SPRT (A. Wald WWII)

Simplest case: compare 2 simple hypotheses
\( X_1, X_2, \ldots \) iid sequence, 2 possible densities

\( H_0: f_0(x) \quad H_1: f_1(x) \)

Optimal FSS well-known: Neyman-Pearson lemma.

Sample of size \( n \). Likelihood ratio

\[ L_n = \frac{\prod_{i=1}^{n} f_1(x_i)}{\prod_{i=1}^{n} f_0(x_i)}. \]

Likelihood ratio test: choose \( r \)
- reject \( H_0 \) if \( L_n > r \)
- accept \( H_0 \) if \( L_n < r \)

Optimal from both Bayesian & frequentist viewpoint.

Frequentist: Let \( \alpha = P_0 \in L_n \geq r \) type I error

Among all tests of \( H_1 \) vs. \( H_0 \) based on the \( n \) \( X_i \)s with type I error \( \leq \alpha \), likelihood ratio test has greatest power.

Sequential version (Sequential Probability Ratio Test)
- large \( L_n \) reject \( H_0 \)
- small \( L_n \) accept \( H_0 \)
- intermediate \( L_n \) collect more data
SPRT

Choose constants \( 0 < A < L < B < \infty \)

Sample \( x_1, x_2, \ldots \) in sequence until \( L_n \notin (A, B) \)

Stopping time \( \tau = \inf \{ n \geq 1 : L_n \notin (A, B) \} \)

\[ = \infty \text{ if } L_n \in (A, B) \text{ for all } n. \]

Decision rule: at \( \tau \) (for \( \tau < \infty \))

- reject \( H_0 \) if \( L_{\tau} \notin B; \) accept \( H_0 \) if \( L_{\tau} \in A \)

More tractable in log scale

Let \( b = \ln B > 0 \) \quad a = \ln A < 0 \)

Let \( Z_i = \ln \{ f_1(x_i)/f_0(x_i) \} \)

Let \( S_0 = 0 \) and \( S_n = \frac{n}{i=1} Z_i; \) \( n > 1. \) \( \bar{S}_n = \ln L_n. \)

\[ S_n \]

\[ \begin{array}{c}
\text{b} \\
\text{reject } H_0 \\
\vdots \\
\text{a} \\
\text{accept } H_0 \\
\end{array} \]

asymmetric, random walk with linear absorbing boundaries

Sampling from one-parameter exponential family, \quad \Rightarrow \text{stopping boundary linear for sufficient statistic}

Ex: \( x_1, x_2, \ldots \) iid \( N(\mu, 1) \)

\[ H_0: \mu = \mu_0 \quad H_1: \mu = \mu_1 \]

\[ L_n = \frac{\prod_{k=1}^{n} e^{-\frac{1}{2} (x_k - \mu)^2}}{\prod_{k=1}^{n} e^{-\frac{1}{2} (x_k - \mu_0)^2}} \]

\[ l_n = \ln L_n = (\mu_1 - \mu_0) \left( \sum_{k=1}^{n} x_k - \frac{n(\mu_1 + \mu_0)}{2} \right) \]

\( \hat{\Sigma}X_k \) sufficient statistic.
Stopping time \( \tau = \inf \{ n \geq 1, (\mu, -M_0) \left[ \sum X_k - \frac{n}{2} (\mu + M_0) \right] \notin (a, b) \} \)

Let \( a^* = \frac{a}{M_1 - M_0} \quad b^* = \frac{b}{\mu_1 - M_0} \)

decision rule: reject \( H_0 \) if \( \sum X_k \geq b^* + \frac{\tau}{2} (\mu_1 + M_0) \)

accept \( H_0 \) if \( \sum X_k \leq a^* + \frac{\tau}{2} (\mu_1 + M_0) \)
Will SPRT terminate?

Open boundaries suggest it may go on however.

However

Let \( Z_i = \log \frac{f_i(x_i)}{f_0(x_i)} \) \( \text{iid} \).

Let \( \mu_j = E(Z_j) \) \( j = 0 \text{ or } 1 \).

