

1.7 Proof of optimality of the SPRT. For the Neyman-Pearson Lemma, we had a statement, Theorem 1.1.3, not involving losses or priors, and another (Theorem 1.1.8) when there are losses and a prior. Here, the statement of Theorem 1.5.1 again did not involve losses or priors, and did not involve the cost c per observation, but the proof to be given for it will do so, much as in Theorems 1.2.5 and 1.2.6 which showed that under some conditions Bayes decision rules are admissible.

For a sequential randomized test $\{\phi_n\}$ of P vs. Q as defined in the last section, we can represent the randomization as follows. Let $A = \{-1, 0, 1\}$. Let $(\Omega, \mathcal{F}, \mu)$ be a probability space on which there are i.i.d. random variables y_1, y_2, \dots , uniformly distributed in $[0, 1]$. Take the y_i to be independent of the X_j . Then given $\omega \in \Omega$, let

$$\phi_n^\omega(X_1, \dots, X_n) := \begin{cases} -1, & \text{if } y_n \leq \phi_n(X_1, \dots, X_n)(-1) \\ 1, & \text{if } y_n > 1 - \phi_n(X_1, \dots, X_n) \quad (1) \\ 0, & \text{otherwise.} \end{cases}$$

If ω is chosen from Ω according to μ , then the non-randomized sequential test ϕ_n^ω is applied, the probabilities of decisions are just as in the randomized test. Randomized tests will, then, be represented by such ϕ_n^ω , where for each ω , $\{\phi_n^\omega\}_{n \geq 0}$ is a sequential test in the previously defined (non-randomized) sense. So, sequential randomized tests of P vs. Q are all “realizable” in a natural analogy with the notion defined in Sec. 1.3 and treated in Sec. 1.4, and $\{\phi_n^\omega(\cdot) : n \geq 0, \omega \in \Omega\}$ will be called a *realized sequential test* for P vs. Q . Here N will be the random n at which a choice of P or Q is made (or $+\infty$ if the choice is never made). So N depends on ω as well as $\{X_j\}_{j \geq 1}$. Error probabilities such as $\alpha(P, \phi)$ and average sample numbers such as $E_{P, \phi} N$ are naturally defined for sequential randomized tests.

For any given losses, let $v := L_{PQ}$, the loss if Q is chosen when P is true, and $w := L_{QP}$, the loss if P is chosen when Q is true. If a prior π is defined, let $p := \pi(P)$ and $q := 1 - p := \pi(Q)$.

For a sequential randomized test $\phi = \{\phi_n\}$, cost c per observation and p , v , etc. as above, the overall *risk* is defined as the total expected cost plus loss:

$$r(p, \phi) := pv\alpha(P, \phi) + qw\alpha(Q, \phi) + pcE_{P, \phi} N + qcE_{Q, \phi} N.$$

Note that as a function of p , for fixed ϕ , P , Q , v and w , $r(\cdot, \phi)$ is an affine (linear) function $r(p, \phi) = a + bp$ (since $q \equiv 1 - p$), i.e. a polynomial of degree 1 in p (or a constant). To see this, note that on the right side in the above definition there are two terms times p and two terms times q , which don't otherwise depend on p or q . Since $q = 1 - p$, this risk is an affine function of p .

Given P , Q , p , $c > 0$, $v > 0$, and $w > 0$, a ϕ which minimizes $r(p, \phi)$ and has finite risk will be called *Bayes*, agreeing with previous definitions except that we now have additional terms depending on the cost per observation.

For any test ϕ and $x \in X$ let $\phi^{(x)}$ be the test defined by

$$\phi_n^{(x), \omega}(y_1, \dots, y_n) := \phi_{n+1}^\omega(x, y_1, \dots, y_n)$$

for all x, y_1, \dots, y_n in X and $n = 1, 2, \dots$.

An equation $\phi_n = u$ for a realized test ϕ will mean that $\phi_n^\omega = u$ for (almost) all ω , where both ϕ_n and u may depend on X_1, \dots, X_n .

It will be shown that for any test taking at least one observation, the risk equals the expected posterior risk after one observation. In other words,

1.7.1 Lemma. If ϕ is a realized sequential test with $\phi_0 \equiv 0$, then

$$r(p, \phi) = c + \int \int r(p_x, \phi^{(x), \omega}) d(pP + qQ)(x) d\mu(\omega)$$

where p_x is the posterior probability of P after observing x .

