STA 711: Probability & Measure Theory
Robert L. Wolpert

12 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 $M(\lambda) = \mathsf{E}[e^{\lambda\xi_j}]$ finite in an interval around $\lambda = 0$. Under suitable regularity conditions the logarithm $m(\lambda) := \log M(\lambda)$ has Taylor expansion $m(\lambda) = \mu\lambda + \sigma^2\lambda^2/2 + o(\lambda^2)$ near zero. Let $\mathcal{F}_n := \sigma\{\xi_j : j \le n\}$ 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

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

and the "right exit probability" by $\alpha := \mathsf{P}[\tau < \infty \text{ and } X_\tau \ge b]$. 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 $x \in (a, b)$. This classical probability problem is called the "gambler's ruin problem", for the example in which x is a gambler's initial fortune and ξ_j the net return on the jth independent play of some game of chance; in that case α is the probability that the gambler attains a fortune of $X_n \ge b$ before going broke (for $a = 0$). Below we will present an approach to sequential statistical significance testing based on the same probability model.

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

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

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

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

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 $a < x < b$, $\alpha = \mathsf{P}[X_n \ge b \text{ before } X_n \le a \mid X_0 = x] = (x-a)/(b-a)$ and the exit time $\tau := \min[n : X_n \notin (a,b)]$ has expectation $\mathsf{E}[\tau \mid X_0 = x] = (b-x)(x-a)$.

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 $m(\lambda) := \log M(\lambda)$ 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 $m(0) = 0$ and $m'(0) = \mu \ne 0$, so there exists some $\lambda^* \ne 0$ (approximately $\lambda^* \approx -2\mu/\sigma^2$) for which $m(\lambda^*) = 0$. For any $\lambda \in \mathbb{R}$, $Y_n := e^{\lambda X_n - n\,m(\lambda)}$ is a martingale (well, any λ for which $M(\lambda) < \infty$). In particular, $e^{\lambda^* X_n}$ is a martingale, so again the optional sampling theorem and UI limit theorem give

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

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,

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

For example, for the standard asymmetric random walk on the integers with $\xi = \pm 1$ with probabilities p, $q = 1-p$, respectively, then $\mu = p-q$, $\sigma^2 = 4pq$, 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 $a < x < b$, $\alpha = \mathsf{P}[X_n \ge b \text{ before } X_n \le a \mid X_0 = x] = ((q/p)^x - (q/p)^a)/((q/p)^b - (q/p)^a)$ and the exit time $\tau := \min[n : X_n \notin (a,b)]$ has expectation $\mathsf{E}[\tau \mid X_0 = x] = (a - x + \alpha(b-a))/\mu$. 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 (against the null)

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

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

$$\mathsf{P}[H_0 \mid Y_1...Y_n] = \frac{\pi_0 f_0(y_1)\cdots f_0(y_n)}{\pi_0 f_0(y_1)\cdots f_0(y_n) + \pi_1 f_1(y_1)\cdots f_1(y_n)} = \frac{(\pi_0/\pi_1)}{(\pi_0/\pi_1)+\Lambda_n}$$
$$\frac{\mathsf{P}[H_1 \mid Y_1...Y_n]}{\mathsf{P}[H_0 \mid Y_1...Y_n]} = \frac{\pi_1}{\pi_0}\Lambda_n,$$

so the posterior odds against H_0 are the prior odds multiplied by Λ_n (which in this context is called the "Bayes factor against H_0"), while the Neyman-Pearson Lemma says that the most powerful (frequentist) test of level α is to reject H_0 whenever $\Lambda_n \ge r$, where r is chosen to ensure that the probability of a "Type-I error" (rejecting a true null hypothesis) is no more than some specified value $\mathsf{P}_0[\Lambda_n \ge r] \le \alpha$ 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 $\mathsf{P}_1[\Lambda_n \ge r]$, the probability of rejecting H_0 when in fact H_1 is true, or one minus the probability $\beta = \mathsf{P}_1[\Lambda_n < r]$ of a "Type-II error", failing to reject a false null hypothesis.