SPRT: plot of \( \sum_{i=1}^{n} Z_i \) vs. \( n \).

\( \mu_j \) is average slope under \( H_j \).

\[ \mu_0 < 0 \text{ if } f_1(\cdot) \neq f_0(\cdot) \]

\[ \mu_1 > 0 \text{ if } f_1(\cdot) \neq f_0(\cdot) \]

\( \mu_0 = \int \log \frac{f_i(x)}{f_0(x)} \frac{f_0(x)}{f_i(x)} \frac{f_i(x)}{f_0(x)} dx \)

but \( \log x \leq x - 1 \) w. equality iff \( x = 1 \).

So \( \mu_0 < \int \frac{f_i(x)}{f_0(x)} - 1 \frac{f_0(x)}{f_i(x)} dx \)

\[ = \int f_i(x) dx - \int f_0(x) dx \]

\[ = 0 \]

Similarly, \( \mu_1 \) w. = iff \( f_0(x) = f_1(x) \).

average path under \( H_1 \)

expect finite termination.
Basic Theorems of Seq. Analysis
& Properties of SPRT.

1) Stein's Lemma
Statement: Wijsman, p.69 in G+S

Let $X_1, X_2, \ldots$ be iid, $P(X=0) < 1$,
Let $S_n = \sum_{i=1}^{n} X_i$

Let $a < 0 < b$ be given real nos.
Let $N = \inf \{ n \geq 1 ; \ S_n \leq a \text{ or } S_n \geq b \}$.

Then $N$ is exponentially bounded,
exists constants $C > 0$, $0 < p < 1$
s.t. $P(N \geq n) < C p^n$, $n=1,2, \ldots$

$\Rightarrow N$ has finite moment generating function
in neighborhood of 0

$\Rightarrow$ all moments exist and are finite

$\Rightarrow P(N < \infty) = 1$.

Stein's proof does not build on previous
intuitive argument. It is very dependent
on independent observations.

These regularity conditions apply to SPRT.
Its stopping time is exp. bounded, has finite moments
of all orders. It is finite w. prob. 1.
SPRT will terminate.
A NOTE ON CUMULATIVE SUMS

BY CHARLES STEIN

Columbia University

Let \( \{Z_i\} \) be a denumerable sequence of identical independent real-valued random variables. Two constants \( a > 0 > b \) are chosen and the random variable \( n \) is defined as the smallest integer for which one of the inequalities \( \sum_i Z_i \geq a \) or \( \sum_i Z_i \leq b \) holds. For any events \( E_1 \) and \( E_2 \), \( P[E_1] \) will denote the probability of the event \( E_1 \) and \( P[E_1 | E_2] \) the conditional probability of the event \( E_1 \) given that \( E_2 \) has occurred.

It will be shown that there exists \( k > 0 \) such that the moment generating function, \( E e^{nt} \) exists for any complex number \( t \) whose real part is less than or equal to \( k \), and as an immediate consequence that \( n \) has finite moments of all orders.

If \( d \) is any constant satisfying \( b < d < a \), then, for fixed \( m \),

\[
P \left( b < \sum_{i=1}^{m} Z_i + d < a \right) \leq P \left( b < \sum_{i=1}^{m} Z_i < c \right)
\]

where \( c = \lceil a \rceil + \lceil b \rceil \). We exclude the case \( P[Z_i = 0] = 1 \). Then there exists \( \epsilon > 0 \) such that either

\[
\delta_1 = P[Z_i \geq \epsilon] > 0 \quad \text{or} \quad \delta_2 = P[Z_i \leq -\epsilon] > 0.
\]

Taking, for example, the former alternative with \( m_1 = \left[ \frac{c}{\epsilon} \right] + 1 \),

\[
P \left( \sum_{i=1}^{m_1} Z_i \geq c \right) \geq P \left( Z_i \geq \epsilon \right. \quad \text{for} \quad i = 1, \ldots, m_1 \right) = \delta_1^{m_1} > 0
\]

where \( [w] \) denotes the largest integer less than or equal to \( w \). For any positive integer \( k \),

\[
P \left( n > km_1 \right) \leq P \left( n > km_1 | n > (k-1)m_1 \right)
\]

\[
\leq P \left( b < \sum_{i=1}^{m_1} Z_i < a \right) \quad \text{for} \quad s = 1, \ldots, (k-1)m_1 \]

since \( n > km_1 \) implies \( b < \sum_{i=1}^{m_1} Z_i < a \).

