can be recursively calculated as follows. If $v_{n-1}(\omega) = x = S_1(n-1)/(S_1(n-1) + S_2(n-1))$ then

$$\begin{aligned}
v_n(\omega) &= S_1(n-1)/(S_1(n-1) + S_2(n-1) + F(n)) \\
&= x - x\delta_n/(1 + \delta_n) \\
&= x/(1 + \delta_n)
\end{aligned}$$

with probability $1 - x$, and

$$\begin{aligned}
v_n(\omega) &= (S_1(n-1) + F(n))/(S_1(n-1) + S_2(n-1) + F(n)) \\
&= x + ((1 - x)\delta_n)/(1 + \delta_n) \\
&= (x + \delta_n)/(1 + \delta_n)
\end{aligned}$$

with probability x. So

$$\begin{aligned} V_n &= \int x^2 dv_n \\ &= \int (1-x)x^2/(1+\delta_n)^2 + x(x+\delta_n)^2/(1+\delta_n)^2 dv_{n-1} \\ &= 1/(1+\delta_n)^2 \int (1-x)x^2 + x(x+\delta_n)^2 dv_{n-1} \\ &= 1/(1+\delta_n)^2 \int x^2 + 2\delta_n x^2 + x\delta_n^2 dv_{n-1} \\ &= (\delta_n^2/2 + S_{n-1}(1+2\delta_n))/(1+\delta_n)^2 \end{aligned}$$

To see how the value of V_n relates to the value of V_{n-1} we let W_k denote $1/2 - V_k$. Then

$$\begin{aligned} W_n &= 1/2 - [(\delta_n^2/2) + S_{n-1}(1+2\delta_n)]/(1+\delta_n)^2 \\ &= (1/2 + \delta_n - S_{n-1} - 2\delta_n S_{n-1})/(1+\delta_n)^2 \\ &= W_{n-1}(1+2\delta_n)/(1+\delta_n)^2 \\ &= W_{n-1}(1 - \delta_n^2/(1+\delta_n)^2). \end{aligned} \tag{2}$$

Thus the value V_n converges to 1/2 if and only if W_n converges to 0 which happens whenever the product of the values $(1 - \delta_n^2/(1+\delta_n)^2)$ converges to 0. This happens whenever $\sum_{n=1}^{\infty} \delta_n^2/(1+\delta_n)^2$ diverges, which in turn happens exactly when $\sum_{n=1}^{\infty} \delta_n^2$ diverges, and theorem 3.2 is proved. □

The proof of theorem 3.3 will be given at the end of the next section but in the case that $\sum_{n=1}^{\infty} F(n) = M < \infty$ there is a combinatorial proof which is enough simpler to be worth including. To prove theorem 3.3 in this case, pick any $x \in [0,1]$ and set $L = Mx = x\sum_{n=1}^{\infty} F(n)$. Then $v_n(\omega)$ converges to x if and only if $S_1(n)(\omega)$ converges to L. Suppose $prob(S_1(n) \to L) > 0$. Since Ω is a product measure space there must be some finite set of values $\xi_1, \ldots \xi_{j-1}$, such that $prob(S_1(n) \to L \mid S_1(k+1) = \xi_1, \ldots, S_1(k+j-1)) = \xi_{j-1})$ is arbitrarily close to 1, say greater than $(1+M)/(2+M)$. The following argument shows that this cannot happen.

Fix any value of $S_1(j-1)$ and consider all probabilities hereafter to be conditional upon this value of $S_1(j-1)$. Now find a k such that $\sum_{n=k+1}^{\infty} F(n) < F(j)$. If $S_1(n, \omega)$

converges to L then $S_1(k,\omega)$ must be in the interval $[L - \sum_{n=k+1}^{\infty} F(n)\,,\, L]$. This interval has length less than $F(j)$. Consequently, if a given sequence of draws from the j^{th} through the k^{th} causes $S_1(k)$ to fall in this interval, then the same sequence with the color of the j^{th} draw reversed will miss the interval. Formally, if

$$A(\omega) = \{i : j \leq i \leq k \text{ and } z_{i-1} \leq S_1(i-1)/(S_1(i-1) + S_2(i-1))\}$$

and

$$\mathcal{S} = \{B \subseteq j, ..., k : S_1(k,\omega) \in [L - \sum_{n=k+1}^{\infty} F(n)\,,\, L] \text{ and } A(\omega) = B\}$$

then

$$B \in \mathcal{S} \text{ and } j \in B \Rightarrow B\backslash\{j\} \notin \mathcal{S} \qquad \text{and}$$
$$B \in \mathcal{S} \text{ and } j \notin B \Rightarrow B \cup \{j\} \notin \mathcal{S}.$$

We need to show $prob(A(\omega) \in \mathcal{S}) \leq (1+M)/(2+M)$. So we consider a subset $\mathcal{T}$ of the complement of $\mathcal{S}$. Let $\mathcal{T}$ be the collection of $B \subseteq \{j, \ldots, k\}$ such that either

$$(i) \quad j \notin B \quad \text{and} \quad \text{if } A(\omega) = B \text{ then } S_1(k,\omega) < L - \sum_{n=k+1}^{\infty} F(n)$$

$$\text{or} \quad (ii) \quad j \in B \quad \text{and} \quad \text{if } A(\omega) = B \text{ then } a_k(\omega) > L.$$