Only a large sample-size n will ensure that both of these error probabilities 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 Wald and Wolfowitz (1948), is to design a **sequential** test in which data are drawn successively until the evidence becomes compelling either that H_0 is false and must be rejected (large values of Λ_n) or that H_0 is true and must not be rejected (small values of Λ_n). A simple process is to select numbers $0 < A < 1 < B$ and continue sampling until either $\Lambda_n \ge B$, in which case we stop and reject H_0, or $\Lambda_n \le A$, in which case we stop sampling and accept H_0.

But this is exactly equivalent to drawing samples until the **random walk** $\ell_n := \log \Lambda_n$, 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 solved. For this random walk and any $\lambda \in \mathbb{R}$, under the hypothesis H_0, the means of $\xi_j := \log\big(f_1(Y_j)/f_0(Y_j)\big)$ and of $\exp(\lambda\xi_j)$ are given by:

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

and under hypothesis H_1 they are

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

so $\mu_0 < 0 < \mu_1$. The quantity $K(f : g) \ge 0$ 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 $\mathsf{No}(\mu, 1)$ distribution is $\mu^2/2$. Also $M_0(\lambda^* := 1) = 1 = M_1(\lambda^* := -1)$, so the exit time $\tau := \min[n \ge 0 : \ell_n \notin (a, b)]$ leads to right-exit probability (and Type-I error probability)

$\alpha = \mathsf{P}_0[\ell_\tau \ge b]$ and to left-exit probability (and Type-II error probability) $\beta = \mathsf{P}_1[\ell_\tau \le a]$ of

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

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 $AB = 1$, $\alpha = \beta = 1/(1+B) = A/(1+A)$, 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 $A^{-1} = B = \log\frac{1-\alpha}{\alpha}$ sufficiently large, but doing so will increase the expected sample size to approximately $\mathsf{E}[\tau] \approx \log\frac{1}{\alpha}/K(f_0:f_1)$.

Exercise 1a: Starting with $X_0 = \$80$ and betting \$1 each turn at even odds, what is the chance of reaching $b = \$100$ before going broke (*i.e.*, reaching $a = \$0$)? On average, how long will it take to reach one of these?

Exercise 1b: Same question, but now playing US Roulette with probability $p = 9/19$ of winning and $q = 10/19$ of losing each turn.

Exercise 2: (Berk, 1966). Suppose that *both* hypotheses are wrong, and that $\xi_j \sim f(x)\,dx$ but $f \ne \{f_0, f_1\}$. 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, $\Lambda_n \to 0$ as $n \to \infty$ if $K(f:f_0) < K(f:f_1)$ and $\Lambda_n \to \infty$ as $n \to \infty$ if $K(f:f_0) > K(f:f_1)$. What do you think would happen if $K(f:f_0) = K(f:f_1)$?

Exercise 3a: Find $K(f_0:f_1)$ if each f_i is $\mathsf{No}(\mu_i, \sigma^2)$ (different means, same variance).

Exercise 3b: Find $K(f_0:f_1)$ if each f_i is $\mathsf{No}(0, \sigma_i^2)$ (same mean, different variances).

Exercise 3c: Find $K(f_0:f_1)$ if each f_i is $\mathsf{Ex}(\lambda_i)$, exponentially distributed with rate λ_i.

Exercise 3d: Find $K(f_0:f_1)$ if each f_i is $\mathsf{Bi}(N, p_i)$, binomial with the same N but possibly different probabilities p_i.

References

Berk, R. H. (1966), "Limiting Behavior of Posterior Distributions when the Model is Incorrect," *Annals of Mathematical Statistics*, 37, 51–58, doi:10.1214/aoms/1177699597.

Wald, A. and Wolfowitz, J. (1948), "Optimal Character of the Sequential Probability Ratio Test," *Annals of Mathematical Statistics*, 19, 326–339, doi:10.1214/aoms/1177730197.

Last edited: July 6, 2017