Martingale Methods: Application to SPRT

RANDOM WALKS AND MARTINGALES

Let $\{\xi_j\}$ be independent, identically distributed random variables, all with the same mean $\mu = \mathsf{E}[\xi_j]$, variance $\sigma^2 = \mathsf{V}[\xi_j]$, and moment generating function $M(\lambda) = \mathsf{E}[e^{\lambda \xi_j}]$. Under suitable regularity conditions the logarithm $m(\lambda) \equiv \log M(\lambda)$ has Taylor expansion $m(\lambda) = \mu\lambda + \sigma^2\lambda^2/2 + o(\lambda^2)$ near zero. Let $\mathcal{F}_n \equiv \sigma\{\xi_j : j \le n\}$ be the filtration generated by $\{\xi_j\}$.

For any $x \in \mathbb{R}$ consider the sequence $X_n = x + \sum_{j \le n} \xi_j$ of partial sums, starting at x; X_n is a *random walk* starting at x. Fix real numbers $a < b$ and define a $\{\mathcal{F}_n\}$-stopping time $\tau = \tau_{a,b}$ by

$$\tau \equiv \inf\{n : X_n \notin (a, b)$$

and the "right exit probability" by $\alpha \equiv \mathsf{P}[\tau < \infty \text{ and } X_\tau \ge b]$. Our object is to compute α and $\mathsf{E}[\tau]$, the probability of exiting on the right and the expected exit time, as functions of $x \in (a, b)$.

The Symmetric Case

First suppose $\mu = 0$. Then X_n is a martingale, and so (by the optional sampling theorem) is $X_{\tau \wedge n}$, which moreover is bounded and hence uniformly integrable. It follows that

$$\begin{aligned} x &= \mathsf{E}[X_{\tau\wedge 0}] \\ &= \lim_{n\to\infty} \mathsf{E}[X_{\tau\wedge n}] \\ &= E[X_\tau] \\ &\approx a(1-\alpha) + b(\alpha), \\ \alpha &\approx \frac{x-a}{b-a} \end{aligned}$$

(the estimates are exact if $\mathsf{P}[X_\tau \in \{a, b\}] = 1$, but only approximate if there is a chance of "overshooting" the boundary). Also $(X_n)^2 - n\sigma^2$ is a martingale, so

$$\begin{aligned} x^2 &= \mathsf{E}[(X_{\tau\wedge 0})^2 - (\tau \wedge 0)\sigma^2] \\ &= E[(X_\tau)^2 - \tau\sigma^2] \\ &\approx a^2(1-\alpha) + b^2(\alpha) - \mathsf{E}[\tau]\sigma^2, \\ \mathsf{E}[\tau] &\approx \frac{a^2 + \alpha(b^2 - a^2) - x^2}{\sigma^2} \\ &\approx (b-x)(x-a)/\sigma^2. \end{aligned}$$

For example, for the standard symmetric random walk on the integers with $\xi = \pm 1$ with probability 1/2 each, then $\mu = 0$, $\sigma^2 = 1$, and for $a < x < b$, $\alpha = \mathsf{P}[X_n \ge b \text{ before } X_n \le a | X_0 = x] = (x-a)/(b-a)$ and the exit time $\tau \equiv \min[n : X_n \notin (a, b)]$ has expectation $\mathsf{E}[\tau | X_0 = x] = (b-x)(x-a)$.

The Asymmetric Case

Now suppose $\mu \ne 0$, that $\mathsf{P}[\xi_j < 0] > 0$ and $\mathsf{P}[\xi_j > 0] > 0$, and that $M(t)$ is smooth enough for the logarithm $m(\lambda) \equiv \log M(\lambda)$ to have Taylor expansion $m(\lambda) = \mu\lambda + \sigma^2\lambda^2/2 + o(\lambda^2)$ near zero. Then $m(\lambda) \to \infty$ as $\lambda \to \pm\infty$ while $m'(0) = \mu \ne 0$, so there exists some $\lambda^* \ne 0$ (approximately $\lambda^* \approx -2\mu/\sigma^2$) for which $m(\lambda^*) = 0$. For any λ, $Y_n \equiv e^{\lambda X_n - n\,m(\lambda)}$ is a martingale (well, any λ for which $M(\lambda) < \infty$) and, in particular, $e^{\lambda^* X_n}$ is a martingale, so again the optional sampling theorem gives

$$\begin{aligned} e^{\lambda^* x} &= \mathsf{E}[e^{\lambda^* X_{\tau\wedge n}}] \\ &= \mathsf{E}[e^{\lambda^* X_\tau}] \\ &\approx e^{\lambda^* a}(1-\alpha) + e^{\lambda^* b}(\alpha), \\ \alpha &\approx \frac{e^{\lambda^* x} - e^{\lambda^* a}}{e^{\lambda^* b} - e^{\lambda^* a}} \end{aligned}$$

To find $P = \mathsf{P}[X_t \textit{ ever} \text{ exceeds } b]$, take $a \to -\infty$ to find $P \approx e^{-\lambda^*(b-x)} < 1$, if $\lambda^* > 0$, or $P = 1$, if $\lambda^* \le 0$.