Then $\mathcal{T}$ has the following properties:

(i) $\{B \in \mathcal{T} : j \notin B\}$ is closed under subset

(ii) $\{B \in \mathcal{T} : j \in B\}$ is closed under superset

(iii) For any $C \subseteq \{j+1, \ldots, k\}$ either $C \in \mathcal{T}$ or $\{j\} \cup C \in \mathcal{T}$

The final step is to show that $prob(A(\omega) \in \mathcal{T}) \geq 1/(2+M)$ for any $\mathcal{T}$ satisfying properties (i) - (iii). We use the following:

Lemma 3.4 *For any collection $\mathcal{U}$ of subsets of $\{j+1,\ldots,k\}$ that is closed under subset, $prob(A(\omega) \in \mathcal{U} \mid j \notin A(\omega)) \geq prob(A(\omega)\backslash\{j\} \in \mathcal{U} \mid j \in A(\omega))$. (Here all probabilities are still conditional on a_{j-1}.) In other words a set of outcomes for draws $j+1,\ldots,k$ that is closed under changing red draws to black is at least as likely after a black ball on the j^{th} draw as it would be after a red ball on the j^{th} draw.*

Proof: Recall from the formal definition of the process the variables $\{z_i\}$ of equation (1). For a given sequence of variables $z_{j+1},\ldots,z_k$, decreasing the value of $S_1(j)$ can only change red draws to black. The lemma follows from the independence of the variables z_i. □

Apply the lemma with $\mathcal{U} = \{B \in \mathcal{T} : j \notin B\}$. Letting $\alpha = prob(A(\omega) \in \mathcal{U} \mid j \notin A(\omega))$, we have

$$\begin{aligned} & prob(A(\omega) \in \mathcal{T} \mid j \in A(\omega)) \\ \geq\ & 1 - prob(A(\omega)\backslash\{j\} \in \mathcal{U} \mid j \in A(\omega)) \\ \geq\ & 1-\alpha. \end{aligned}$$

Then $prob(A(\omega) \in \mathcal{T}) \geq \alpha\, prob(j \notin A(\omega)) + (1-\alpha)\, prob(j \epsilon A(\omega)) \geq \min\{prob(j \in A(\omega)),\, prob(j \notin A(\omega))\}$ which is at least $1/(2+M)$ as required. □

Knowing that the distribution of v is nonatomic on $(0,1)$, it is logical to ask when the distribution is absolutely continuous with respect to Lebesgue measure. Nothing is known about this except when the distribution of v is known explicitly or when $F(n)$ goes to zero very fast.

3.2 Processes with non-attracting points

The object of this section is to state conditions (3) - (5) on a process $X_1, X_2, \ldots$ that prevent the partial sums $S_n = X_1 + \cdots + X_n$ from converging to a preassigned point,

p. One application will take S_n to be the variables v_n from the previous section. By showing the conditions to hold for all $p \in (0,1)$ we will conclude that the law of $\lim_{n\to\infty} S_n$ is nonatomic on $(0,1)$ and theorem 3.3 will be proved. Note that we cannot here conclude that S_n never converges to a point of $(0,1)$; in fact it often does. Later, in chapter five, we will apply the results of this section in cases where S_n is known to converge almost surely to one of a finite set of points. In this case, we will indeed be able to rule out any point satisfying conditions (3) - (5) as possible points of convergence of S_n.

Consider a sequence of real random variables $X_1, X_2, \ldots$ and their partial sums, $S_n = S_0 + \sum_{i=1}^{n} X_i$ for some fixed S_0. Let p be any real number and let $\mathcal{N}$ be a neighborhood of p. Suppose that the following properties hold for sufficiently large n:

$$|X_n| \leq \frac{c_0}{n^\gamma} \tag{3}$$

$$(S_n - p)(\mathbf{E}(X_{n+1}|\mathcal{F}_n) \geq 0 \tag{4}$$

$$\frac{c_1}{n^2} \geq \mathrm{Var}(X_{n+1}|\mathcal{F}_n) \geq \frac{c_2}{n^2} \tag{5}$$

where c_0, c_1, c_2, ϵ are positive constants, the sequence $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots$ is a filtration with $\mathcal{F}_n$ containing the σ-algebra generated by $\{S_1, \ldots, S_n\}$, γ is a constant greater than 1/2, and $\mathrm{Var}(X_{n+1}|\mathcal{F}_n) = \mathbf{E}({X_{n+1}}^2|\mathcal{F}_n) - \mathbf{E}(X_{n+1}|\mathcal{F}_n)^2$. Condition (4) states that the S_{n+1} is expected to lie farther from from p than S_n, and on the same side. Hill, Lane and Sudderth say p is a splitting point for $\{S_n\}$ when this holds.

Remark: These conditions may be taken to hold almost surely; that is, they may fail on a set of measure zero in $\mathcal{F}_n$. Indeed, condition (4) uses a regular conditional probability that may only be well-defined up to sets of measure zero in $\mathcal{F}_n$.