Proof. It will be enough to prove this for non-randomized tests, then integrate. Let $r(P, \phi) := r(1, \phi)$ and $r(Q, \phi) := r(0, \phi)$. For $f(x) := R_{Q/P}(x)$ we have $p_x = p/(p + qf(x))$ and $q_x := 1 - p_x = qf(x)/(p + qf(x))$. Note that in this case the predictive measure γ on X is $pP + qQ$. Recall that by Theorem 1.3.7, the posterior is well-defined for γ -almost all x and doesn't depend on the dominating measure. Next,

$$r(p_x, \phi^{(x)}) = p_x r(P, \phi^{(x)}) + q_x r(Q, \phi^{(x)}).$$

Then, decomposing the integral with respect to x over the two sets $\{f < \infty\}$ and $\{f = \infty\}$ gives

$$\begin{aligned} & \int_{f < \infty} r(p_x, \phi^{(x)})(p + qf(x)) dP(x) + \int_{f = \infty} r(p_x, \phi^{(x)}) q dQ(x) \\ &= \int_{f < \infty} pr(P, \phi^{(x)}) + qf(x)r(Q, \phi^{(x)}) dP(x) + q \int_{f = \infty} r(p_x, \phi^{(x)}) dQ(x) \\ (1.7.2) \quad &= p \int r(P, \phi^{(x)}) dP(x) + q \int r(Q, \phi^{(x)}) dQ(x). \end{aligned}$$

On the other hand, $r(p, \phi) = pr(P, \phi) + qr(Q, \phi)$. For $x \in X^\infty$ write $x = \langle x_1, x^{(1)} \rangle$ where $x^{(1)} := \{x_j\}_{j \geq 2}$ and $\phi(x) = \langle \phi_0, \{\phi_n(x_1, \dots, x_n)\}_{n \geq 1} \rangle$. Let $N := N(x)$ be the least n such that $\phi_n(x_1, \dots, x_n) \neq 0$, in other words the time a decision is made. Let $1\{\dots\}$ denote the indicator function of the set $\{\dots\}$. Then

$$r(P, \phi) = \int L(P, \phi(x)) dP^\infty(x) \quad \text{with} \quad L(P, \phi(x)) := cN(x) + v1\{\phi_N = 1\},$$

and where P^∞ is the distribution for which x_1, x_2, \dots are i.i.d. with distribution P . By the Tonelli-Fubini theorem,

$$r(P, \phi) = \int \int L(P, \phi(\langle x_1, x^{(1)} \rangle)) dP^\infty(x^{(1)}) dP(x_1).$$

For $y = x_1$ fixed, and since $\phi_0 = 0$, the inner integral equals $c + r(P, \phi^{(y)})$. The corresponding fact for Q then combines to give

$$r(p, \phi) = (p + q)c + p \int r(P, \phi^{(y)})dP(y) + q \int r(Q, \phi^{(y)})dQ(y),$$

and using (1.7.2) completes the proof of Lemma 1.7.1. \square

Now, if we multiply the cost c per observation and the losses v and w all by a constant $h > 0$, leaving p fixed, then we multiply $r(p, \phi)$ by h for any test ϕ . (This can be considered a choice of units or scale for costs and losses.) So if ϕ is Bayes for c , v and w , it is also for hc , hv and hw . So in this proof (the proof of Theorem 1.5.1), letting $h := 1/(v + w)$, we can and will assume $v + w = 1$.

Let $r(p) := \inf\{r(p, \phi) : \phi \in \mathcal{C}\}$ where \mathcal{C} is the class of all realized sequential tests ϕ with $\phi_0 \equiv 0$, so that $N \geq 1$ always. Then for all $\phi \in \mathcal{C}$, $r(p, \phi) \geq c$ for $0 \leq p \leq 1$. If V is a test with $V_0 \equiv 0$ and $V_1 \equiv 1$, then $r(p, V) = c + pv$, so $c \leq r(p) \leq c + pv$ and $r(\cdot)$ is continuous at 0 (see Fig. 1.7A). Likewise, $c \leq r(p) \leq c + (1 - p)w$ implies r is continuous at 1. If $0 < t < 1$ and $x, y \in [0, 1]$, then for any test ϕ , we have by the affine form of $r(p, \phi)$ as a function of p (with other quantities fixed), as noted just after the definition of $r(p, \phi)$, that

$$(1.7.3) \quad r(tx + (1 - t)y, \phi) \equiv tr(x, \phi) + (1 - t)r(y, \phi),$$

and so

$$r(tx + (1 - t)y) = \inf\{tr(x, \phi) + (1 - t)r(y, \phi) : \phi \in \mathcal{C}\} \geq tr(x) + (1 - t)r(y).$$

Thus $r(\cdot)$ is concave, and so continuous on the open interval $(0, 1)$ (RAP, Theorem 6.3.4; $-r$ is convex). So $r(\cdot)$ is continuous on the closed interval $[0, 1]$.

