Doob's h-transform: theory and examples

Alex Bloemendal

DRAFT: May 18, 2010

Abstract

We discuss what it means to condition a Markov process on its exit state, using the resulting theory to consider various examples of conditioned random walk and Brownian motion. The treatment is informal with an emphasis on computations rather than proofs.

Consider a *time-homogeneous Markov process* X_t on a *state space* S with *transition kernel* $P^t(x, dy)$. Up to some measurability issues, this means the following: For each $x \in S$ there is a probability measure $\mathbf{P}_x$ on the space Ω of *paths* $\{X_t\} \subseteq S$ indexed over $t \in \mathbb{Z}_+$ or $t \in \mathbb{R}_+$, i.e. time can be discrete or continuous; for each $x \in S$ and $t \geq 0$, $P^t(x, \cdot)$ is a probability measure on S; if P^t operates on bounded functions $f : S \to \mathbb{R}$ by $P^t f = \int_S P^t(\cdot, dy) f(y)$, then $\{P^t\}$ forms a semigroup of operators, i.e. $P^s P^t = P^{s+t}$ for $s, t \geq 0$; and, of course, the *Markov property* holds:

$$\mathbf{E}_x[\, f(X_{t+s}) \mid \mathcal{F}_t \,] = P^s f(X_t), \qquad \mathbf{P}_x\text{-a.s.},$$

where $\mathcal{F}_t$ is the *natural filtration* generated by $\{X_s\}_{0 \leq s \leq t}$. Recall that a more general Markov property holds in which $f(X_{t+s})$ is replaced with a bounded function of the future $\{X_s\}_{s \geq t}$: if $F : \Omega \to \mathbb{R}$ is bounded and $\theta_t : \Omega \to \Omega$, $\{X_s\}_{s \geq 0} \mapsto \{X_{t+s}\}_{s \geq 0}$ is the *time shift*, then

$$\mathbf{E}_x[\, F \circ \theta_t \mid \mathcal{F}_t \,] = \mathbf{E}_{X_t} F, \qquad \mathbf{P}_x\text{-a.s.} \tag{1}$$

We want a general way of discussing the "exit state" of X_t, that is, "where it goes when $t \to \infty$." It is natural to consider *shift-invariant* functions $H = H \circ \theta_t$ or events $A = \theta_t^{-1} A$ (all $t \geq 0$), as these depend only on the infinite future. Let $\mathcal{I}$ denote the corresponding sigma field. As it turns out, $\mathcal{I}$ is intimately connected with the *bounded harmonic functions* on S; these are bounded h satisfying $P^t h = h$ for all t, or equivalently, such that $h(X_t)$ is a martingale under $\mathbf{P}_x$ for each x. On the one hand, a bounded $H \in \mathcal{I}$ gives rise to a bounded function $h(x) = \mathbf{E}_x H$ which is harmonic by (1): $h(X_t) = \mathbf{E}_x[\, H \mid \mathcal{F}_t \,]$, manifestly a martingale. On the other hand, given a bounded harmonic function h, the limit $H = \lim_{t \to \infty} h(X_t)$ exists $\mathbf{P}_x$-a.s. by the martingale convergence theorem, and clearly $H \in \mathcal{I}$; we can then use bounded convergence to recover $h(x) = \mathbf{E}_x H$. As long as all the states

x communicate, the measures induced on $\mathcal{I}$ will be mutually absolutely continuous; in this case $L^\infty(\mathcal{I})$ is a well-defined object, known as the *Poisson boundary.*

It is elementary to condition our process pathwise on an event $A \in \mathcal{I}$ of positive probability. Let $h(x) = \mathbf{P}_x(A)$ be the corresponding harmonic function and $\widetilde{S} = \{x \in S : h(x) > 0\}$ the set of states from which A is accessible. For $x \in \widetilde{S}$, the conditioned path measure $\widetilde{\mathbf{P}}_x \ll \mathbf{P}_x$ and is given by

$$d\widetilde{\mathbf{P}}_x = \frac{\mathbf{1}_A}{h(x)} d\mathbf{P}_x.$$

Restricted to $\mathcal{F}_t$, this becomes

$$d\widetilde{\mathbf{P}}_x\big|_{\mathcal{F}_t} = \mathbf{E}_x\Big[\frac{\mathbf{1}_A}{h(x)} \Bigm| \mathcal{F}_t\Big] d\mathbf{P}_x\big|_{\mathcal{F}_t} = \frac{h(X_t)}{h(x)} d\mathbf{P}_x\big|_{\mathcal{F}_t}.$$

Our starting point is the observation is that the conditioned process is also Markov.