To see how this generalizes the 2-color urn model, suppose we have a two-color urn. At time n we add a random number of balls of each color, where the joint distribution of the two random numbers of balls added may depend on the entire history of the process. If we let $S_n = S(n)$ be the fraction of balls of color 1 in the urn at time n, and $X_n = S_n - S_{n-1}$, then (3) - (5) are satisfied for a wide class of urn schemes including the one in the previous section.

Theorem 3.5 *Let* $\{X_i, S_i : i \geq 1\}$ *satisfy (3) - (5) above. Then* $prob(S_n \to p \text{ as } n \to \infty) = 0.$

Later we will sketch a proof of essentially the same theorem but with weaker hypotheses. The reader interested in reading only a sketch of this proof should read that version beginning after lemma 5.18 on page 102.

Proof of theorem 3.5: If $prob(S_n \to p) > 0$ then there is some n and some event $\mathcal{A} \in \mathcal{F}_n$ such that $prob(S_n \to p \,|\, \mathcal{A})$ is arbitrarily close to 1. In fact, n can be taken to be as large as desired and in particular, we can assume n is large enough so that conditions (3) - (5) are satisfied. So it suffices to show that there is a constant $a > 0$ determined by c_0, c_1, c_2 and p so that $prob(S_n \to p \,|\, \mathcal{F}_n) \leq 1 - a$ for sufficiently large n. To do this, we establish two claims.

$$\text{Claim 1: } prob(\sup_{k \geq n} |S_k - p| > \frac{c_3}{\sqrt{n}} \,|\, \mathcal{F}_n) \geq 1/2 \tag{6}$$

where

$$c_3 = \sqrt{\frac{c_2}{32}}. \tag{7}$$

$$\text{Claim 2: } prob(\inf_{k \geq n} |S_k - p| \geq \frac{c_3}{2\sqrt{n}} \,|\, \mathcal{F}_n, \mathcal{A}) \geq 2a \tag{8}$$

where $\mathcal{A}$ is the event $|S_n - p| \geq c_3/\sqrt{n}$ and $a = 1/2 \min(1/2, c_3{}^2/16c_1)$.

Putting these two claims together, we see that for any value of S_n, the probability is at least a that some S_{n+k} will be at least $c_3/\sqrt{n}$ away from p and that no subsequent S_{n+k+l} will ever return to the interval $[p - c_3/2\sqrt{n}\,,\, p + c_3/2\sqrt{n}]$. The theorem follows.

Proof of claim 1: Let $\tau = \inf\{k \geq n : |S_k - p| > c_3/\sqrt{n}\}$. We will calculate the variance of S stopped at τ. On the one hand, this is limited by the fact that $|S_k - p|$ is never much more than $c_3/\sqrt{n}$ for $k \leq \tau$. On the other hand, condition (5) forces the variance to keep up a certain minimum growth on the order of n^{-2} until the stopping timeis reached. When c_3 is small enough, these two facts together imply that the stopping time is reached often enough for (15) to hold. To calculate how $\mathbf{E}((S_{\tau\wedge M} - p)^2|\mathcal{F}_n)$ increases with M, we fix any $M \geq n$.

$$\begin{aligned}
&\mathbf{E}((S_{\tau\wedge(M+1)} - p)^2|\mathcal{F}_n) \\
&= \mathbf{E}((S_{\tau\wedge M} - p)^2\,|\,\mathcal{F}_n) \;+\; \mathbf{E}(\mathbf{1}_{\tau>M}(2X_{M+1}(S_M - p) + X_{M+1}{}^2)|\mathcal{F}_n) \\
&= \mathbf{E}((S_{\tau\wedge M} - p)^2\,|\,\mathcal{F}_n) \;+\; \mathbf{E}(\mathbf{E}(\mathbf{1}_{\tau>M}X_{M+1}2(S_M - p)|\mathcal{F}_M)|\mathcal{F}_n) \\
&\qquad\qquad\qquad\qquad\quad\; +\; \mathbf{E}(\mathbf{E}(\mathbf{1}_{\tau>M}X_{M+1}{}^2|\mathcal{F}_M)|\mathcal{F}_n) \\
&= \mathbf{E}((S_{\tau\wedge M} - p)^2\,|\,\mathcal{F}_n) \;+\; \mathbf{E}(\mathbf{E}(\mathbf{1}_{\tau>M})\mathbf{E}(X_{M+1}(S_M - p)|\mathcal{F}_M)|\mathcal{F}_n) \\
&\qquad\qquad\qquad\qquad\quad\; +\; \mathbf{E}(\mathbf{E}(\mathbf{1}_{\tau>M})\mathbf{E}(X_{M+1}{}^2|\mathcal{F}_M)|\mathcal{F}_n) \qquad (9) \\
&\geq \mathbf{E}((S_{\tau\wedge M} - p)^2\,|\,\mathcal{F}_n) \;+\; \mathbf{E}(\mathbf{1}_{\tau>M}\frac{c_2}{(M+1)^2}|\mathcal{F}_n) \qquad (10)
\end{aligned}$$

since the middle term in (9) is non-negative by (4) and using the lower bound in (5),

$$\geq \mathbf{E}((S_{\tau\wedge M} - p)^2\,|\,\mathcal{F}_n) + \frac{c_2}{(M+1)^2}\,prob(\tau = \infty\,|\,\mathcal{F}_n).$$