Since $(X_n - n\mu)$ is also a martingale,

$$\begin{aligned} x &= \mathsf{E}[X_{\tau\wedge n} - (\tau \wedge n)\mu] \\ &= \mathsf{E}[X_\tau - \tau\mu] \\ &\approx a(1-\alpha) + b(\alpha) - \mathsf{E}[\tau]\mu, \\ \mathsf{E}[\tau] &\approx \frac{a - x + \alpha(b-a)}{\mu}. \end{aligned}$$

For example, for the standard asymmetric random walk on the integers with $\xi = \pm 1$ with probabilities p, $q = 1-p$, respectively, then $\mu = q-p$, $\sigma^2 = 4pq$, and $M(\lambda) = pe^{\lambda} + qe^{-\lambda} = 1 + (pe^{\lambda} - q)(1 - e^{-\lambda})$ so $m(\lambda^*) = 0$ for $\lambda^* = \log q/p$. Thus for $a < x < b$, $\alpha = \mathsf{P}[X_n \ge b \text{ before } X_n \le a | X_0 = x] = ((q/p)^x - (q/p)^a)/((q/p)^b - (q/p)^a)$ and the exit time $\tau \equiv \min[n : X_n \notin (a,b)]$ has expectation $\mathsf{E}[\tau | X_0 = x] = (a - x + \alpha(b-a))/\mu$. If $\alpha \approx 0$ (resp., $\alpha \approx 1$) this is close to $\mathsf{E}[\tau] \approx (x-a)/\mu$ (resp., $\mathsf{E}[\tau] \approx (b-x)/\mu$), just what you would expect for a heavily biased random walk.

Sequential Probability Ratio Test

Let $\{Y_j\}$ be independent, identically-distributed random variables with absolutely-continuous distributions and density function $f(y)$, and consider the statistical problem of trying to tell from observed values $y_1, ..., y_n$ which of two possible density functions $\{f_0, f_1\}$ governs the distribution of the $\{Y_j\}$. All of the standard statistical tests of the hypothesis $H_0 : f = f_0$ against its alternative $H_1 : f = f_1$ make use of the Likelihood Ratio

$$L_n \equiv \frac{f_1(y_1) \cdots f_1(y_n)}{f_0(y_1) \cdots f_0(y_n)} = \prod_{j \le n} \frac{f_1(y_j)}{f_0(y_j)}$$

or, equivalently, its logarithm $\ell_n = \sum_{j \le n} \log(f_1(y_j)/f_0(y_j))$. For example, the Bayesian posterior probability of H_0 (starting with prior $\pi_0 = \mathsf{P}[H_0]$, $\pi_1 = 1 - \pi_0 = \mathsf{P}[H_1]$) is given by

$$\begin{aligned} \mathsf{P}[H_0 | Y_1 ... Y_n] &= \frac{\pi_0 f_0(y_1) \cdots f_0(y_n)}{\pi_0 f_0(y_1) \cdots f_0(y_n) + \pi_1 f_1(y_1) \cdots f_1(y_n)} = \frac{(\pi_0/\pi_1)}{(\pi_0/\pi_1) + L_n} \\ \frac{\mathsf{P}[H_1 | Y_1 ... Y_n]}{\mathsf{P}[H_0 | Y_1 ... Y_n]} &= \frac{\pi_1}{\pi_0} L_n, \end{aligned}$$

so the posterior odds in favor of H_1 are the prior odds multiplied by L_n (which in this context is called a "Bayes factor"), while the Neyman-Pearson Lemma says that the most powerful (frequentist) test of level α is to reject H_0 whenever $L_n \ge r$, where r is chosen to ensure that the probability of a "type-I error" (rejecting a true null hypothesis) is no more than $\mathsf{P}_0[L_n \ge r] \le \alpha$ if $H_0 : f = f_0$ is true (the subscript zero on P_0 indicates that this probability should be computed assuming H_0); the "power" of the test is then $\mathsf{P}_1[L_n \ge r]$, the probability of rejecting H_0 when in fact H_1 is true, or one minus the probability $\beta = \mathsf{P}_1[L_n < r]$ of a "type-II error", failing to reject a false null hypothesis.

Only a large sample-size n will ensure that the probability of either of these errors will be small, but *how* large n must be will depend on how different f_0 and f_1 are, something that may be difficult to anticipate. One possibility, initially proposed by Abraham Wald, is to design a **sequential** test in which we draw data successively until the evidence becomes compelling either that H_0 is false and must be rejected (large values of L_n) or that H_0 is true and must not be rejected (small values of L_n). A simple process is to select numbers $0 < A < 1 < B$ and continue sampling until either $L_n \ge B$, in which case we stop and reject H_0, or $L_n \le A$, in which case we stop sampling and accept H_0.