Let $R(p) := \inf_{\phi} r(p, \phi)$ (the infimum over all tests, not just those in \mathcal{C}). Then $R(\cdot)$ can be approximated as follows:

1.7.4 Lemma. For any $\varepsilon > 0$ there is a function F from $[0, 1]$ to the set of all realized sequential tests, where F has finitely many tests as possible values, each on an interval, and where for $0 \leq p \leq 1$, $R(p) \leq r(p, F(p)) \leq R(p) + \varepsilon$.

Proof. As with $r(\cdot)$, $R(\cdot)$ is concave. Let S be the test with $S_0 \equiv 1$ and T the test with $T_0 \equiv -1$. So S chooses Q and T chooses P , each with no observations ($N \equiv 0$). Then $r(p, S) = pv$ and $r(p, T) = qw$. If at least one observation is taken, the risk is $r(p)$. The overall risk, as a function of p , is the minimum of these three risks (and could not be reduced by choosing at random among S, T and a test in \mathcal{C}). So as shown in Fig. 1.7A,

$$(1.7.5) \quad R(p) \equiv \min(pv, r(p), (1 - p)w),$$

so $R(\cdot)$ is continuous for $0 \leq p \leq 1$. Also, for any fixed test ϕ , $r(p, \phi)$ is affine (linear) in p by (1.7.3) and so continuous. For each $p \in [0, 1]$ there is a test ϕ_p with $r(p, \phi_p) < R(p) + \varepsilon$, and so $r(u, \phi_p) < R(u) + \varepsilon$ for all u in some relatively open interval J_p containing p in

c = cost per observation
 v = L_{PQ} = loss when P is true, Q is chosen
 w = L_{QP} = loss when Q is true, P is chosen
 (c, v, w) normalized so that $v + w = 1$

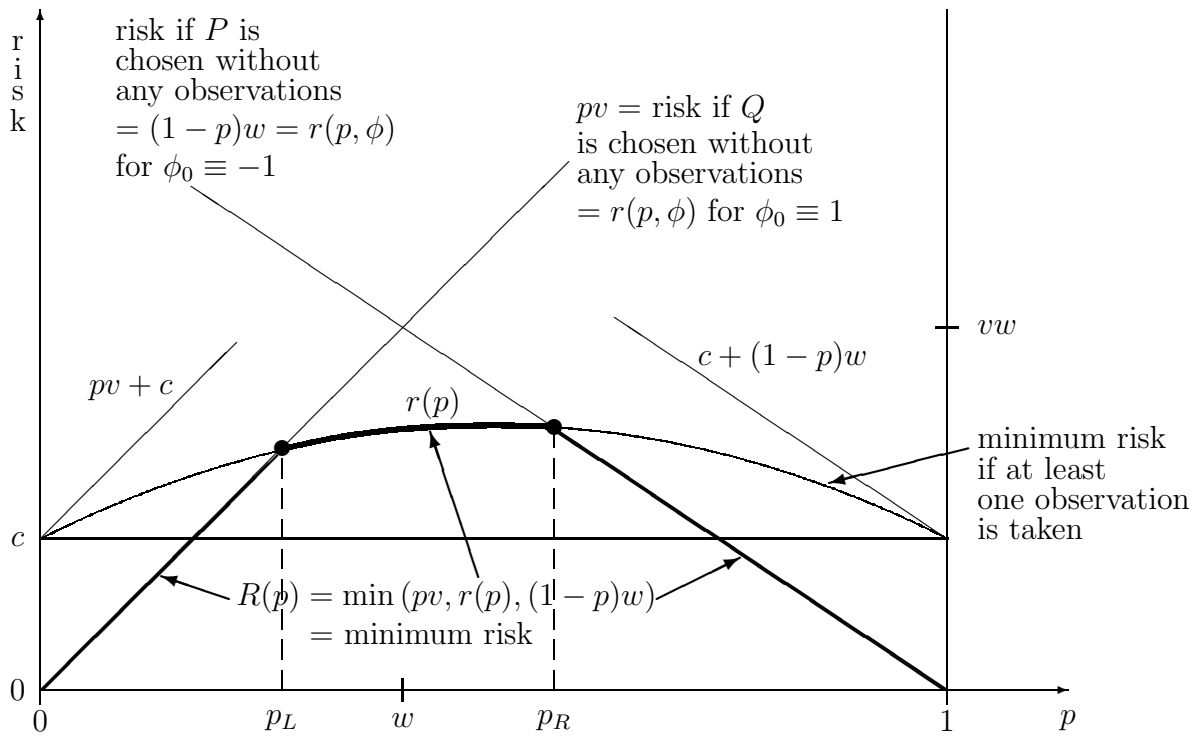


Figure 1.7A.