Theorem 1. *Under $\widetilde{\mathbf{P}}_x$ with $x \in \widetilde{S}$, X_t is a time-homogeneous Markov process on $\widetilde{S}$ with transition kernel*

$$\widetilde{P}^t(x, dy) = \frac{h(y)}{h(x)}\, P^t(x, dy). \tag{2}$$

This formula is known as Doob's *h-transform.* In terms of measures, (2) expresses that the conditioned transition probability is absolutely continuous with respect to the unconditioned one and gives a formula for the Radon-Nikodym derivative. In terms of operators, (2) writes $\widetilde{P}^t = h^{-1} P^t h$; here h is acting diagonally, i.e. by multiplication.

Proof. First, $\widetilde{P}^t(x, \cdot)$ is a probability measure on $\widetilde{S}$ because h is harmonic and zero off $\widetilde{S}$; the semigroup property holds because $\widetilde{P}^t$ is just a conjugate of P^t. As for $\widetilde{\mathbf{P}}_x$, it is a probability measure on paths in $\widetilde{S}$ because $\widetilde{\mathbf{P}}_x\big(h(X_t) > 0\big) = \mathbf{E}_x\, \mathbf{1}_{h(X_t)>0} h(X_t)/h(x) = 1$.

Informally, the Markov property is inherited by the conditioned process and (2) is just Bayes' rule:

$$\mathbf{P}_x(X_{t+s} \in dy \mid A, \mathcal{F}_t) = \frac{\mathbf{P}_x(A \mid X_{t+s} = y, \mathcal{F}_t)\, \mathbf{P}_x(X_{t+s} \in dy \mid \mathcal{F}_t)}{\mathbf{P}_x(A \mid \mathcal{F}_t)} = \frac{h(y)}{h(X_t)} P^s(X_t, y).$$

Rigorously, use the Markov property (1) for the unconditioned process (and the absolute continuity) to write the desired Markov property for the conditioned process as

$$\widetilde{\mathbf{E}}_x[\, f(X_{t+s}) \mid \mathcal{F}_t\,] = h^{-1}(X_t)\, \mathbf{E}_x[\, h(X_{t+s}) f(X_{t+s}) \mid \mathcal{F}_t\,], \qquad \widetilde{\mathbf{P}}_x\text{-a.s.}$$

To establish it, note the right-hand side is an $\mathcal{F}_t$-random variable, let $B \in \mathcal{F}_t$, and compute

$$\begin{aligned}
&\widetilde{\mathbf{E}}_x\, h^{-1}(X_t)\, \mathbf{E}_x[\, h(X_{t+s}) f(X_{t+s}) \mid \mathcal{F}_t\,] \mathbf{1}_B \\
&= \mathbf{E}_x\ h^{-1}(X_t)\, \mathbf{E}_x[\, f(X_{t+s}) h(X_{t+s}) \mathbf{1}_B \mid \mathcal{F}_t\,] h(X_t)/h(x) \\
&= \mathbf{E}_x\ f(X_{t+s}) \mathbf{1}_B h(X_{t+s})/h(x), \\
&= \widetilde{\mathbf{E}}_x\, f(X_{t+s}) \mathbf{1}_B.
\end{aligned}$$

□

Many of our examples will fit into the above framework by considering an *absorbing boundary* $\partial S \subseteq S$, meaning $P^t(x,\cdot)$ is just δ_x for $x \in \partial S$. The process stops when encountering ∂S in the sense that $X_t = X_{t\wedge T}$ where $T = \inf\{t : X_t \in \partial S\}$ is the hitting time. Observe that information about where X_t lands in ∂S is contained in $\mathcal{I}$: if $Z \subseteq \partial S$, then $A = \{X_T \in Z\} \in \mathcal{I}$. (Sometimes, for example with a slit domain, one has to be more careful and redefine ∂S appropriately.)