Then by induction

$$\begin{aligned}
&\mathbf{E}((S_{\tau\wedge M} - p)^2\,|\,\mathcal{F}_n) \geq (S_n - p)^2 + prob(\tau = \infty\,|\,\mathcal{F}_n)c_2 \textstyle\sum_{k=n+1}^{M} \frac{1}{k^2} \\
&\geq c_2\, prob(\tau = \infty\,|\,\mathcal{F}_n)\left(\frac{1}{4n} - \frac{1}{4M}\right).
\end{aligned}$$

But $|S_{\tau\wedge M} - p| < c_3/\sqrt{n} + c_0/n^\gamma$, so for n sufficiently large $c_3/\sqrt{n} > c_0/n^\gamma$, so $2c_3/\sqrt{n} >$

$|S_{\tau\wedge M} - p|$ and therefore

$$\frac{4c_3{}^2}{n} \geq \mathrm{E}((S_{\tau\wedge M} - p)^2 \,|\, \mathcal{F}_n) \geq c_2 \, prob(\tau = \infty \,|\, \mathcal{F}_n) \left(\frac{1}{4n} - \frac{1}{4M} \right)$$

and letting $M \to \infty$, $prob(\tau = \infty \,|\, \mathcal{F}_n) \leq 16c_3{}^2/c_2 = 1/2$, so $prob(\sup_{k\geq n} |S_k - p| > c_3/\sqrt{n} \,|\, \mathcal{F}_n) = 1 - prob(\tau = \infty \,|\, \mathcal{F}_n) \geq 1 - 16c_3{}^2/c_2 = 1/2$.

Proof of Claim 2: The idea this time is that the variance of the variables $\{X_k : k \geq n\}$ is not enough to give a high probability of getting back to within $c_3/2\sqrt{n}$ of p. The inequality we use is a one-sided Tchebysheff inequality relying on the fact that the expectation of S_n is not getting any closer to p, so if S_n has a probability of $1 - \epsilon$ of coming back within $c_3/2\sqrt{n}$ of p, then ϵ of the time it must get to a distance of order ϵ^{-1} away from p, contributing to the variance on the order of ϵ^{-2}.

Assume without loss of generality that $S_n < p - c_3/\sqrt{n}$ since the case where $S_n > p + c_3/\sqrt{n}$ is identical. We are trying to show that $prob(\sup_{k\geq n} S_k < p - c_3/2\sqrt{n}) \geq 2a = \min(1/2, c_3{}^2/16c_1)$. So let $\tau = \inf\{k \geq n : S_k \geq p - c_3/2\sqrt{n}\}$. Define a sequence of variables $Y_{n+1}, Y_{n+2}, \ldots$ by $Y_k = 0$ for $k > \tau$ and $Y_k = X_k + \mu_k$ for $\tau \geq k > n$, where $\mu_k = -\mathrm{E}(X_k | S_{k-1}) \geq 0$ by (4). So the sequence $\{Z_k : k > n\}$ is a martingale, where $Z_k = \sum_{j=n+1}^k Y_j$. By (5), $\{Z_k\}$ is L^2-bounded. It suffices to show

$$prob\left(\sum_{k=n+1}^{\tau} Y_k \geq c_3/2\sqrt{n} \right) \leq 1 - 2a \tag{11}$$

since

$$\begin{aligned}
\sup_{k\geq n} S_k \geq p - c_3/2\sqrt{n} \;&\Rightarrow\; S_\tau \geq p - c_3/2\sqrt{n} \\
&\Rightarrow \sum_{k=n+1}^{\tau} X_k \geq c_3/2\sqrt{n} \\
&\Rightarrow \sum_{k=n+1}^{\tau} Y_k \geq c_3/2\sqrt{n}
\end{aligned}$$

But

$$\mathrm{Var}\left(\sum_{k=n+1}^{\tau} Y_k\right) \leq \sum_{k=n+1}^{\infty} \frac{c_1}{n^2} \leq \frac{c_1}{n}. \tag{12}$$

Also,

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{k=n+1}^{\tau} Y_k\right) \\
\geq\ & prob(\tau < \infty)\left(\frac{c_3}{2\sqrt{n}}\right)^2 + prob(\tau = \infty)\mathbf{E}({Z_\infty}^2|\tau = \infty) \\
\geq\ & prob(\tau = \infty)\mathbf{E}(Z_\infty|\tau = \infty)^2 \qquad (13) \\
\geq\ & prob(\tau = \infty)\left(-\frac{c_3}{2\sqrt{n}}\frac{prob(\tau < \infty)}{prob(\tau = \infty)}\right)^2 \\
=\ & \frac{{c_3}^2}{4n}\frac{prob(\tau < \infty)^2}{prob(\tau = \infty)}
\end{aligned}$$

where the penultimate term is calculated from the fact that $\mathbf{E}(Z_\infty \,|\, \tau < \infty) > c_3/2\sqrt{n}$ while $\mathbf{E}(Z_\infty)$ must be zero. Combining inequalities (12) and (13) gives

$$\frac{prob(\tau = \infty)}{prob(\tau < \infty)^2} \geq \frac{{c_3}^2}{4c_1}$$

so either $prob(\tau = \infty) \geq 1/2$ or $prob(\tau < \infty) \leq ({c_3}^2/4c_1)(1/2)^2 = {c_3}^2/16c_1$ and in both cases we are done by the definition of a. □