Fig. 1.7.A. Minimum risk for sequential tests

$[0, 1]$. By compactness these intervals have a finite subcover $J_{p(i)}$, $i = 1, \dots, m$. Then we can take disjoint subintervals $I_i \subset J_{p(i)}$ which cover $[0, 1]$ (but are no longer open), and set $F(u) = \phi_{p(i)}$ for $u \in I_i$ for each $i = 1, \dots, m$. \square

We also see in (1.7.5) that if a Bayes test ϕ of P vs. Q exists, and pv is strictly less than either of the other two terms on the right, we must have $\phi_0 \equiv 1$. If $(1-p)w$ is strictly the smallest of the three terms, then $\phi_0 \equiv -1$. And if $r(p)$ is strictly smallest, then $\phi_0 \equiv 0$.

Now if $r(w) \leq vw$, then the function $g(p) := r(p) - pv$ is concave and continuous on $[0, 1]$, $g(0) > 0$, and $g(w) \leq 0$. Thus by the intermediate value theorem, g has a root p_L with $0 < p_L \leq w$ (again, see Fig. 1.7A). Since $g(0) > 0$ and g is concave, the root is unique (if $0 < a < b < 1$ and $g(b) = 0$ then $g(a) > 0$ by concavity). Likewise, the function $h(p) := r(p) - (1-p)w$ has a unique root p_U with $w \leq p_U < 1$. Or if $r(w) > vw$, then let $p_L := p_U := w$.

1.7.6 Proposition. If $p_L < p < p_U$ then

$$R(p) = c + \int R(p_x) d(pP + qQ)(x).$$

Proof. In the given range, $R(p) = r(p)$. Let ϕ be any test with $\phi_0 = 0$. Then by Lemma 1.7.1,

$$r(p, \phi) = c + \int r(p_x, \phi^{(x)}) d(pP + qQ)(x) \geq c + \int R(p_x) d(pP + qQ)(x).$$

Taking the infimum over such ϕ ($\phi \in \mathcal{C}$) gives

$$R(p) = r(p) \geq c + \int R(p_x) d(pP + qQ)(x).$$

(Note that since $x \mapsto p_x$ is measurable and R is continuous, $x \mapsto R(p_x)$ is measurable.)

For the reverse inequality, given $\varepsilon > 0$ and a fixed p , take F from Lemma 1.7.4 and define a test η with $\eta_0 \equiv 0$ and $\eta^{(x)} := F(p_x)$ for all x . Then by Lemma 1.7.1 again,

$$\begin{aligned} R(p) &\leq r(p, \eta) = c + \int r(p_x, \eta^{(x)}) d(pP + qQ)(x) \\ &\leq c + \varepsilon + \int R(p_x) d(pP + qQ)(x). \end{aligned}$$

Letting $\varepsilon \downarrow 0$ finishes the proof. \square

The Neyman-Pearson lemma on admissibility (Theorem 1.1.3) was proved above before the characterization of Bayes tests (Theorem 1.1.8). In contrast, for sequential tests, the following sufficient condition for the Bayes property will be proved, then used in the rest of the proof of Theorem 1.5.1.

1.7.7 Theorem. Suppose given P, Q, c, p and v with $0 < w := 1 - v < 1$. Let $q := 1 - p, A := p(1 - p_U)/(qp_U)$ and $B := p(1 - p_L)/(qp_L)$. Then $\psi := \text{SPRT}(A, B)$ is Bayes for P vs. Q .

Conversely, any Bayes sequential test of P vs. Q must, almost surely for P^n and Q^n for each n , if no previous decision was made, continue (take at least one more observation) when $A < r_n < B$, or choose P if $r_n < A$, or Q if $r_n > B$.

Proof. If $p \geq p_U$, then $A \geq 1$ and ψ chooses P immediately, attaining $r(p, \psi) = R(p) = (1 - p)w$. Likewise if $p \leq p_L$ and $p < p_U$, then $A < B \leq 1$ and ψ chooses Q immediately, attaining $r(p, \psi) = R(p) = pv$. So ψ is Bayes in these cases.