Example 1. Simple random walk on $\mathbb{Z} \cap [0, M]$ conditioned to hit M before 0. Here $S = \{0, \dots, M\}$, $\partial S = \{0, M\}$, $Z = \{M\}$. Solving the discrete Dirichlet problem gives $h(i) = i/M$, $0 \le i \le M$; this is "gambler's ruin." Hence for $0 < i < M$ we have $\widetilde{P}(i,j) = (j/2i)\mathbf{1}_{|j-i|=1}$. Notice that M does not appear. For the asymmetric simple walk whose steps are positive with probability $0 < p < 1$, one can find h by solving the linear recurrence; the result is $h(i) = (1-r^i)/(1-r^M)$, where $r = (1-p)/p$. Once again, M does not appear in the transition probabilities.

Example 2. Simple random walk on $\mathbb{Z}^3$ conditioned to hit some finite set Z. Here $S = \mathbb{Z}^3$, $\partial S = Z$, so $A = \{T < \infty\}$ and $h(x) = \mathbf{P}_x(T < \infty)$. Interesting fact (stated in discrete time): $(I - P)h(x) = \mathbf{P}_x(X_0 \in Z \text{ but } X_t \notin Z \text{ for } t > 0)$. In the language of electrical networks, if Z is a conductor held at unit voltage with respect to a ground at infinity, then h is the induced potential, and its "discrete Laplacian" represents the source of the induced current flow through the network (which is the gradient of h); the total current flowing, which equals the total mass of the source, is called the "capacity" of Z; this quantity is maximal over all sources on Z whose potentials (normalized to be zero at infinity) nowhere exceed one (Dynkin and Yushkevich 1969).

Example 3. Random walk on the rooted d-regular tree, conditioned to end up among the descendants of a given child v of the root u. The symmetry makes it easy to compute h. For example, $h(u) = 1/d$ and $h(v) = (d-1)/d$; in general, h decreases by a factor of $d-1$ as you step "up" toward u or away from v, and $1-h$ decreases by a factor of $d-1$ as you step "down" toward v or away from u. The words "up" and "down" can be pictured in terms of the flow induced by h, which incidentally has total current $(d-2)/d$.

For a general reversible Markov Chain on a countable state space, i.e. a random walk on a network, if P^t is induced by conductances c_{xy}, then $\widetilde{P}^t$ is induced by conductances $\widetilde{c}_{xy} = h(x)h(y)c_{xy}$. In a sense, the conditioned walk behaves as the unconditioned walk but is biased by h, "going with the flow."

Turning now to the continuous time setting where X_t has infinitesimal generator $L = \frac{d}{dt}\big|_{t=0} P^t$, i.e. $Lf(x) = \lim_{t\downarrow 0} \mathbf{E}_x\big(f(X_t) - f(x)\big)/t$ (or equivalently $P^t = e^{tL}$), the conditioned process has generator

$$\widetilde{L} = h^{-1}Lh. \tag{3}$$

Here h is also L-*harmonic*, meaning $Lh = 0$. These statements are obtained by differentiating the corresponding statements about P^t at $t = 0$.

In the special case of a diffusion $dX_t = \sigma(X_t)dB_t + b(X_t)dt$ in $\mathbb{R}^d$, we have

$$L \;=\; \tfrac{1}{2}\sigma\sigma^\dagger\!:\!\nabla\nabla^\dagger + b\!\cdot\!\nabla \;=\; \tfrac{1}{2}\sum_{i,j} a_{ij}\frac{\partial^2}{\partial x_i \partial x_j} + \sum_i b_i \frac{\partial}{\partial x_i}$$

where $a = \sigma\sigma^\dagger$. We can use (3) and L-harmonicity to compute that

$$\widetilde{L} = L + a\frac{\nabla h}{h}\cdot\nabla. \tag{4}$$

We can also see this using stochastic calculus. On the one hand, the absolute continuity $\widetilde{\mathbf{P}}_x\big|_{\mathcal{F}_t} \ll \mathbf{P}_x\big|_{\mathcal{F}_t}$ already implies the diffusion coefficients coincide, and the change of measure induced by an additional drift term $\widetilde{b} - b$ is given by the Cameron-Martin-Girsanov formula:

$$\frac{d\widetilde{\mathbf{P}}_x}{d\mathbf{P}_x}\bigg|_{\mathcal{F}_t} = \exp\left(\int_0^t a^{-1}(\widetilde{b}-b)(X_s)\cdot dX_s - \tfrac{1}{2}\int_0^t (\widetilde{b}-b)\cdot a^{-1}(\widetilde{b}-b)(X_s)ds\right).$$

