March 21, 2003

1.5 The sequential probability ratio test. Sec. 1.1 treated tests between two laws P and Q on a sample space X. If we have independent, identically distributed (i.i.d.) observations $X_1, \ldots, X_n$ with distribution P or Q, then X can be replaced by the Cartesian product X^n of n copies of X on which one decides between two laws P^n and Q^n, so that we are back in the situation of deciding between two laws on a sample space. But now, decision procedures will be considered where the statistician, instead of having to decide between P and Q for a fixed n, is allowed to gather additional observations. An infinite amount of data would allow a correct decision to be made between any two different probability measures with zero error probabilities, but now it will be assumed that each observation has a *cost* c, measured on the same scale as loss functions. So if $c > \max(L_{PQ}, L_{QP})$, it would not be worth taking any observations; for a prior π one should choose P if $\pi(Q)L_{QP} < \pi(P)L_{PQ}$, choose Q if $\pi(Q)L_{QP} > \pi(P)L_{PQ}$, and make an arbitrary choice if $\pi(Q)L_{QP} = \pi(P)L_{PQ}$. More typically, the cost c per observation is rather small relative to L_{PQ} and L_{PQ}, so it will be worth taking some observations. A decision rule will choose among three options after the nth observation: decide in favor of P or of Q (and stop taking observations), or take at least one more observation. Note that if $nc > \max(L_{PQ}, L_{QP})$, more would be spent on n observations than could be lost by making a wrong decision, so for a good strategy, the probability of taking n or more observations should not be too high. Still, if one has already made $n-1$ observations, the $(n-1)c$ spent on them is already lost and it may well be that the next observation, with cost c, is worth taking.

We will have a sequence of possible observations $X_1, X_2, \ldots$, i.i.d. with distribution $\mu = P$ or Q. Let f be the likelihood ratio $R_{Q/P}$ on X. Then $0 \leq f \leq +\infty$. After n observations the likelihood ratio of Q^n to P^n is

$$R_{Q^n/P^n}(X_1, \ldots, X_n) \;:=\; r_n(X_1, \ldots, X_n) \;:=\; \textstyle\prod_{j=1}^n f(X_j)$$

for $n = 1, 2, \ldots$. Let $r_0 \equiv 1$.

The probability that r_n is an undefined product $0 \cdot \infty$ will be 0 under either P^n or Q^n, since $f = 0$ has probability 0 for Q and $f = \infty$ has probability 0 for P. For a fixed n, the Neyman-Pearson lemma (Sec. 1.1) tells us to choose P if r_n is small and Q if r_n is large. In the sequential probability ratio test (SPRT) to be defined, the idea is that if r_n is in an intermediate range, we make no decision and continue sampling. Specifically, for $0 \leq A \leq B \leq \infty$, the SPRT$(A, B)$ is the decision rule which calls for taking observations $X_1, \ldots, X_n$ until the least n such that $r_n \leq A$ or $r_n \geq B$, then to choose P if $r_n \leq A$ or Q if $r_n \geq B$ (and $r_n > A$, which will be true automatically except in the unusual case that $r_n = A = B$). We will usually have $0 < A < 1 < B < \infty$. If $A \geq 1$ the test chooses P, or for $B \leq 1$ and $A < 1$ it chooses Q, for $n = 0$, that is, without taking any observations.

In general, a *sequential test* (non-randomized) will be a sequence of measurable functions $\{\phi_n(X_1, \ldots, X_n)\}_{n \geq 0}$. Each ϕ_n has possible values $-1, 0$ and 1, where we stop sampling at N, the least n such that $\phi_n \neq 0$, and then choose P if $\phi_n = -1$ and Q if $\phi_n = 1$. Here N is random. We would like to choose a test so as to minimize the expectations of N under P or Q as well as the error probabilities. The total cost + loss will be Nc plus the loss L_{PQ} or L_{QP}, if any.

For any sequential test $\phi = \{\phi_n\}$ of P vs. Q, let $\alpha(\mu, \phi)$ be the probability that ϕ

rejects μ when it is true, and let $E_{\mu,\phi}N$ be the expectation of N if μ is true, both for $\mu = P$ or Q. Sequential probability ratio tests have an optimality property among all sequential tests, as follows:

1.5.1 Theorem (A. Wald, J. Wolfowitz, T. Ferguson, D. Burkholder, R. Wijsman). Let $\psi = \mathrm{SPRT}(A, B)$ where $0 < A < 1 < B < \infty$ and let ϕ be any other sequential test. Suppose that for two laws P and Q, $\alpha(P, \phi) \le \alpha(P, \psi)$ and $\alpha(Q, \phi) \le \alpha(Q, \psi)$. Then $E_{P,\psi}N \le E_{P,\phi}N$ and $E_{Q,\psi}N \le E_{Q,\phi}N$.