If $r(w) \geq vw$, then $p_L = p_U = w$ and $p \geq p_U$ or $p \leq p_L$ for all p . So we can assume $r(w) < vw$,

We are left with the main case $p_L < p < p_U$. Let $p(x) := p_x$. Now $p_x \geq p_U$ is equivalent to $p \geq p_U(p + qf(x))$ or $f(x) \equiv r_1(x) \leq A$, while $p_x \leq p_L$ is likewise equivalent to $r_1(x) \geq B$. Thus if $p_L < p < p_U$, but $p(X_1) \leq p_L$ or $p(X_1) \geq p_U$, then ψ makes a choice after one observation, achieving a risk $c + R(p(X_1))$, matching the corresponding part of the integral in Proposition 1.7.6. The case $p_L < p(X_1) < p_U$ remains. After n observations, the posterior probability of P is

$$p^{(n)} := p^{(n)}(x_1, \dots, x_n) := p/(p + r_n q),$$

where $r_n := \prod_{1 \leq j \leq n} f(X_j)$. Now $p_L < p^{(n)}$ if and only if $B > r_n$. Also, $p^{(n)} < p_U$ if and only if $r_n > A$.

Moreover, by the remarks after the proof of Lemma 1.7.4, applied after the n th observation, if a Bayes sequential test has not made a decision previously, it must choose P if $r_n < A$, choose Q if $r_n > B$, and continue if $A < r_n < B$, proving the last statement in Theorem 1.7.7.

Let C_n be the event that $A < r_j < B$ for $j = 0, 1, \dots, n$, so that ψ takes at least $n + 1$ observations. For $n \geq 1$ let $V_n := C_{n-1} \cap \{r_n \leq A\}$ and $W_n := C_{n-1} \cap \{r_n \geq B\}$.

1.7.8 Lemma. If $p_L < p < p_U$ and $n = 1, 2, \dots$, then

$$R(p) \geq T_n(p) := \sum_{j=1}^n p[P^j(W_j)(v + jc)] + q[Q^j(V_j)(w + jc)] + jc[pP^j(V_j) + qQ^j(W_j)].$$

Proof. The proof will be by induction on n . For $n = 1$,

$$T_1(p) = c[(pP + qQ)(W_1 \cup V_1)] + vpP(W_1) + wqQ(V_1) \leq R(p)$$

by Proposition 1.7.6 as follows. The coefficient of c above is ≤ 1 . For $p_x \geq p_U$, which is the event V_1 , $R(p_x) = (1 - p_x)w$. Likewise for $p_x \leq p_L$, which is W_1 , $R(p_x) = p_x v$.

For the induction step, note that $T_n(p)$ is the risk restricted to the complement of C_n . If $x(n) := (x_1, \dots, x_n)$, then on C_n , $p_L < \pi_{x(n)} < \pi_R$, so Proposition 1.7.6 applies to the induction step from n to $n + 1$ just as in the proof for $n = 1$. \square

On the other hand,

$$\begin{aligned} r(p, \psi) &= pr(P, \psi) + qr(Q, \psi) = p \int L(P, \psi(x)) dP^\infty(x) + q \int L(Q, \psi(x)) dQ^\infty(x) \\ &\leq T_n(p) + p \int_{C_n} L(P, \psi(x)) dP^\infty(x) + q \int_{C_n} L(Q, \psi(x)) dQ^\infty(x). \end{aligned}$$

As $n \rightarrow \infty$, the events C_n decrease, $P^\infty(C_n) \downarrow 0$ and $Q^\infty(C_n) \downarrow 0$ by Lemma 1.5.2. On $C_{n-1} \setminus C_n$, $L(\mu, \psi(x)) \leq nc + \max(v, w)$ for $\mu = P$ or Q . By Lemma 1.5.2 and the fact that $\sum_n (a + nc)\gamma^n$ converges whenever $\gamma < 1$, for any constants a, c , it follows that the integrals in the last display are finite and decrease to 0 by dominated convergence or monotone convergence. So $r(p, \psi) \leq R(p)$, and $r(p, \psi) = R(p)$ and ψ is Bayes, finishing the proof of Theorem 1.7.7. \square

Now, the dependence of p_L and p_U on c and v (with $v + w = 1$) will be indicated in the notation, setting $p_L := p_L(c, v)$, $p_U := p_U(c, v)$.