On the other hand, the Radon-Nikodym derivative is just $\mathbf{E}_x[\,\mathbf{1}_A/h(x)\mid\mathcal{F}_t\,] = h(X_t)/h(x)$. Applying Itô's lemma and $Lh = 0$ to $\log h(X_t)$, we identify $\widetilde{b} - b = a\nabla\log h$, recovering (4). To summarize, *conditioning just adds a drift in the direction of increasing h, with magnitude given by its relative increase.*

Example 4. Brownian motion on the interval $[0, c]$ conditioned to hit c before 0. Here $L = \frac{1}{2}d^2/dx^2$, so $h(x) = x/c$ and $\widetilde{L} = \frac{1}{2}d^2/dx^2 + (1/x)d/dx$; in other words, the conditioned process is the diffusion $dX_t = dB_t + (1/X_t)dt$ on $[0, c]$. Notice that c does not appear in the SDE. For the generalization to a domain in $\mathbb{R}^d$, one would solve the corresponding Dirichlet problem for the Laplacian.

Example 5. Brownian motion on the interval $[0, \pi]$ conditioned to remain in $(0, \pi)$ up to time t_1. This example initially appears to fall outside of our framework. The key is the "space-time trick": *we can recover the time-homogeneous setting in a trivial way, by enlarging the state space to include the time variable.* Here, $S = [0, \pi] \times [0, t_1]$, $\partial S = \big(\{0, \pi\} \times (0, t_1]\big) \cup \big((0, \pi) \times \{t_1\}\big)$, and $Z = (0, \pi) \times \{t_1\}$. (The boundary is absorbing as above, so the state actually includes $t \wedge T$, i.e. the process remembers both where and when it stopped.)

The 2-dimensional generator becomes $L = \frac{1}{2}d^2/dx^2 + d/dt$, i.e. $a = \mathrm{diag}\big(1, 0\big)$ and $b = \big(0, 1\big)^\dagger$. From (4), we already see that the drift term $a(\nabla h/h)\cdot\nabla = (dh/dx)(1/h)d/dx$ has only a spatial component, but that it is time-dependent.

The equation $Lh(x, t) = 0$ is the heat equation with space variable x and time $-t$. We are to solve it with initial data $h(x, t_1) = 1$ and Dirichlet boundary conditions $h(0, t) = h(\pi, t) = 0$ for $t < t_1$. The solution can be written as the Fourier series $h(x, t) = \sum_{k=1}^\infty c_k e^{-k^2(t_1 - t)}\sin(kx)$ with $c_k = (4/k\pi)\mathbf{1}_{k \text{ is odd}}$. For the generalization to a domain, one would solve the heat equation with constant initial data and Dirichlet boundary conditions.

It is natural to ask what happens as $t_1 \to \infty$; presumably the process converges to some time-homogeneous positive recurrent process. As we will see in Example 13, the answer involves the ground state eigenfunction of the Dirichlet Laplacian.

We now come to the interesting situation where we would like to condition on $A \in \mathcal{I}$, but $\mathbf{P}_x(A) = 0$ everywhere. Here the "singular" conditioning must in general be interpreted as some limit of ordinary conditionings, say on A_n with $\mathbf{P}_x(A_n) > 0$ and $\bigcap_n A_n = A$. The key observation is that *only ratios of values of h ever matter*, which allows us to renormalize h_n by scaling it up—hopefully so as to converge to a positive finite limit h. When the result of this procedure is unique (up to a multiplicative constant), it is reasonable to simply *define the conditioned process by*

$$d\widetilde{\mathbf{P}}_x\big|_{\mathcal{F}_t} = \frac{h(X_t)}{h(x)} d\mathbf{P}_x\big|_{\mathcal{F}_t}, \tag{5}$$

even though the measures themselves are now mutually singular. Significantly, Theorem 1 still holds! The proof used only (5) and the fact that h is a *positive harmonic function.* This is called "conditioning in the sense of the h-transform." (Actually, one should be more careful. While (2) certainly still defines a Markov semigroup, $h(X_t)$ is in general only a local martingale under $\mathbf{P}_x$; in (5), t may need to be replaced with $t \wedge \tau_n$ where τ_n is a localizing sequence like $\min\{t : h(X_t) \geq n\}$.)