In other words, if ψ is a sequential probability ratio test with $\alpha(P, \psi) = \alpha$ and $\alpha(Q, \psi) = \gamma$, then in the class of all sequential tests ϕ with $\alpha(P, \phi) \le \alpha$ and $\alpha(Q, \phi) \le \gamma$, the SPRT ψ minimizes $E_{P,\phi}N$, *and* it minimizes $E_{Q,\phi}N$. This optimality property is unexpectedly strong, since one might have thought that by allowing $E_{P,\phi}N$ to be larger one could make $E_{Q,\phi}N$ smaller, or vice versa.

Theorem 1.5.1 will be proved in Sec. 1.7. First, here are some other facts.

1.5.2 Lemma. For any $\psi = \mathrm{SPRT}(A, B)$ of P vs. Q with $0 < A \le B < \infty$, there is a δ with $0 < \delta < 1$ and a $C < \infty$ such that $\mu^n(N > n) \le C(1 - \delta)^n$ for $\mu = P$ or Q.

Proof. If $A = B$ then $N \equiv 0$ and there is no problem. So assume $A < B$. Let $Z_i := \log f(X_i)$. Then Z_i are independent, identically distributed random variables with $-\infty \le Z_i \le +\infty$. Let $S_n := Z_1 + \cdots + Z_n$, which is well-defined almost surely for P^n or Q^n. Then $N > n$ if and only if $\log A < S_j < \log B$ for $j = 1, \dots, n$. Since $P \ne Q$, there is a $b > 0$ such that $p := \mu(|Z_1| \ge b)/2 > 0$ for $\mu = P$ or Q. So for some m large enough, there is probability at least p^m that $|S_m| > \log(B/A)$. The events $D_i := \{|S_{im} - S_{(i-1)m}| > \log(B/A)\}$ are independent for $i = 1, 2, \dots$, and if D_i occurs then $N \le im$. So the probability that $N > im$ is $\le (1 - p^m)^i$ for $\mu = P$ or Q. Thus for all $j = 1, 2, \dots$, $\Pr(N \ge j) \le (1 - p^m)^{[j/m]}$, where $[x]$ is the largest integer $\le x$ and $\Pr = \mu^n$, $\mu = P$ or Q. Thus $\Pr(N > j) \le (1 - p^m)^{j/m - 1}$, so let $\delta := 1 - (1 - p^m)^{1/m}$ and $C := 1/(1 - p^m)$. □

It follows from Lemma 1.5.2 that $E_{P,\psi}N < \infty$ and $E_{Q,\psi}N < \infty$. To apply sequential probability ratio tests it is useful to know relations between A, B and the error probabilities $\alpha(P, \psi)$ and $\alpha(Q, \psi)$ such as the following.

1.5.3 Proposition. For the test $\psi = \mathrm{SPRT}(A, B)$ of P vs. Q, with $0 < A < 1 < B < +\infty$, let $\alpha_0 := \alpha(P, \psi)$ and $\alpha_1 := \alpha(Q, \psi)$. Then

$$\alpha_0 \ \le\ (1 - \alpha_1)/B \ \le\ 1/B \ \text{ and } \ \alpha_1 \ \le\ (1 - \alpha_0)A \ \le\ A.$$

Proof. Let F_n be the event that $N = n$ and $r_n \ge B$. Then the events F_n are disjoint and

$$B\alpha(P, \psi) \ =\ B\sum_{n=1}^{\infty} P^n(F_n) \ \le\ \sum_{n=1}^{\infty} Q^n(F_n) \ =\ 1 - \alpha(Q, \psi).$$

The other inequality is proved symmetrically, since $\mathrm{SPRT}(A, B)$ for P vs. Q is equivalent to $\mathrm{SPRT}(1/B, 1/A)$ for Q vs. P. □

Examples. Here is a class of examples to show that of the four inequalities in Proposition 1.5.3, the first and third are sharp (they become equalities in special cases). Of course, the second inequality is an equation only for $\alpha_1 = 0$ and the fourth for $\alpha_0 = 0$. Zero error probabilities are possible but unusual.

There are P and Q such that $r_1 := R_{Q/P}$ takes only two values t and $1/t$, where $1 < t < \infty$. Specifically, if $p := P(r_1 = t)$, then $1 = Q(X) = pt + (1-p)/t$ so $p = 1/(t+1)$.