1.7.9 Lemma. The functions $p_L(\cdot, v)$ and $p_U(\cdot, v)$ are continuous in c for $c > 0$ and each fixed v .

Proof. For each p and ϕ , the risk $r(p, \phi)$ is a non-decreasing function of c , which will be called $r(p, \phi; c)$ or $r(p, \phi; c, v)$. Thus, $r(p; c, v) := r(p; c) := r(p)$ is non-decreasing in c for each fixed p and v . Given p, c and $\delta > 0$, let ϕ be a test such that $r(p, \phi; c) < r(p; c) + \delta$. Let $M := \max(E_{P, \phi} N, E_{Q, \phi} N) < \infty$ since $c > 0$ and

$$M \leq [r(p; c) + \delta] / (c \cdot \min(p, q)).$$

Take any $c_1 \leq c \leq c_2$. Then

$$M \leq M_1 := [r(p; c_2) + \delta] / (c_1 \cdot \min(p, q)).$$

Then if $c_1 \leq c < c_2$, for any $h > 0$ such that $c_1 \leq c < c + h \leq c_2$,

$$r(p; c + h) \leq r(p, \phi; c + h) \leq r(p, \phi; c) + M_1 h \leq r(p; c) + M_1 h + \delta.$$

Letting $\delta \downarrow 0$ gives that $0 \leq r(p; c + h) - r(p; c) \leq M_1 h$. In other words $r(p; c)$ is, for $c > 0$, locally Lipschitz in c . So, it is continuous in c . Recall that $r(p; c) \geq c$ for all p . Thus there is a unique smallest $c := c_0 := c_0(v) := c_0(v, p, P, Q)$ such that $r(w; c) \geq vw$, with $r(w; c_0) = vw$, or equivalently such that $p_L = p_U = w$. Thus for $c \geq c_0$, $p_L(c, v) = w$, which is continuous (being constant) in c . For $0 < c \leq c_0$, $r(w; c) \leq vw$ and $p_L(c, v)$ is the unique solution p for $0 < p < 1$ of $r(p; c) = pv$, with $r(p; c) > pv$ if $p < p_L$ and $r(p; c) < pv$ if $p > p_L$. For any $d \geq c$, we have

$$vp_L(d, v) = r(p_L(d, v); d, v) \geq r(p_L(d, v); c, v),$$

and hence $p_L(d, v) \geq p_L(c, v)$, so p_L is non-decreasing in c .

Suppose $c_n \downarrow c < c_0$ and $p_L(c_n, v) \downarrow p' > p_L(c, v)$. Then

$$r(p'; c, v) < vp' \leq r(p'; c_n, v) \rightarrow r(p'; c, v),$$

a contradiction. So p_L is continuous in c from the right. Likewise if $c_n \uparrow c \leq c_0$ and $p_L(c_n, v) \uparrow p' < p_L(c, v)$, then

$$r(p'; c, v) > vp' \geq r(p'; c_n, v) \rightarrow r(p'; c, v),$$

again a contradiction. Thus p_L is continuous in c . The proof for p_U is symmetrical. \square

1.7.10 Lemma. For any $\varepsilon > 0$, $P \neq Q$, $0 < A < 1 < B < \infty$, and $\psi = \text{SPRT}(A, B)$, there exist $c > 0$, $v \in (0, 1)$, and $p \in (0, \varepsilon)$ such that ψ is Bayes for the given P , Q , p , c and v , and another such p with $1 - \varepsilon < p < 1$.

Proof. Fix $v \in (1 - \varepsilon/B, 1)$ and let

$$j(c) := p_L(c, v)[1 - p_U(c, v)]/[p_U(c, v)(1 - p_L(c, v))].$$

Then j is a continuous function of $c > 0$, by Lemma 1.7.9, since $0 < p_L < 1$ and $0 < p_U < 1$. For c large enough, $p_L = p_U$, and then $j(c) = 1$. As $n \rightarrow \infty$, the infimum of error probabilities

$$\inf\{\max(\alpha(P, \phi), \alpha(Q, \phi)) : \phi \in \mathcal{C}_n\} \rightarrow 0$$

where \mathcal{C}_n is the class of ordinary (non-sequential) tests $\phi(x_1, \dots, x_n)$, as can be seen from 1.5.3 with B and $1/A$ (in 1.5.3, not here) large, then 1.5.2. Thus as $c \rightarrow 0$, $\sup_{0 \leq p < 1} r(p; c) \rightarrow 0$, so $p_L(c, v) \rightarrow 0$ and $p_U(c, v) \rightarrow 1$. So $j(c) \rightarrow 0$. Thus by the intermediate value theorem there is a c such that $j(c) = A/B$.

