Random Walks and Hitting Probabilities

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be the unit interval $\Omega = (0,1]$ with Lebesgue measure $\mathsf{P}(dx) = dx$, and define a sequence of random variables b_n on $(\Omega, \mathcal{F}, \mathsf{P})$ by

$$b_n(\omega) = (\lfloor 2^n \omega \rfloor) \bmod 2,$$

the n^{th} "bit" in the binary expansion $\omega = \sum_n b_n 2^{-n}$. The random variable b_1 is equal to zero on $(0, 1/2]$ and one on $(1/2, 1]$; b_2 is equal to zero on $(0, 1/4] \cup (1/2, 3/4]$ and one on $(1/4, 1/2] \cup (3/4, 1]$; b_n is equal to zero on a union of 2^{n-1} intervals, each of length 2^{-n} (for a total length of $1/2$), and is equal to one on the complementary set, also of length $1/2$. Thus $\mathsf{P}[d_n = 0] = \mathsf{P}[d_n = 1] = 1/2$. (Note about simulation on 32-bit computers).

The smallest $\sigma-$algebra $\mathcal{F}_n$ over which $d_1, ..., d_n$ are all measurable is just the field consisting of all (finite) unions $\mathcal{F}_n = \{\cup_i (a_i/2^n, b_i/2^n]\}$ of left-open intervals with dyadic-rational endpoints. Each set in $\mathcal{F}_n$ can be specified by listing which of the 2^n intervals $(\frac{i}{2^n}, \frac{i+1}{2^n}]$ $(0 \le i < 2^n)$ it contains, so there are 2^{2^n} sets in $\mathcal{F}_n$ altogether.

Lebesgue measure P assigns the length $b - a$ to any interval $(a, b]$, so it is determined on $\mathcal{F}_n$ easily, but the restriction of P is also determined on $\mathcal{F}_n$ simply by the joint probability distribution of $b_1, ..., b_n$: IID Bernoulli RV's, each with $\mathsf{P}[b_i = 1] = 1/2$. For each number p, $0 < p < 1$, we can make a similar measure μ_p on $(\Omega, \mathcal{F}_n)$ by forcing $\mu_p[b_n = 1] = p$; for example, the four intervals in $\mathcal{F}_2$ would have probabilities $[(1-p)^2, p(1-p), , p(1-p)$, and $p^2]$ instead of $[1/4, 1/4, 1/4, 1/4]$. This determines a measure on each $\mathcal{F}_n$, which uniquely determine a measure μ_n on $\mathcal{F} = \bigvee_n \mathcal{F}_n$. While $\mu_{0.5}$ is just Lebesgue measure P, the other μ_p's are new.

Now let x be any integer and define two stochastic processes S_n and X_n for $n \ge 0$ by $S_n = \sum_{i=1}^n b_i$ and $X_n = x + \sum_{i=1}^n [2b_i - 1] = x + 2S_n - n$. The sum S_n is called the "partial sums of $\{b_n\}$," and X_n is called a one-dimensional "Random Walk starting at x." If the probability measure is $\mathsf{P} = \mu_{0.5}$ the random walk is called "symmetric," and otherwise asymmetric. As before set $T_j = \inf[n : X_n = j]$, the first hitting time of state j. For some integers $a \le x$ and $b \ge x$, I want to calculate (in several different ways) the quantities

$$\begin{aligned} F_{a,b}(x) &= \Pr[X_n = b \text{ before } X_n = a | X_0 = x] = \Pr[T_a > T_b | X_0 = x] \\ E_{a,b}(x) &= \mathsf{E}[T_a \wedge T_b | X_0 = x] \end{aligned}$$

One way: Difference Equations

Note that $F_{a,b}(a) = 0$ and $F_{a,b}(b) = 1$, while for $a < x < b$, $F_{a,b}(x) = (1-p)F_{a,b}(x-1) + (p)F_{a,b}(x+1)$. Thus $[F_{a,b}(x+1) - F_{a,b}(x)] = [F_{a,b}(x+2) - F_{a,b}(x+1)]\frac{p}{1-p} = c_1\big(\frac{p}{1-p}\big)^{-x}$, or $F(x) = \sum_{a \le i < x}[F_{a,b}(i+1) - F_{a,b}(i)] = c_2\big[(\frac{p}{1-p})^{-a} - (\frac{p}{1-p})^{-x}\big]$. The solution is:

$$F_{a,b}(x) = \begin{cases} \frac{x-a}{b-a} & \text{if } p = 1/2 \\ \frac{\left(\frac{p}{1-p}\right)^{b-x} - \left(\frac{p}{1-p}\right)^{b-a}}{1 - \left(\frac{p}{1-p}\right)^{b-a}} & \text{if } p \ne 1/2 \end{cases}$$

Starting with a fortune of $x = 10$ and playing Roulette, which offers a chance $p = \frac{18}{38} \approx 0.47368$ of winning and $1-p = \frac{20}{38} \approx 0.52632$ of losing (so $\frac{p}{1-p} = \frac{18}{20} = 0.9$), the chance of gaining \$1 before going broke (reaching a fortune of $b = 11$ before falling to $a = 0$) is $\frac{0.9^{11-10} - 0.9^{11-0}}{1-0.9^{11-0}} = \frac{0.9-0.3138}{1.0-0.3138} = 0.8543$, while a "fair" game with $p = 1-p = 1/2$ would give $\frac{10-0}{11-0} = 0.9091$; starting with a fortune of $x = 1000$ and playing until winning \$100 or losing everything, the chances are $\frac{0.9^{100} - 0.9^{1100}}{1-0.9^{1100}} \approx 0.9^{100} \approx 0.00002656$, while again a fair game would yield $F_{a,b}(1000) = 10/11$. Biased random walks don't move very far against the bias.