Theorem 3.5 is really a theorem about the behaviour of S_n near the point p. The next proposition is a version of the same theorem with slightly weaker hypotheses that reflect the local nature of the result.

Proposition 3.6 *With notation as in the statements of conditions (3) - (5), suppose there exists a neighborhood $\mathcal{N}$ of p and a constant N such that (3) - (5) hold whenever $S_n \in \mathcal{N}$ and $n > N$. Then $prob(S_n \to p \text{ as } n \to \infty) = 0$.*

Proof: If $prob(S_n \to p) > 0$ then there is some $n > N$ and some event $\mathcal{A} \in \mathcal{F}_n$ so that $prob(S_k \in [p - \epsilon, p + \epsilon]$ for all $k \geq n \,|\, \mathcal{A}) > 0$. Let $\tau = \inf\{k \geq n : S_n \notin [p - \epsilon, p + \epsilon]\}$.

Construct a sequence of random variables, $Y_1, Y_2, \ldots$ by letting $Y_k = S_k$ for $k \le \tau$ and letting Y evolve independently from X after time τ in any way that makes the process $\{Y_i - Y_{i-1}\}$ satisfy (3) - (5) whether or not $S_n \in \mathcal{N}$. Then $Y_k \to p$ with non-zero probability which is a contradiction. □

At this point, we could derive as a corollary a case of theorem 3.3, namely the case where $n\delta_n$ is bounded between two constants. This has as a subcase any F that is approximately polynomial in n. Instead, we will derive theorem 3.3 in the general case by essentially repeating the proof of theorem 3.5. The tradeoff will be that because $F(n)$ is fixed, we know the approximate sizes of ${X_n}^2$ in advance. We are then free to replace the $c_3/\sqrt{n}$ in claims 1 and 2 above by appropriate other values and the condition (5) becomes irrelevant. It should be noted that condition (5) is vital in the absence of such advance information, since without it the sequence $\{S_n\}$ is just a $[0,1]$-valued martingale whose increments get small fast enough; it is easy to find one of these that converges to some point in $(0,1)$ with nonzero probability.

Proof of theorem 3.3: Define

$$\alpha_n = \sum_{i=n}^{\infty} {\delta_i}^2.$$

According to theorem 3.2 there is no loss of generality in assuming α_n to be finite. Fix $p \in (0,1)$. Also assume without loss of generality that $p \le 1/2$ since the case $p > 1/2$ is identical but with red balls and black balls interchanged. We will prove versions of claims 1 and 2 above. Since $\alpha_n \to 0$ there is an N for which $n \ge N$ implies $\alpha_n < p/10$. Choose c small enough so that $c < 1$ and

$$9c^2 \le 81p^2/800. \tag{14}$$

$$\text{Claim 1}': \; prob(\sup_{k \ge n} |S_k - p| > c\sqrt{\alpha_n} \mid \mathcal{F}_n) \ge \min\{1/2, 9p/10\} \tag{15}$$

Proof: Let $\tau = \inf\{k \ge n : |S_k - p| > c\sqrt{\alpha_n}\}$. We need to show that $prob(\tau < \infty) \ge \min\{1/2, 9p/10\}$.

Case 1: $\delta_i > 2c\sqrt{\alpha_n}/(1 - p - c\sqrt{\alpha_n})$ for some $i \geq n$. Basically what happens in this case is that there is a good enough chance of stopping on the $i + 1^{st}$ draw:

$$\tau > i \text{ and draw } i \text{ is red}$$

$$\Rightarrow \quad v_i \geq p - c\sqrt{\alpha_n} \text{ and draw } i \text{ is red}$$

$$\Rightarrow \quad v_{i+1} > p + c\sqrt{\alpha_n}$$

$$\Rightarrow \quad \tau = i + 1.$$

Since the probability of a red draw is always at least $p - c\sqrt{\alpha_i} > 9p/10$ until τ is reached, this easily implies that $prob(\tau < \infty) \geq 9p/10$.