But \( \sum_{i=1}^{m_1} Z_i = \sum_{i=1}^{(k-1)m_1} Z_i + \sum_{i=(k-1)m_1+1}^{m_1} Z_i \) and the second sum on the right hand side is independent of all terms in the first sum.

Thus the distribution of \( \sum_{i=1}^{km_1} Z_i \) given \( \sum_{i=1}^{(k-1)m_1} Z_i \) for \( s = 1, \ldots, (k-1)m_1 \) depends only on \( \sum_{i=1}^{(k-1)m_1} Z_i \) so that

\[
P \left[ n > km_1 \right] \leq P \left[ \sum_{i=1}^{m_1} Z_i + \sum_{i=(k-1)m_1+1}^{m_1} Z_i < a \right]
\]

\[
\leq P \left[ \sum_{i=(k-1)m_1+1}^{m_1} Z_i < c \right] \leq 1 - \delta_1^{m_1} \quad \text{by (1) and (2)}.
\]

Consequently, by induction on \( k \),

\[
P \left[ n > m \right] \leq P \left[ n > \left\lceil \frac{m}{m_1} \right\rceil \right] \leq \left( 1 - \delta_1^{m_1} \right)^{\left\lceil \frac{m}{m_1} \right\rceil}.
\]

Let \( t_0 \) be any positive number less than \( \frac{1}{m_1} \log (1 - \delta_1^{m_1}) \).

Then

\[
E e^{nt} = \sum_{n=0}^{\infty} e^{nt} P [n = m] \leq \sum_{n=0}^{\infty} e^{nt} P [m \leq n \leq km_1]
\]

\[
\leq \sum_{n=0}^{\infty} e^{nt} P [n > (k-1)m_1]
\]

\[
\leq \sum_{n=0}^{\infty} e^{nt} (1 - \delta_1^{m_1})^{k-1}
\]

\[
= \frac{1}{1 - \delta_1^{m_1}} \sum_{n=0}^{\infty} (e^{nt} (1 - \delta_1^{m_1}))^k.
\]

But this is a geometric series with decreasing terms, and is consequently convergent. Thus for any \( t \) whose real part \( R(t) \leq t_0 \), the moment generating function \( E e^{nt} \) exists. Since, for all positive \( l \), \( m_1^l < e^{nt} \) for sufficiently large \( m, n \) has finite moments of all orders.

Vol. 17 1946
A.1 Proof that the Probability is 1 That the Sequential Probability Ratio Test Will Eventually Terminate

The sequential probability ratio test terminates at the $n$th trial where $n$ is the smallest integer for which either

$$z_1 + \cdots + z_n \geq \log A$$

or

$$z_1 + \cdots + z_n \leq \log B$$

Let $c = |\log B| + |\log A|$. We shall subdivide the infinite sequence $z_1, z_2, z_3, \cdots$, ad inf., into segments of length $r$ where $r$ is some positive integer. Thus, the first segment $S_1$ will consist of the elements $z_1, \cdots, z_r$, the second segment $S_2$ will contain the elements $z_{r+1}, \cdots, z_{2r}$, etc. In general, the $k$th segment $S_k$ will consist of the elements $z_{(k-1)r+1}, \cdots, z_{kr}$. Let $\xi_k$ denote the sum of the elements in the $k$th segment. It can be seen that if the infinite sequence $z_1, z_2, \cdots$, ad inf., is such that the sequential process never terminates, then we must have

$$|\xi_k| < c \text{ for } k = 1, 2, \cdots, \text{ad inf.}$$

Inequality (A:1) can also be written

$$(\xi_k)^2 < c^2 \text{ for } k = 1, \cdots, \text{ad inf.}$$

(A:2)