The expectation of the length of time $T_{a,b} = T_a \wedge T_b$ until hitting a or b can be calculated in the same way. For the symmetric random walk, we have $G_{a,b}(a) = 0 = G_{a,b}(b)$ and

$$\begin{aligned} G_{a,b}(x) &= \mathsf{E}[T_a \wedge T_b | X_0 = x] = 1 + (1-p)G_{a,b}(x-1) + pG_{a,b}(x+1) \\ &= \begin{cases} (x-a)(b-x) & p = 1/2 \\ \frac{x-a}{1-2p} - \frac{b-a}{1-2p}\frac{(\frac{p}{1-p})^{b-x} - (\frac{p}{1-p})^{b-a}}{1-(\frac{p}{1-p})^{b-a}} & p \ne 1/2 \end{cases} \end{aligned}$$

A symmetric random walk starting at 1000 would be ten times more likely to hit 1100 first than 0, and would take on average about 100,000 steps to do either, while a roulette player starting with $1000 and playing until he wins another $100 or loses it all, is almost certain to lose but it will take an average of 18,999.4487 tries. If $p < 1/2$ we can take the limit as $b \to \infty$ above, and see that the probability of eventual "ruin" tends to one and the expected wait until ruin tends to $\mathsf{E}\, T_a = \frac{x-a}{1-2p}$, or $\frac{1000}{1-2*18/38} = 19000$; evidently we are very close to the limit already.

Another way: Martingales

The process $M_n = \big(\frac{p}{1-p}\big)^{-X_n}$ is a Martingale, since

$$\begin{aligned}
\mathsf{E}[M_{n+1}|M_n] &= M_n \times \mathsf{E}\big[\big(\frac{p}{1-p}\big)^{X_n - X_{n+1}}\big] \\
&= M_n \times \big[p\big(\frac{p}{1-p}\big)^{-1} + (1-p)\big(\frac{p}{1-p}\big)^{+1}\big] \\
&= M_n \times \big[(1-p) + p\big] = M_n.
\end{aligned}$$

We will see later from Doob's so-called *optional sampling theorem* that it follows not only that $\mathsf{E} M_t = M_0 = (\frac{p}{1-p})^{-x}$ for fixed times t, but also for certain random times including $t = T_a \wedge T_b$. This implies

$$\begin{aligned}
\big(\frac{p}{1-p}\big)^{-x} = M_0 &= \mathsf{E}\, M_{T_a \wedge T_b} \\
&= \mathsf{P}[T_a < T_b]\mathsf{E}[M_{T_a}|T_a < T_b] + \mathsf{P}[T_a > T_b]\mathsf{E}[M_{T_b}|T_a > T_b] \\
&= [1 - F_{a,b}(x)]\,\big(\frac{p}{1-p}\big)^{-a} + [F_{a,b}(x)]\,\big(\frac{p}{1-p}\big)^{-b}, \\
&= \big(\frac{p}{1-p}\big)^{-a} + [F_{a,b}(x)]\,\Big[\big(\frac{p}{1-p}\big)^{-b} - \big(\frac{p}{1-p}\big)^{-a}\Big],
\end{aligned}$$

leading to a shorter and more direct calculation of $F_{a,b}(x)$ when $p \neq 1/2$; a simpler argument using $M_n = X_n$ gives the result when $p = 1/2$.

The process $M_n = X_n - n[2p-1]$ is also a Martingale, since we have exactly subtracted the expected change in X_n; this implies that

$$\begin{aligned}
x = M_0 &= \mathsf{E}\, M_{T_a \wedge T_b} \\
&= \mathsf{P}[T_a < T_b]\, \mathsf{E}[X_{T_a} - T_a[2p-1]|T_a < T_b] + \mathsf{P}[T_a > T_b]\, \mathsf{E}[X_{T_b} - T_b[2p-1]|T_a > T_b] \\
&= (1 - F_{a,b}(x))\, a + F_{a,b}(x)\, b - [2p-1]\mathsf{E}[T_a \wedge T_b | X_0 = x],
\end{aligned}$$

leading to a shorter and more direct calculation of $G_{a,b}(x)$ when $p \neq 1/2$; a simpler argument using $M_n = (X_n)^2 - n$ gives the result when $p = 1/2$. The art lies in *recognizing* the martingales that will lead to simple solutions; for most processes X_n there are almost mechanical ways of locating "predictable" processes A_n, B_n, and C_n (depending only on $X_1,\ldots,X_{n-1}$) so that the three processes

$$\begin{aligned}
M_n^1 &= X_n - A_n \\
M_n^2 &= (X_n)^2 - B_n \\
M_n^3 &= e^{\alpha X_n - \frac{\alpha^2}{2} C_n}
\end{aligned}$$

are each martingales; for our example we needed M_1 with $A_n = n[2p-1]$, M_2 (for the case $p = 1/2$) with $B_n = n$, and M_3 (with $\alpha = -\log \frac{p}{1-p}$, so that $C_n \equiv 0$). In fact these three almost always will do the trick!