Let $A = 1/t^j$ and $B = t^k$ for some positive integers j and k. For all n, clearly $r_n = t^i$ for some $i \in \mathbb{Z} := \{0, \pm 1, \pm 2, \ldots\}$. If $r_n = t^i$ then $r_{n+1} = t^{i\pm 1}$. So if $r_N \geq B$ then $r_N = B$ (recalling the definition of N) and if $r_N \leq A$ then $r_N = A$. So the only inequality in the proof of Proposition 1.5.3 for B becomes an equation, and likewise for A. So $\alpha_0 = (1-\alpha_1)/B$ and $\alpha_1 = (1-\alpha_0)A$. The two linear equations can then be solved for α_0 and α_1, giving $\alpha_0 = (1-A)/(B-A)$ and $\alpha_1 = (B-1)A/(B-A)$.

In general, however, it can easily happen that $r_N > B$ ("overshoot") or $r_N < A$ ("undershoot"). If r_1 is bounded above by K, then $r_N \leq KB$, and if r_1 is bounded below by ε, then $r_N \geq \varepsilon A$.

Proposition 1.5.3 gave bounds for error probabilities for SPRTs, which became equations in the last example. In that example, it is also possible to find explicitly the average sample numbers $E_{\mu,\psi}N$ for $\mu = P$ and Q.

Definition. A random variable τ which is a function of $X_1, X_2, \ldots$ is a *stopping time* if τ has nonnegative integer values and for all $n = 1, 2, \ldots$ there is an event A_n such that $\tau \leq n$ if and only if $(X_1, \ldots, X_n) \in A_n$, while for $n = 0$, $\{\tau = 0\}$ is either empty or the whole space.

For example, if $\tau = N$ for an SPRT or any sequential test ϕ, $N \leq n$ if and only if for some $j \leq n$, $\phi_j(X_1, \ldots, X_j) \neq 0$, so N is a stopping time. Recall that "i.i.d." means "independent and identically distributed."

1.5.4 Wald's identity. If $Y_1, Y_2, \ldots$ are i.i.d., $E|Y_1| < \infty$, $T_n := Y_1 + \cdots + Y_n$ for $n \geq 1$, $T_0 := 0$, and $E\tau < \infty$, then $ET_\tau = E\tau EY_1$.

Proof. The identity is clear if $\tau \equiv 0$, so we can assume $\tau \geq 1$ almost surely. A mathematical note: the following proof will apply first in the case where $Y_1 \geq 0$ a.s., replacing Y_i by $|Y_i|$ for all i, which will justify interchanging sums, and sums with expectations, then for general Y_i. We have

$$ET_\tau = \sum_{n=1}^{\infty} P(\tau = n) E(Y_1 + \cdots + Y_n | \tau = n) = \sum_{n=1}^{\infty} P(\tau = n) \sum_{j=1}^{n} E(Y_j | \tau = n)$$

where the conditional expectation is replaced by 0 if $P(\tau = n) = 0$. Note: if Y_j were independent of $\{\tau = n\}$ for $j \leq n$ then $E(Y_j | \tau = n) = EY_j$ and Wald's identity would follow. But this independence does not hold in general. Instead,

$$ET_\tau = \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} E(Y_j | \tau = n) P(\tau = n) = \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} E(Y_j 1_{\tau=n})$$

$$= \sum_{j=1}^{\infty} E(Y_j 1_{\tau \geq j}) \;=\; \sum_{j=1}^{\infty} E(Y_j(1 - 1_{\tau \leq j-1})).$$

The event $\{\tau \leq j-1\}$ *is* independent of Y_j, so

$$ET_\tau \;=\; \sum_{j=1}^{\infty} E(Y_j)P(\tau \geq j) \;=\; EY_1 \sum_{j=1}^{\infty} P(\tau \geq j).$$

Now

$$\sum_{j=1}^{\infty} P(\tau \geq j) \;=\; \sum_{j=1}^{\infty}\sum_{k\geq j} P(\tau = k) \;=\; \sum_{k=1}^{\infty} P(\tau = k)\sum_{j=1}^{k} 1 \;=\; \sum_{k=1}^{\infty} kP(\tau = k) \;=\; E\tau$$

and the identity follows. □

Recall the examples just after Prop. 1.5.3, where $r_1 = f = R_{Q/P}$ only has values $1/t$, 1 or t for some $t > 1$, and we do a test $\psi = \text{SPRT}(A, B)$ for $A = 1/t^j$ and $B = t^k$ for some positive integers j, k. Then r_N has only two possible values A, B. Take $Y_i := \log f(X_i)$, which has possible values 0 or $\pm \log t$, and T_N has possible values $\log A$ or $\log B$. Then $P^\infty(r_N = B) = \alpha(P, \psi) = \alpha_0$ and $Q^\infty(r_N = A) = \alpha(Q, \psi) = \alpha_1$. Recall that α_0 and α_1 can be found explicitly in this case in terms of A, B. Thus we can find $E_\mu T_\tau$ and by Wald's identity, we can evaluate $E_{\mu,\psi} N = E_\mu T_\tau / E_\mu Y_1$ where E_μ is the expectation when μ is the true probability law and $\mu = P$ or Q.