So when does it work? The full story is that of the *Martin boundary* of a Markov chain, a topological object consisting of the ideal points of a certain compactification of the state space; its Borel measures are in correspondence with the positive harmonic functions, and the extreme measures (Dirac masses)—which correspond to extremal positive harmonic functions in the usual sense—represent exit states we can condition on. References include Kunita and Watanabe (1965), Dynkin and Yushkevich (1969), Kemeny, Snell and Knapp (1976).

Even with all the theory it is often not a simple matter to show that some part of a Martin boundary reduces to a point, much less completely analyze a given example. If one is willing to settle for heuristically justified computations, however, interesting examples abound. In practice, once you find a positive harmonic function that "does the right thing at the boundary"—and convince yourself that it is the only one—you're off to the races.

Example 6. Simple random walk on $\mathbb{Z}_+$ conditioned never to hit 0. $S = \mathbb{Z}_+$, $\partial S = \{0\}$, $Z = \varnothing$, $h(i) = i$. The conditioned walk is transient, as the resistance to infinity is finite: $\sum_{i=1}^{\infty} 1/\widetilde{c}_{i,i+1} = \sum 1/i(i+1) < \infty$. If we condition the conditioned process to return to zero, however, we recover the simple random walk: $\widetilde{h}(i) = 1/h(i) = 1/i$ is harmonic for the conditioned walk. (There are two more surprising dualities between these two processes: one involves a space- and time-reversal, and the other involves reflection about the past maximum or future minimum.)

Example 7. Brownian motion on $\mathbb{R}_+$ conditioned never to hit 0. $S = \mathbb{R}_+$, $\partial S = \{0\}$, $Z = \varnothing$, $h(x) = x$, $\widetilde{b} = 1/x$. The conditioned process is the Bessel-3 process, which is distributed like the radial process of a 3-dimensional Brownian motion. The same dualities as in the previous example are present here as well; see Williams (1974), Pitman (1975).

Example 8. Asymmetric simple random walk on $\mathbb{Z}$ with probability p of a positive step. Assuming $\frac{1}{2} < p < 1$, there are two extremal positive harmonic functions: the constant, and the exponential $h(i) = \rho^{-i}$ with $\rho = p/(1-p)$. Transforming by the latter should be interpreted as conditioning the walk to end up at $-\infty$ instead of $+\infty$; it is curious that the conditioned walk is again an asymmetric random walk but with the asymmetry exactly reversed! (The same phenomenon occurs for a one-dimensional Brownian motion with constant drift.)

Example 9. Simple random walk on the d-regular tree conditioned to end up at a given point u_* at infinity. Fix any infinite path u_n defining our point at infinity; A is the event that the walk is eventually "below" each u_n. Making the path bi-infinite, we can define the generation of a vertex v by $|v| = n - m$ where the unique path from v to u_* takes m steps to get to u_n. Then $h(v) = (d-1)^{|v|}$, inducing a constant bias toward u_* that exactly flips the bias away from u_* present in the unconditioned walk.

Example 10. Brownian motion in the hyperbolic plane, conditioned to tend to a given point at infinity. We could use the Poincaré disk or upper half-plane model; h would be the corresponding Poisson kernel. A particularly convenient choice is the upper-half plane $S = \{(x, y) : y > 0\}$ with the distinguished boundary point at infinity. The metric is conformal with linear distortion factor y, so the hyperbolic Brownian motion satisfies $(dX_t, dY_t) = Y_t(dB_t^1, dB_t^2)$; the generator has $a = \operatorname{diag}(y^2, y^2)$ and $b = 0$. The Poisson kernel for the point at infinity is simply $h(x, y) = y$. We get $\widetilde{b} = \big(0, y\big)^\dagger$, i.e. a drift of $Y_t\,dt$; in the intrinsic metric, this is a constant drift in the direction of the hyperbolic straight line to the given point at infinity.