Case 2: No δ_i is that big. Then the increment on which τ is reached cannot be bigger than $2c\sqrt{\alpha_n}$, and so

$$|v_{i\wedge\tau} - p| \leq 3c\sqrt{\alpha_n} \text{ for all } i \geq n. \tag{16}$$

Pick any $i \geq n$ and use the fact that v_i is a martingale to get

$$\mathbf{E}((v_{(i+1)\wedge\tau} - p)^2|\mathcal{F}_n) = \mathbf{E}(v_{i\wedge\tau}{}^2|\mathcal{F}_n) + \mathbf{E}(\mathbf{1}_{\tau>i}(v_{i+1} - v_i)^2|\mathcal{F}_n). \tag{17}$$

But

$$(v_{i+1} - v_i)^2 = \begin{cases} v_i{}^2(\delta_i/(1+\delta_i)^2) \text{ with probability } 1 - v_i \\ (1 - v_i)^2(\delta_1/(1+\delta_i))^2 \text{ with probability } v_i \end{cases} . \tag{18}$$

Now since $1 + \delta_i < 2$ by the assumption that $\alpha_n < p/10$ and since $\tau > i \Rightarrow \min\{v_i, 1 - v_i\} \geq p - c\sqrt{\alpha_n} \geq 9p/10$, it follows that

$$(v_{i+1} - v_i)^2 \geq 81p^2\delta_i{}^2\mathbf{1}_{\tau>i}/400.$$

So the right hand side of equation (17) is at least

$$\mathbf{E}((v_{i\wedge\tau})^2|\mathcal{F}_n) + 81\delta_i{}^2 prob(\tau = \infty|\mathcal{F}_n)/400$$

and summing over i gives

$$\mathbf{E}((v_{(n+M)\wedge\tau} - p)^2|\mathcal{F}_n) \geq \left(81p^2 \sum_{i=n}^{M-1} {\delta_i}^2/400\right) prob(\tau = \infty|\mathcal{F}_n).$$

But equation (16) implies that

$$\mathbf{E}((v_{(n+M)\wedge\tau} - p)^2|\mathcal{F}_n) \leq 9c^2\alpha_n = 9c^2 \sum_{i=n}^{\infty} {\delta_i}^2,$$

so letting $M \to \infty$ gives

$$prob(\tau = \infty) \leq 9c^2/(81p^2/400) \leq 1/2$$

by the choice of c in (14) above. So claim $1'$ is proved. Now assume $|v_n - p| > c\sqrt{\alpha_n}$ with the same conditions on c, n and α_n as in claim $1'$.

$$\text{Claim } 2'\text{: } prob(\inf_{k \geq n} |v_k - p| \geq c\sqrt{n}/2 \,|\, \mathcal{F}_n) \geq 1 - \min\{1/2, c^2/16\} \tag{19}$$

Proof: From (17) again, calculate

$$\mathrm{Var}(v_{n+M}|\mathcal{F}_n) = \sum_{i=n}^{M-1} \mathbf{E}(\mathbf{1}_{\tau > i}(v_{i+1} - v_i)^2|\mathcal{F}_n) \leq \sum_{i=n}^{\infty} {\delta_i}^2$$

according to the values for $(v_{i+1} - v_i)^2$ given in (18). So $\{v_{n+i}\}$ is an L^2-bounded martingale with variance at most $\sum_{i=n}^{\infty} {\delta_i}^2 = \alpha_n$. Then the one-sided Tschebysheff inequality used in equation (13) and following shows that

$$\alpha_n \geq (c^2/4)\alpha_n \frac{prob(\tau < \infty)^2}{prob(\tau = \infty)}$$

so $prob(\tau = \infty) \geq \min\{1/2, c^2/16\}$ and claim $2'$ is proved.

Theorem 3.3 follows from claims $1'$ and $2'$ in the same way that theorem 3.5 follows from claims 1 and 2. □

3.3 Non-linear draws and random additions of balls

The most general urn scheme we consider is the following d-color scheme. Let $\vec{F}(n)$ be a random vector and let $F_i(n)$ balls of color i be added at time n. By conditioning on the past, we can think of this random vector as having a distribution which is itself an $\mathcal{F}_n$-measurable function. In the examples below, this distribution will often only depend on $\vec{v}(n)$, but in general the distribution of $\vec{F}(n)$ and other related quantities may depend on the entire past.

The formal definition (1) be generalized in the obvious way; instead of $S_1(n)$ and $S_2(n)$, we will have a vector $\vec{S}(n) = S_1(n), \ldots, S_d(n)$, keeping track of the numbers of balls of each color present at time n. Let $\vec{v}(n) = \vec{S}(n)/\sum_{i=1}^d S_i(n)$ keep track of the proportions of each color present at time n, so $\vec{v}(n)$ is analagous to q_n. Note that $\vec{v}(n)$ must lie on the unit simplex $\Delta \stackrel{def}{=} \{\vec{v} : \sum v_i = 1 \text{ and for each } i, v_i \geq 0\}$. Let $\delta_n = \sum_{k=1}^d F_k(n)/\sum_{k=1}^d S_k(i)$ be the fractional addition vector, let $\hat{F}(n)$ be the unit vector in the direction $\vec{F}(n)$, and let $\vec{G}(n) = \frac{\delta_n}{1+\delta_n}(\hat{F} - I)\vec{v}(n) = \vec{v}(n+1) - \vec{v}(n)$. Note that $\vec{G}(n)$ always lies in the subspace $W \stackrel{def}{=} \{\vec{v} : \sum_{k=1}^d v_k = 0\}$ since $\vec{v}(n+1)$ and $\vec{v}(n)$ are both in Δ. Write $\overline{G}(n) = \mathbf{E}(\vec{G}(n) \,|\, \mathcal{F}_n)$. Typically the direction of $\overline{G}$ is a function solely of $\vec{v}(n)$.