Thus, in order to show that the probability is 1 that the sequential process will eventually terminate, it is sufficient to prove that the probability is 0 that (A:2) holds for all integral values $k$. For any given positive integer $i$ denote by $P_i$ the probability that $\xi_i^2 < c^2$. Since $z_1, z_2, \cdots$, are independently distributed, each having the same distribution, the distribution of $\xi_i$ must be the same for all values $i$. Hence, also $P_i$ is independent of $i$ and we shall denote it by $P$. Since $\xi_1, \xi_2, \cdots, \text{etc.}$, are independently distributed, the probability of the joint event that (A:2) holds for $k = 1, 2, \cdots, j$ is equal to $P^j$. Hence, in order to show that the probability is 0 that (A:2) holds for all values $k$, it is sufficient to show that $P < 1$. Clearly, if the expected value of $\xi_i^2$ is $> c^2$, then $P$ must be $< 1$. Since the variance of $z_i$ is assumed to be positive, the expected value of $\xi_i^2$ can be made arbitrarily large by choosing $r$, i.e., the number of elements in a segment, sufficiently large. Thus, $P < 1$, and we have proved the proposition: The probability is 1 that the sequential probability ratio test procedure will eventually terminate.
2) Wald's likelihood ratio identity (LRI)

(G+S, p. 70, proof Siegmund, p. 13).

Let $Z_1, Z_2, \ldots$ be arbitrary sequence of r.v.'s (stoch. process) let $T$ be stopping time with respect to $\mathcal{F}_{T_3}$

Defn: a r.v. $Y$ is prior to $T$ if $Y I_{\mathcal{F}_{T_3}}$ (equiv. $Y I_{\mathcal{F}_{T \leq T_3}}$) is a fn. of $(Z_1, Z_2, \ldots, Z_n)$ fn.

Interp: By the time $T$, observer who knows values of $Z_1, Z_2, \ldots, Z_T$ also knows value of $Y$.

$T$ is prior to itself

LRI: Let $P_0, P_1$ denote 2 probabilities for $\mathcal{F}_{T_3}$

Let $Z_1, \ldots, Z_n$ have st. pdf or mass fn. $P_{1, n}$ under $P_1$ (i.e., $P_1$)

then $l_n = P_{1, n}/P_{0, n}$ likelihood ratio.

For any stopping time $T$ non-neg. r.v. $Y$ prior to $T$

$$E_1 (Y I_{\mathcal{F}_{T < T_3}}) = E_0 (Y I_T I_{\mathcal{F}_{T < T_3}})$$
Siegmund's proof of LRI (filled in)

\[ E_{I_i}(Y_{I_i} \mid \xi T = n_3) = E_{I_i}(\sum_{n=1}^{\infty} Y_{I_i} \mid \xi T = n_3) \]

\[ = \sum_{n=1}^{\infty} E_{I_i}(Y_{I_i} \mid \xi T = n_3) \]

Interchange order of integration and summation is always valid provided function in this case \( Y_{I_i} \mid \xi T = n_3 \) is non-neg. So, explains need for non-neg \( Y \) as regularity condition.


Since \( T \) is a stopping time w.r.t. \( \xi Z_n \) and \( Y \) is prior to \( T \), both \( Y \) \& \( I_i \mid \xi T = n_3 \) are functions of \( \xi Z_n \), \( Z_n \) \& \( T \).

So

\[ = \sum_{n=1}^{\infty} \int_{\xi T = n_3} Y_{(i_1, \ldots, i_n)} p_{I_n}(z_{i_1}, \ldots, z_n) dz_{i_1} \cdots dz_n \]

\[ = \sum_{n=1}^{\infty} \int_{\xi T = n_3} Y_{(i_1, \ldots, i_n)} \frac{p_{I_n}(z_{i_1}, \ldots, z_n)}{p_{I_n}(z_{i_1}, \ldots, z_n)} p_{I_n}(z_{i_1}, \ldots, z_n) dz_{i_1} \cdots dz_n \]