Example 11. Brownian bridge. $S = \mathbb{R} \times [0, t_1]$, $\partial S = \mathbb{R} \times \{t_1\}$, $Z = \{(a, t_1)\}$, $h(x, t) = \frac{1}{\sqrt{t_1 - t}} e^{-(x-a)^2/2(t_1-t)}$, $\widetilde{b} = \big(\widetilde{b}_x, 1\big)^\dagger$, $\widetilde{b}_x = (a - x)/(t_1 - t)$. Here there is an alternative way to condition $\{B_t\}_{0\le t\le t_1}$ on $B_{t_1} = a$, using its structure as a Gaussian process: with $t_1 = 1$ and $a = 0$ for simplicity, one checks that B_1 and $\{B_t - tB_1\}_{0\le t\le 1}$ are uncorrelated and hence independent; then $\big(\{B_t\} \bigm| B_1 = 0\big) \stackrel{d}{=} \big(\{B_t - tB_1\} \bigm| B_1 = 0\big) \stackrel{d}{=} \{B_t - tB_1\}$. It is not difficult to verify that this process coincides with the solution of the linear SDE $dX_t = dB_t - x/(1-t)\,dt$: both are mean-zero Gaussian with covariance structure $\mathbf{E}\,X_sX_t = s(1-t)$ where $0 \le s \le t \le 1$.

Example 12. Scaled Brownian excursion. $S = \mathbb{R}_+ \times [0, t_1]$, $\partial S = \big(\{0\} \times [0, t_1]\big) \cup \big(\mathbb{R}_+ \times \{t_1\}\big)$, $Z = \{(0, t_1)\}$, $h(x, t) = \frac{x}{(t_1-t)^{3/2}} e^{-x^2/2(t_1-t)}$, $\widetilde{b}_x = 1/x - x/(t_1 - t)$. Interpretation of h as density of hitting time of zero (Rogers and Williams 1987).

Example 13. Brownian motion conditioned to remain in an interval forever. $S = [0, \pi] \times \mathbb{R}^+$, $\partial S = \{0, \pi\} \times \mathbb{R}^+$, $Z = \varnothing$, $h(x, t) = e^t \sin x$, $\widetilde{b}_x = \cot x$. Time homogeneity; stationary distribution $\frac{2}{\pi} \sin^2 x$ (cf. quasi-stationary distribution of killed process, $\frac{1}{2} \sin x$). General domain in $\mathbb{R}^d$: ground state conditioning $h = e^{\lambda_0 t}\varphi_0(x)$ (Pinsky 1985).

Example 14. Brownian motion in a Weyl chamber and Dyson's Brownian motion (for GUE). $S = \{x \in \mathbb{R}^d : x_1 \le \cdots \le x_d\}$, $\partial S = \{x \in S : x_i = x_j,\ \text{some } i \ne j\}$, $Z = \varnothing$, $h(x) = \prod_{i<j}(x_j - x_i)$, $\widetilde{b}_i = \sum_{j\ne i} 1/(x_i - x_j)$. Relation to Karlin-McGregor formula; generalizations (Grabiner 1999, Biane 2009).

We finish with an example that features an interplay of all of its extremal positive harmonic functions; the characterization of the Martin boundary here goes back to Watanabe

(1960), Blackwell and Kendall (1964). In passing, we note that the positive harmonic functions of an h-transformed process are in a simple correspondence with those of the original process: If $Ph_0 = h_0$ and $h_0 > 0$, let $\widetilde{P} = h_0^{-1} P h_0$ as operators; then $Ph = h$ iff $\widetilde{P}(h/h_0) = h/h_0$. The map $h \mapsto h/h_0$ is a linear bijection, and in particular it preserves extremality. (Essentially, the Martin boundary is invariant under h-transforms.)

Example 15. Independent die-rolls and Polya's urn. Consider an urn containing balls of d different colours; the state space is $S = \mathbb{Z}_+^d$. To begin, fix a biased d-sided die which comes up colour i with probability p_i, where

$$p = (p_1, \ldots, p_d) \ \in \ \Delta = \{(q_1, \ldots, q_d) \in \mathbb{R}^d : q_i > 0, \sum_i q_i = 1\},$$

the standard simplex. The transition dynamics are given by rolling the die independently and adding a ball of the colour that comes up. In other words, $X_n = X_0 + B_1 + \ldots + B_n$ where $B_1, B_2, \ldots \in \{\mathbf{e}_1, \ldots, \mathbf{e}_d\}$ (the standard basis vectors) are i.i.d. with distribution determined by $\mathbf{E}^p B_1 = p$. The strong law of large numbers then says that $\mathbf{P}^p(X_n/n \to p) = 1$.

Now given $q \in \Delta$, $h_q(x) = \prod_i (q_i/p_i)^{x_i}$ is a positive harmonic function for this process. In fact, the h_q-transformed process has law $\mathbf{P}^q$, i.e. it is the same process except that the p-die is replaced with a q-die! Thus $X_n/n \to q$ a.s. under the transformed dynamics, and we can identify this transform as "conditioning on $\{X_n/n \to q\}$". (It turns out that $\{h_q : q \in \Delta\}$ is precisely the set of extremal positive harmonic functions for the original process.)