This scheme is general enough so its features are really that of a stochastic approximation scheme, although nothing is being approximated. It has most of the urn schemes we have already discussed as special cases.

(1) When $\vec{F}(n)$ is $f(n)$ times the standard basis vector e_j with probability $v_j(n)$ for $1 \leq j \leq d$, this gives the model discussed in section 1.

(2) When $d = 2$ and $\vec{F}(n) = (1, 0)$ with probability $f(\vec{v}(n))$ and $(0, 1)$ with probability $1 - f(\vec{v}(n))$, this gives the model in [HLS].

(3) When $f_i(n)$ are independent, identically distributed positive integer random variables and $\vec{F}(n) = f(n)e_j$ with probability $v_j(n)$ for $1 \leq j \leq d$, this gives the urn scheme in [At].

Of course these generalizations may be combined in countless ways, and the methods of the previous section will yield various theorems about which points cannot be the limit of $\vec{v}(n)$ with positive probability. Instead of trying to state a theorem of maximal generality, we will derive a reasonably broad result that is adequate for use in chapter five. The exposition will be terse in places, referring the reader ahead to chapter five for details. The reason is that the application of this material in chapter five is to a non-linear urn process for which we know the urn function only approximately. Because it is not completely straightforward, we cannot avoid repeating the calculations of this section in detail, and therefore choose only to summarize them here.

In the two cases we consider, namely the ones in section 1 of this chapter and the one in chapter five, we already have convergence theorems. In several other cases that have been studied, convergence theorems have also already been obtained (see [NH]). Instead we will try to generalize [HLS] and [Ar] by finding conditions on $\vec{F}$ near a point $\vec{p}$ that imply $prob(\vec{v}(n) \to \vec{p}) = 0$. The question of when $\vec{v}(n)$ converges for a generalized urn process is too difficult to address here; it seems to depend a lot on the details of the random vector fields $\vec{F}$ and $\vec{G}$.

Impose the following restrictions on $\vec{F}$ via restrictions on $\vec{G}$ and $\overline{G}$.

$$\begin{array}{c}\text{There is a } \gamma > 1/2 \text{ such that for all } \epsilon > 0 \text{ there is an} \\ N \text{ such that } prob(|\vec{G}(n)| < 1/n^\gamma \text{ for all } n > N) > 1 - \epsilon.\end{array} \tag{20}$$

$$\mathbf{E}(|\vec{G}(n)|^2 \,|\, \mathcal{F}_n) \leq c/n^2 \text{ for some constant } c. \tag{21}$$

Condition (20) is a multidimensional version of condition (3) that has been weakened by saying it only has to hold on sets of measure arbitrarily close to 1. Condition (21) is a version of one of the first inequality in condition (5). Note that these are satisfied with $\gamma = 1$ when $|\vec{F}| \equiv 1$ as in cases (2) and (3) above, or more generally when $|\vec{F}|$ is bounded above and below by polynomials of the same degree as in case (1).

The other two conditions we impose on $\vec{F}$ are local conditions. Let $\vec{p}$ be any point in the interior of Δ, so $p_i > 0$ for all i, and let $\mathcal{N}$ be any neighborhood. We require that There is some constant c for which, whenever $\vec{v}(n) \in \mathcal{N}$,

$$prob(\vec{G}(n) \cdot \vec{\theta} > c|\vec{\theta}|/n) > c \tag{22}$$

for all $\vec{\theta} \in W$. Again observe that this condition is satisfied in cases (1) - (3) above. Basically, the condition states that $\vec{G}(n)$ is of order $1/n$ and is not restricted to any narrow band inside W. It is the counterpart to the second inequality in condition (5). To generalize the notion of an upcrossing in [HLS], suppose that there is a twice continuously differentiable function η with $\vec{\nabla}\eta(\vec{p})$ nonzero such that either

$$\vec{\nabla}\eta(\vec{v}(n)) \cdot \overline{G}(n) \text{ has the same sign and a greater magnitude than } \lambda\eta(\vec{v}(n))/n \tag{23}$$

for some constant $\lambda > 0$, or

$$\eta \text{ is linear and } (\vec{\nabla}\eta(\vec{v}(n)) \cdot \overline{G}(n))\eta(\vec{v}(n)) \geq 0 \tag{24}$$

whenever $\vec{v}(n) \in \mathcal{N}$. Note that (24) is satisfied in cases (1) and (3) for any nondegenerate linear η vanishing at $\vec{p}$ because $\overline{G} \equiv 0$. In case (2), $\eta(v_1, v_2)$ can be taken to be $v_1 - p_1$. Then whenever the hypotheses of Hill, Lane and Sudderth's theorem 2.4 are satisfied, condition (23) is satisfied with λ being any positive lower bound for $F'(p)$ on a neighborhood of p. Conditions (23) and (24) are of course the counterparts to the splitting condition (4).