For that c , let $p_L := p_L(c, v)$, $p_U := p_U(c, v)$ and $p := p_L B / (1 - p_L + p_L B)$, with $q := 1 - p$. Then $p/q = p_L B / (1 - p_L) = p_U A / (1 - p_U)$, so by Theorem 1.7.7, $\text{SPRT}(A, B)$ is Bayes. We have $0 < p \leq p_L B \leq wB < \varepsilon$ as desired. To find another $p > 1 - \varepsilon$, choose instead $v \in (0, A\varepsilon)$, proceeding otherwise as above. Then $q = (1 - p_U) / (1 - p_U + p_U A) \leq (1 - p_U) / A \leq v/A < \varepsilon$. \square

Proof of Theorem 1.5.1. The optimality of SPRTs will be proved also as compared to sequential randomized tests of P vs. Q . Given $\psi = \text{SPRT}(A, B)$, with $0 < A < 1 < B < \infty$, and $\varepsilon > 0$, Lemma 1.7.10 gives a v , $w = 1 - v$, $c > 0$, and a p with $0 < p < \varepsilon$ for which ψ is Bayes. Let ϕ be another (randomized) test with $\alpha(P, \phi) \leq \alpha(P, \psi)$ and $\alpha(Q, \phi) \leq \alpha(Q, \psi)$.

Case I. Suppose $E_{P, \phi} N < \infty$ and $E_{Q, \phi} N < \infty$. Then by the Bayes property of ψ ,

$$0 \leq r(p, \phi) - r(p, \psi) = pv(\alpha(P, \phi) - \alpha(P, \psi)) + qw(\alpha(Q, \phi) - \alpha(Q, \psi))$$

$$+ pc(E_{P, \phi} N - E_{P, \psi} N) + qc(E_{Q, \phi} N - E_{Q, \psi} N) \leq \varepsilon c(E_{P, \phi} N - E_{P, \psi} N) + c(E_{Q, \phi} N - E_{Q, \psi} N).$$

Letting $\varepsilon \downarrow 0$, it follows that $E_{Q, \phi} N \geq E_{Q, \psi} N$. Likewise, taking $p > 1 - \varepsilon$ from Lemma 1.7.10 so that $q < \varepsilon$ and letting $\varepsilon \downarrow 0$ gives $E_{P, \phi} N \geq E_{P, \psi} N$ as desired.

Case II. Suppose $E_{P,\phi}N = +\infty$. (Note: this case has little practical interest, as one would not want to use a test with infinite risk (expected cost of observations) under P , even if it had a smaller expectation for N under Q .) For a proof by contradiction, suppose that $E_{Q,\phi}N < E_{Q,\psi}N$.

Case IIa: Suppose also $\alpha(P, \psi) > 0$ and $\alpha(Q, \psi) > 0$. Let $C := \alpha(Q, \psi)/2$, $D := 2/\alpha(P, \psi)$, and $T := \text{SPRT}(C, D)$. Then by Lemma 1.5.2, $E_{P,T}N < \infty$ and $E_{Q,T}N < \infty$. By Lemma 1.5.3, $\alpha(P, T) < \alpha(P, \psi)$ and $\alpha(Q, T) < \alpha(Q, \psi)$. Let $T(t)$ be a randomized sequential test with $\Pr(T(t) = T) = t$ and $\Pr(T(t) = \phi) = 1 - t$. For t small enough,

$$\alpha(P, T) < \alpha(P, \psi), \quad \alpha(Q, T(t)) < \alpha(Q, \psi), \quad \text{and} \quad E_{Q, T(t)}N < E_{Q, \psi}N.$$

For each $k = 1, 2, \dots$, let $T(t, k)$ be the test with $T(t, k)_j = T(t)_j$ for $j < k$ and $T(t, k)_k = 1$, so that $T(t, k)$ acts the same as $T(t)$ through the first $k-1$ observations, then $T(t, k)$ chooses Q . For $T(t, k)$ we have $N \leq k$, so $E_{P, T(t, k)}N < \infty$ for any k . We have

$$\alpha(Q, T(t, k)) \leq \alpha(Q, T(t)) < \alpha(Q, \psi) \quad \text{and} \quad E_{Q, T(t, k)}N \leq E_{Q, T(t)}N < E_{Q, \psi}N,$$

while for k large enough, $\alpha(P, T(t, k)) < \alpha(P, \psi)$. This contradicts Case I.

Case IIb: $\alpha(P, \psi) = 0$. So also $\alpha(P, \phi) = 0$. Note that $\alpha(Q, \psi) < 1$ by Lemma 1.5.3 since $A < 1$. If $P(r_1 > 1) > 0$, then for some $\varepsilon > 0$, $P(r_1 > 1 + \varepsilon) > \varepsilon$. For each n , $P^n(r_j > (1 + \varepsilon)^j \text{ for } j = 1, \dots, n) > \varepsilon^n$. For some n , $(1 + \varepsilon)^n > B$ so $\alpha(P, \psi) > \varepsilon^n > 0$, a contradiction. Thus $P(r_1 > 1) = 0$. Since $Q \neq P$, $Q\{r_1 < \infty\} = \int r_1 dP < 1$. Let $F := \{r_1 = \infty\}$, so $Q(F) > 0$.