Finally, we consider a uniform mixture of the laws $\mathbf{P}_0^q$ of these processes started from $X_0 = 0$. Let λ be normalized $(d-1)$-dimensional Lebesgue measure on Δ, and let $\mathbf{P}_0^\lambda = \int_\Delta \mathbf{P}_0^q \, \lambda(dq)$. Now

$$d\mathbf{P}_0^\lambda\big|_{\mathcal{F}_n} = \int_\Delta d\mathbf{P}_0^q\big|_{\mathcal{F}_n} \lambda(dq) = \int_\Delta \frac{h_q(X_n)}{h_q(0)} \lambda(dq) \, d\mathbf{P}_0^p\big|_{\mathcal{F}_n}.$$

Since $h_q(0) = 1$ for all q, the $\mathbf{P}_0^\lambda$-process is the h_λ-transform of the $\mathbf{P}_0^p$-process where

$$h_\lambda(x) = \int_\Delta h_q(x) \, \lambda(dq) = \textstyle\prod_i p_i^{-x_i} \displaystyle\int_\Delta \textstyle\prod_i q_i^{x_i} \, \lambda(dq) = \prod_i p_i^{-x_i} (d-1)! \dfrac{\prod_i x_i!}{(\sum_i x_i + d - 1)!}$$

and the integral is evaluated as the well-known normalizing factor of the Dirichlet distribution. We now easily compute the transition probabilities $P_\lambda(x, x + \mathbf{e}_i) = (x_i + 1)/\sum_j (x_j + 1)$, which are precisely those for Polya's urn started with 1 extra ball of each colour! If we replaced $\lambda(dq)$ with the more general Dirichlet$(\alpha_1, \ldots, \alpha_d)$ distribution $\prod_i q_i^{\alpha_i - 1} \lambda(dq)$ (normalized), we would obtain the transition probabilities $P_\lambda(x, x + \mathbf{e}_i) = (x_i + \alpha_i)/\sum_j (x_j + \alpha_j)$ for Polya's urn started with α_i extra balls of colour i.

We have thus stumbled upon the description of Polya's urn as an exchangeable process: First pick a biased die according to the Dirichlet distribution with parameters given by the initial condition, and then just add balls according to a sequence of independent die-rolls.

References

Biane, P. (2009). Matrix valued Brownian motion and a paper by Pólya, *Séminaire de probabilités XLII*, Vol. 1979 of *Lecture Notes in Math.*, Springer, pp. 171–185.

Blackwell, D. and Kendall, D. (1964). The martin boundary of Pólya's urn scheme, and an application to stochastic population growth, *J. Appl. Probability* **1**: 284–296.

Dynkin, E. B. and Yushkevich, A. A. (1969). *Markov processes: Theorems and problems*, Plenum Press.

Grabiner, D. J. (1999). Brownian motion in a Weyl chamber, non-colliding particles, and random matrices, *Ann. Inst. H. Poincaré Probab. Statist.* **35**(2): 177–204.

Kemeny, J. G., Snell, J. L. and Knapp, A. W. (1976). *Denumerable Markov chains*, Springer-Verlag.

Kunita, H. and Watanabe, T. (1965). Markov processes and Martin boundaries. I, *Illinois J. Math.* **9**: 485–526.

Pinsky, R. G. (1985). On the convergence of diffusion processes conditioned to remain in a bounded region for large time to limiting positive recurrent diffusion processes, *Ann. Probab.* **13**(2): 363–378.

Pitman, J. W. (1975). One-dimensional Brownian motion and the three-dimensional Bessel process, *Advances in Appl. Probability* **7**(3): 511–526.

Rogers, L. C. G. and Williams, D. (1987). *Diffusions, Markov processes, and martingales. Vol. 2*, John Wiley & Sons Inc.

Watanabe, T. (1960). A probabilistic method in Hausdorff moment problem and Laplace-Stieltjes transform, *J. Math. Soc. Japan* **12**: 192–206.

Williams, D. (1974). Path decomposition and continuity of local time for one-dimensional diffusions. I, *Proc. London Math. Soc. (3)* **28**: 738–768.