PROBLEMS

1. Let $X = \{0, 1, 2\}$, $P(0) = Q(2) = 1/3$, $P(1) = Q(1) = 1/2$, and $P(2) = Q(0) = 1/6$. For $\psi = \text{SPRT}(1/4, 4)$ of P vs. Q, find upper bounds (as small as possible) for the error probabilities α_0 and α_1.
2. For ψ in Problem 1, evaluate the average sample numbers $E_{P,\psi}N$ and $E_{Q,\psi}N$. Hint: use Wald's identity, applied to $Y_i = \log R_{Q/P}(X_i)$.
3. Suppose that for some $K > 1$, $R_{Q/P}(x) \leq K$ for all x. Show that on the event F_n in the proof of Proposition 1.5.3, $B \leq r_N < KB$. Then show that $\alpha_0 \geq (1 - \alpha_1)/(KB)$. Similarly if $R_{Q/P}(x) \geq 1/M$ for all x, show that $\alpha_1 \geq (1 - \alpha_0)A/M$.
4. Pairs of patients participate in a trial of a blood pressure drug. One of each pair is chosen at random to get the drug while the other gets a placebo which has no effect. Later, the patients' blood pressures are measured. For the ith pair, let X_i be the measurement [average of systolic and diastolic] for the patient getting the drug minus that of the other member of the pair. Hypothesis P is the "null" hypothesis that the drug is ineffective, so $P(X_i < 0) = 1/2$. Hypothesis Q has $Q(X_i < 0) = 0.6$. Let $Y_i = 1$ if $X_i < 0$ and $Y_i = 0$ otherwise. Based on the Y_i data, we want to find a test $\psi = \text{SPRT}(1/B, B)$ of P vs. Q such that the error probabilities α_0 and α_1 are both ≤ 0.05.
   (a) Give a B_1 such that $B \geq B_1$ is sufficient, using $\alpha_0 \leq 1/B$ and $\alpha_1 \leq A$ in Proposition 1.5.3.
   (b) Give a B_0 such that $B \geq B_0$ is necessary, using the previous problem.

5. In the example at the end of the section with $t = 2$, and $R_{Q/P} = t$ or $1/t$ (not 1),
   (a) Find the smallest sample size n for which a (possibly randomized but non-sequential) test for n observations given by the Neyman-Pearson lemma (Sec. 1.1) has both error probabilities ≤ 0.05 (in other words, size ≤ 0.05 and power ≥ 0.95). Hint: if $b(k, n, p)$ is the probability of exactly k successes in n independent trials with probability p of success on each trial, and $E(k, n, p)$ the probability of k or more successes, then $E(12, 23, 1/3) \doteq 0.04805$ while $E(12, 22, 1/3) + \frac{1}{2}b(11, 22, 1/3) \doteq 0.05572$.
   (b) Find an SPRT(A, B) for the same P and Q with both error probabilities ≤ 0.05.
   (c) To evaluate the relative efficiency of the SPRT, find EN/n for P and for Q with n from (a) and N from (b), or give an upper bound for it as small as possible (we hope, less than 1).
6. In problem 2 of Sec. 1.1, find an SPRT(A, B) for P vs. Q having both error probabilities ≤ 0.05, with B as small as possible, $B > 1$, for which the inequality in Proposition 1.5.3 becomes an equality. Then, find the actual error probabilities (which may be less than 0.05).
7. This relates to Wald's identity 1.5.4 and its proof. Let $Y_1, Y_2, \ldots$ be independent, identically distributed random variables with $P(Y_1 = -1) = P(Y_1 = 3) = 1/2$. Let τ be the least j such that $|Y_j| > 2$. Which of the following equations are valid? Show in each case what the left and right sides do equal.
   (a) $E(Y_1|\tau = 1) = EY_1$, (b) $E(Y_1|\tau = 2) = EY_1$,
   (c) $E(Y_2|\tau = 2) = EY_2$, (d) $E(Y_1 + Y_2|\tau = 2) = 2EY_1$,
   (e) $E(Y_1|\tau \geq 1) = EY_1$, (f) $E(Y_2|\tau \geq 2) = EY_2$,
   (g) $E(Y_1 + Y_2|\tau \geq 2) = 2EY_1$.