Theorem 3.7 *Let $\vec{p}, \mathcal{N}, \vec{F}, \vec{G}, \overline{G}$ and η satisfy conditions (20) - (22) and either (23) or (24). Then $prob(\vec{v}(n) \to \vec{p}) = 0$.*

Proof: Letting $S_n = \eta(\vec{v}(n))$, we use theorem 3.5, only with slightly weaker conditions. The modified hypotheses of the theorem are as follows:

For any $\epsilon > 0$ there is an n and a neighborhood $\mathcal{N}$ of 0 such that

$$prob(\mathcal{B} \mid \mathcal{F}_n) > 1 - \epsilon \tag{25}$$

where $\mathcal{B} = \bigcap_{k \geq 0} \mathcal{B}_k$ and $\mathcal{B}_k \in \mathcal{F}_k$ is the event that either (26) - (29) below are satisfied, or $S_n \notin \mathcal{N}$.

$$\mathbf{E}({X_{n+1}}^2 + 2X_{n+1}S_n \, \mathcal{F}_n) \geq b_1/n^2 \tag{26}$$

$$\mathbf{E}(X_{n+1}S_n 1_{|S_n|>c/n} \mid \mathcal{F}_n) \geq 0 \tag{27}$$

$$|X_n| \leq 1/n^\gamma \tag{28}$$

$$\mathbf{E}({X_{n+1}}^2 \mid \mathcal{F}_n) \leq b_2/n^2 \tag{29}$$

where

$$\begin{aligned} b_1, b_2, c &> 0 \\ \gamma &> 1/2 \\ \mathcal{F}_n &= \sigma(\text{all events up to time } \tau_n). \end{aligned}$$

The proof of this version of theorem 3.5 is given in chapter 5, section 5. It is almost identical to the proof of the first version of theorem 3.5 and at any rate, we will not repeat it here. We need to show now that (26) - (29) are satisfied when $S_n = \eta(\vec{v}(n))$. First note that by continuity, η is a bounded operator in a neighborhood of $\vec{p}$, so (28) follows from (20) and (29) follows from (21).

By a linear estimate we get

$$\begin{aligned} S_{n+1} &= \eta(\vec{v}(n+1)) \\ &= \eta(\vec{v}(n) + \vec{G}(n)) \\ &= S_n + \vec{\nabla}\eta(\vec{v}(n)) \cdot \vec{G}(n) + O(|\vec{G}(n)|^2). \end{aligned}$$

The last term drops out if η is linear. Now if (24) holds, taking the expected value and subtracting S_n gives a quantity of the same sign as S_n, so (27) is satisfied. And if (23) holds, we use (21) to get $\mathbf{E}(S_{n+1} \,|\, \mathcal{F}_n) - S_n > \lambda S_n/n + O(1/n^2)$. This is the same sign as S_n whenever $S_n > c/n$ for some appropriate c, and (27) follows.

Finally, it follows from (22) and the fact that $|\vec{\nabla}\eta|$ is bounded away from 0 near $\vec{p}$ that $\mathbf{E}({X_{n+1}}^2)$ is at least b/n^2 for some $b > 0$ on a neighborhood of $\vec{p}$. Then picking any $b_1 < b$, the condition (26) holds when $|S_n| > c/n$ by (27) and when $|S_n| < c/n$ for sufficiently large n because $\mathbf{E}(2X_{n+1}S_n \,|\, \mathcal{F}_n) \geq \lambda {S_n}^2 + O(1/n^2)S_n \geq O(1/n^3)$. □

We conclude this section with a discussion of condition (23). As previously indicated, we would like condition (23) to be satisfiable whenever the linear approximation to $\overline{G}$ near $\vec{p}$ has at least one eigenvalue with positive real part, or in terms of the continuous flow, whenever the autonomous system $d(\vec{v}(t))/dt = \overline{G}(\vec{v}(t))$ has an unstable equilibrium at $\vec{p}$. We can obtain this result by placing an extra condition on the eigenvalues of the linear approximation to $\overline{G}$ near $\vec{p}$.

Theorem 3.8 *Suppose $\overline{G}(n) = f(n, \vec{v}(n))\hat{G}(\vec{v}(n))$ where f is a scalar function and $\hat{G}$ is a vector function of $\vec{v}(n)$ alone. Pick any $\vec{p}$ with $\hat{G}(\vec{p}) = \vec{0}$ and let $T\vec{v}$ be the linear approximation to $\hat{G}(\vec{p} + \vec{v})$. Suppose the eigenvalues of T are distinct and linearly independent over the rationals with none purely imaginary and at least one having positive real part. Then $prob(\vec{v}(n) \to \vec{p}) = 0$.*

Proof: An argument is given in chapter five following proposition 5.23 that is valid whenever the eigenvalue with posive real part is real. If not, then the eigenvector corresponding to this eigenvalue has a conjugate corresponding to the conjugate eigenvalue, and together they span a two-dimensional space over $\mathbf{C}$ which contains a real two-dimensional space. For this case, use the argument in chapter five, but replace projection onto the direction of the eigenvector with magnitude of the projection onto this two-dimensional real space. □