But this is exactly equivalent to drawing samples until the **random walk** $\ell_n \equiv \log L_n$, which starts at $x = \log 1 = 0$, reaches either a lower boundary $a = \log A < 0$ or an upper boundary $b = \log B > 0$, a problem we have just addressed. For this random walk we have under hypothesis H_0 that

$$\begin{aligned}\mu_0 &= \mathsf{E}_0[\xi_j = \log(f_1(Y_j)/f_0(Y_j))] \\ &= \int \log\frac{f_1(y)}{f_0(y)}\, f_0(y)\, dy = -K(f_0, f_1) \\ M_0(\lambda) &= \mathsf{E}_0[e^{\lambda\xi_j} = (f_1(Y_j)/f_0(Y_j))^\lambda] \\ &= \int f_1(y)^\lambda\, f_0(y)^{1-\lambda}\, dy\end{aligned}$$

and under hypothesis H_1 that

$$\begin{aligned}\mu_1 &= \mathsf{E}_1[\xi_j = \log(f_1(Y_j)/f_0(Y_j))] \\ &= \int \log\frac{f_1(y)}{f_0(y)}\, f_1(y)\, dy = K(f_1, f_0) \\ M_1(\lambda) &= \mathsf{E}_1[e^{\lambda\xi_j} = (f_1(Y_j)/f_0(Y_j))^\lambda] \\ &= \int f_1(y)^{1+\lambda}\, f_0(y)^{-\lambda}\, dy\end{aligned}$$

so $\mu_0 < 0$ and $\mu_1 > 0$ ($K(f,g) > 0$ is the Kullback-Leibler divergence from f to g, a measure of how different f and g are; for example, the K-L divergence from a standard normal distribution to the $N(\mu, 1)$ distribution is $\mu^2/2$). Also $M_0(\lambda^* = 1) = 1$ and $M_1(\lambda^* = -1) = 1$, so the exit time $\tau \equiv \min[n : \ell_n \notin (a, b)]$ leads to right-exit probability (and Type-I error probability) $\alpha = \mathsf{P}_0[\ell_\tau \geq b]$ and to left-exit probability (and Type-II error probability) $\beta = \mathsf{P}_1[\ell_\tau \leq a]$ of

$$\begin{aligned}\alpha = \mathsf{P}_0[\ell_\tau \geq b] &\approx \frac{e^0 - e^a}{e^b - e^a} = \frac{1 - A}{B - A} \\ \beta = \mathsf{P}_1[\ell_\tau \leq a] &\approx 1 - \frac{e^{-0} - e^{-a}}{e^{-b} - e^{-a}} = \frac{AB - A}{B - A}\end{aligned}$$

with expected sample-size

$$\begin{aligned}\mathsf{E}_0[\tau] &\approx -\frac{(B-1)\log A + (1-A)\log B}{(B-A)\, K(f_0, f_1)} \\ \mathsf{E}_1[\tau] &\approx \frac{A(B-1)\log A + B(1-A)\log B}{(B-A)\, K(f_1, f_0)}\end{aligned}$$

In the symmetric case $AB = 1$, $\alpha = \beta = 1/(1 + B)$, and

$$\mathsf{E}_0[\tau] = \frac{(B-1)\log B}{(B+1)K(f_0, f_1)} \qquad \mathsf{E}_1[\tau] = \frac{(B-1)\log B}{(B+1)K(f_1, f_0)}.$$

Evidently α and β may be made as small as desired by taking $B = \log\frac{1-\alpha}{\alpha}$ sufficiently large, but doing so will increase the expected sample size to approximately $\mathsf{E}[\tau] \approx -\log\alpha/K(f_0, f_1)$.

Exercise 1: Suppose that *both* hypotheses are wrong, and that $\xi_j \sim f(x)\,dx$ but $f \neq \{f_0, f_1\}$. Show that ℓ_n is again a random walk, now with drift $\mu = K(f, f_1) - K(f, f_0)$ and conclude that, almost surely, $L_n \to 0$ as $n \to \infty$ if $K(f, f_0) < K(f, f_1)$ and $L_n \to \infty$ as $n \to \infty$ if $K(f, f_0) > K(f, f_1)$. What do you think would happen if the two divergences are equal?

Exercise 2a: Find $K(f_0, f_1)$ if each f_i is $\mathsf{No}(\mu_i, \sigma^2)$ (different means, same variance).

Exercise 2b: Find $K(f_0, f_1)$ if each f_i is $\mathsf{No}(0, \sigma_i^2)$ (same mean, different variances).

Exercise 2c: Find $K(f_0, f_1)$ if each f_i is exponentially distributed with possibly-different means λ_i.

Exercise 2d: Find $K(f_0, f_1)$ if each f_i is binomially distributed with the same N but possibly different probabilities p_i.