\[ = \sum_{n=1}^{\infty} E_{0}(Y_{I_n} I_{\xi T = n_3}) \]

\[ = E_0(\sum_{n=1}^{\infty} Y_{I_n} I_{\xi T = n_3}) \text{ Fubini again since } Y_{I_n} I_{\xi T = n_3} \text{ for all } n \]

\[ = E_0(Y_{I_{\xi T < \omega_3}}) \text{ QED.} \]
Special Case of LRI

\[ Y = \mathbf{1}_A \text{ indicator for event } A. \]

\[ E_1(Y \mathbf{1}_z \mathbf{1}_A) = \mathbb{P}_1(A \cap z \mathbf{1}_A) = E_0 \left( L_k^\top \mathbf{1}_z \mathbf{1}_A \right) \]

Intuition from FSS: Consider an example.

Suppose \( z_1, \ldots, z_n \) iid; pdfs either \( f_1 \) or \( f_0 \).

\[ L_n = \prod_{i=1}^n f_1(z_i) / \prod_{i=1}^n f_0(z_i). \]

Let \( T = K \) degenerate

\( A: \) event of interest, may involve all \( k \) \( z \)’s.

LRI: \( \mathbb{P}_1(A) = E_0 (L_k \mathbf{1}_A) \).

pf: \( \mathbb{P}_1(A) = \sum_A \prod_{i=1}^k f_1(z_i) \, dz_1 \cdots dz_k \)

\[ = \sum_A \prod_{i=1}^k f_1(z_i) \prod_{i=1}^k f_0(z_i) \, dz_1 \cdots dz_k \]

\[ = \sum_A L_k \prod_{i=1}^n f_0(z_i) \, dz_1 \cdots dz_k \]

\[ = E_0 (L_k \mathbf{1}_A). \]
RI gives stopping bounds for SRT:

Using LRI, one shows:

Type I error \( \alpha = P_0 \xi > \mathcal{A}_1 \mathcal{B}_2 \leq \frac{1 - \beta}{\beta} \) H.W.

Type II error \( \beta = P_1 \xi > \mathcal{A}_1 \mathcal{B}_2 \leq \mathcal{A}(1 - \alpha) \)

Note: this does not require that data pts. be indep., but does require \( \xi \leq 0 \).

Inequalities because \( L \xi \) usu. overshoots boundary.

Ignore overshoot (assume \( L \xi \) hits boundary), algebra gives.

\[
\beta = (1 - \beta)/\alpha \quad \Leftrightarrow \quad \alpha = (1 - \alpha)/(\beta - \alpha)
\]

\[
A = \beta/(1 - \alpha) \quad \Rightarrow \quad \beta = A(\beta - 1)/(\beta - A)
\]

Pre-specify error sizes \( \alpha \& \beta \); LRT gives boundaries \( A + B \).

True error sizes not precisely \( \alpha \& \beta \) because of overshoot.

But near nominal \( \alpha \& \beta \), particularly when \( \alpha \& \beta \) small.

Let \( \alpha' = \text{true type I error} \) and \( \beta' = \text{true type II error} \).

\[
\alpha' \leq \frac{1 - \beta'}{\beta} = \frac{1 - \beta'}{(1 - \alpha')(B)} \leq \frac{\alpha'(1 - \beta')}{1 - \beta}
\]

Similarly, \( \beta' \leq \beta/(1 - \alpha) \) and \( \alpha' + \beta' \leq \alpha + \beta \).

At least one of \( \alpha' \leq \alpha + \beta' \leq \beta \) holds (because of sum). Usu. both hold. More strengthen then nominal.
It is interesting to note that the calculation of the Wald boundaries $A$ and $B$ depends only on the choice of $\alpha$ and $\beta$ and not on the hypotheses or the distributions of $Z_i$. Thus the same values for $A$ and $B$ would be used whenever one requires $\alpha \leq .01$ and $\beta \leq .05$. The hypotheses enter only in the formula for $Z_i$.

$$Z_i = \log \frac{f_i(x_i)}{f_0(x_0)}$$

from Eisenberg & Ghosh, chap. 3, p. 52 in G&S.