Case IIb1: $\alpha(Q, \psi) = 0$ also. Then just as $P(r_1 > 1) = 0$ we also have $Q(r_1 < 1) = 0$. So $r_1 = 0, 1$ or $+\infty$ almost everywhere for $P + Q$, and the values of r_n are also 0, 1 or $+\infty$ almost surely for P^n or Q^n . Both error probabilities are 0 for both tests. So on the event that $r_1 = r_2 = \dots = r_n = 1$, almost surely neither test has made a decision yet. Thus ψ , which decides the first time $r_n \neq 1$, has $E_{Q, \psi}N \leq E_{Q, \phi}N$ as desired.

Case IIb2: $\alpha(Q, \psi) > 0$. Let $U(k)$ be the test which takes k observations and chooses Q if $r_k = +\infty$, otherwise P . For k large enough, $\alpha(Q, U(k)) = (1 - Q(F))^k < \alpha(Q, \psi)$, while $\alpha(P, U(k)) = 0$. Let U_s do ϕ with probability s and $U(k)$ with probability $1 - s$, where $0 < s < 1$. Then $\alpha(P, U_s) = 0$, $\alpha(Q, U_s) < \alpha(Q, \psi)$, and $E_{Q, U_s}N < \infty$.

Let $V(j)$ be the test which equals U_s before time j , and after the j th observation, if no decision was made previously, $V(j)$ chooses P . Then $\alpha(P, V(j)) = 0$. For j large enough, $\alpha(Q, V(j)) < \alpha(Q, \psi)$. So by Case I

$$E_{Q, \psi}N \leq E_{Q, V(j)}N \leq E_{Q, U_s}N = sE_{Q, \phi}N + (1 - s)E_{Q, U(k)}N \leq sE_{Q, \phi}N + (1 - s)k.$$

Letting $s \uparrow 1$, we get $E_{Q, \psi}N \leq E_{Q, \phi}N$, finishing cases IIb2, IIb and II.

Case III, interchanging P and Q in Case II, is symmetrical. So Theorem 1.5.1 is proved. \square

Risks can be defined without priors, as in Sec. 1.2, as follows. These will be denoted here by $\rho(\cdot, \cdot)$ to distinguish them from risks $r(p, \cdot)$ depending on priors. Let

$$\rho(\mu, \phi) := L_{\mu\nu}\alpha(\mu, \phi) + cE_{\mu, \phi}N$$

where $(\mu, \nu) = (P, Q)$ or (Q, P) and ϕ is any sequential test of P vs. Q . Recall that for two such tests ϕ, η , $\phi \preceq \eta$ means $\rho(\mu, \phi) \leq \rho(\mu, \eta)$ for $\mu = P$ and $\mu = Q$, and $\phi \prec \eta$ means that in addition, the inequality is strict for $\mu = P$ or Q . A sequential test η is *inadmissible* if there exists $\phi \prec \eta$, otherwise *admissible*.

1.7.11 Corollary. For given c, v, w, P and Q , if A and B are defined as in Theorem 1.7.7, for any $p = 1 - q$ with $0 < p < 1$, then $\psi := \text{SPRT}(A, B)$ is admissible.

Proof. We have for $0 \leq p \leq 1$ and any sequential test ϕ , $r(p, \phi) \equiv p\rho(P, \phi) + q\rho(Q, \phi)$. If $\phi \prec \psi$ then since $0 < p < 1$ it would follow that $r(p, \phi) < r(p, \psi)$, contradicting Theorem 1.7.7. \square

The Corollary is really a special case of Theorem 1.2.6.

PROBLEM

1. Suppose that in a sequential test of P vs. Q , there is a loss of $v = L_{PQ} = \$20$ if P is true and Q is chosen, and a loss $w = L_{QP} = \$20$ if Q is true and P is chosen. Let the cost per observation be $c = \$1$. Let p be the prior probability of P . Suppose that the minimum risk over all tests taking at least one observation is $r(p) = 1 + 5p(1 - p)$ dollars.
 - (a) Find p_L , such that if $p < p_L$ but not if $p > p_L$, we should choose Q without taking any observation.
 - (b) Find p_R , such that if $p > p_R$ but not if $p < p_R$, we should choose P without taking any observation.