Wald's Lemma.

Let $Z_1, Z_2, \ldots$ be iid w. mean $\mu < \omega$ and $\text{var} \sigma^2 < \omega$.

Let $Z_0 = S_0 = 0$ and $S_n = \sum_{i=0}^{n} Z_i$.

Let $\tau$ be a stopping time w.r.t. $\mathcal{F}_n$ s.t. $E(\tau) < \infty$.

Then:

\begin{align*}
E(S_{\tau+}) &= \mu E(\tau) \quad (a) \\
E(S_{\tau+} - \tau \mu)^2 &= \sigma^2 E(\tau) \quad (b)
\end{align*}

Of course $E(S_n) = n \mu$ and $E[(S_n - n \mu)^2] = \sigma^2 n$.

Proof:

a) $S_n - n \mu$ is a martingale. Call it $M_n$.

Apply simpler optional stopping theorem. 2. $E(\tau) < \omega$ by assumption.

The martingale difference $M_{n+1} - M_n = Z_{n+1} - \mu$.

$E[|Z_{n+1} - \mu| \mid \mathcal{F}_n] = E[|Z_{n+1} - \mu|] \text{ by indep}$

since $|Z_{n+1} - \mu| \leq 1 + |Z_{n+1} - \mu|^2$.

$E[|Z_{n+1} - \mu|] \leq 1 + \sigma^2 = k$.

We have mart regularity conditions:

$\Rightarrow E(S_{\tau+} - \tau \mu) = 0 \Rightarrow E(S_{\tau+}) = \mu E(\tau)$. 


For proof based on Doob's Optional Sampling Theorem, see Ghosh, Mukh. & Sen, p. 27.

Also, Siegmund, prop. 2.18, p. 12 gives nm martingale pf.
It does not require $\sigma^2 < \infty$.

\[ b) \ (S_n - ny)^2 - n\sigma^2 \text{ is a martingale} \]

As above, let $\tau$ be a stopping time wrt $\{Z_n, \tau \leq \infty\}$.
Assume $\sigma^2 < \infty$.

Then $E[(S_{\tau} - \tau \mu)^2] < \sigma^2 E(\tau)$.

However, cannot simply prove result by a single application of (some version of) Optional Stopping Th'm.

Proof more complicated.

Proofs based on optional stopping:

Applications of Wald's Lemma.

1) Infinite expected sampling size in "Sampling to foregone conclusion"

\(X_1, X_2, \ldots:\) iid unknown mean \(\mu\) and known var. \(\sigma^2 < \infty\).

\(H_0: \mu = 0.\)

Let \(S_n = \sum_{i=1}^{n} X_i.\)

\(Z\) test stat. for sample of size \(n\): \(\frac{|S_n|}{\sigma \sqrt{n}}\)

Sequential Approach: \(T_C = \inf \{ n \geq 1 : |S_n| > C \sigma \sqrt{n} \} \)

By LIL, \(P(T_C < \infty) = 1.\)

- For \(C > 1, E(T_C) = \infty.\)

\(Pf: by \ contradiction \)

By defn of \(T_C, S^2_{T_C} > C^2 \sigma^2 T_C.\)

So \(E(S^2_{T_C}) > C^2 \sigma^2 E(T_C) > \sigma^2 E(T_C) \) for \(C > 1.\)

But if \(E(T_C) < \infty\) Wald's lemma b has

\(E(S^2_{T_C}) = \sigma^2 E(T_C)\) Contradiction

Ref: Woodroofe M (1982) p. 10
2) ASN of SPRT

\[ X_1, X_2, \ldots \text{ iid } Z_i = \log \frac{f_1(x_i)}{f_0(x_i)} \]

\[ \mu_i = E_i(Z) \quad i = 0, 1. \]

\[ \tau = \inf \{ n \geq 1 : l_n \notin (a, b) \} \quad l_n = \sum_{i=1}^{n} Z_i \]

Then \[ E_0 \tau \leq 3 \alpha \log \left( \frac{1-\beta}{\alpha} \right) + (1-\alpha) \log \left( \frac{\beta}{1-\alpha} \right) \frac{3}{\mu_0} \]

\[ = 3 \alpha b + (1-\alpha) a^3 / \mu_0 \]

\[ E_1 \tau \leq 3 (1-\beta) \log \left( \frac{1-\beta}{\alpha} \right) + \beta \log \left( \frac{\beta}{1-\alpha} \right) \frac{3}{\mu_1} \]

\[ = 3 (1-\beta) b + \beta a^3 / \mu_1 \]

Approx. signs mean ignoring overshoot.

pf: Since \[ E_i \tau < \infty \quad i = 0, 1 \text{ (from Stein's Lemma).} \]

If \[ \text{var}(Z) < \sigma \quad i = 0, 1 \text{ can apply Wald's Lemma.} \]

\[ E_i \tau = E_i l \tau / \mu_i \]

\[ = \frac{[E_i(l \tau \mid l \tau > b) P_i(l \tau > b) + E_i(l \tau \mid l \tau \leq a) P_i(l \tau \leq a)]}{\mu_i} \]

\[ = \frac{[b P_i(l \tau > b) + a P_i(l \tau \leq a)]}{\mu_i} \text{ ignoring overshoot.} \]

Assume \[ H_0. \quad P_0(l \tau > b) = \alpha \text{ and } P_0(l \tau \leq a) = 1-\alpha. \]

Since \[ b \leq \log \left( \frac{1-\beta}{\alpha} \right) \text{ and } a = \log \left( \frac{\beta}{1-\alpha} \right) \text{ result follows.} \]

Same idea for \[ H_1. \text{ see also Siegmund, p.11 (2.15) and (2.16) slightly different approach.} \]
Optimality

In iid case, SPRT minimizes both $E_{1}\gamma + E_{2}\gamma$ among all tests of $H_{0}$ vs. $H_{1}$ (including FSS) with the same or smaller error probabilities, $\alpha \& \beta$.

pf: Wijsman p. 76 in G&S

2.1.3 Optimum Property of the SPRT

Wald and Wolfowitz (1948, 1950) proved that among all sequential tests with given error probabilities and finite $E_{i}N$ ($i = 1, 2$) the SPRT minimizes simultaneously $E_{1}N$ and $E_{2}N$.

Another proof was given by Arrow, Blackwell, and Girshick (1949). There are essentially two parts to the proof: first, under a given loss and cost structure show that a Bayes test is a SPRT; second, given a SPRT show that there is a loss and cost structure that makes it Bayes. In Arrow, Blackwell, and Girshick (1949) there is a flaw in the second part of the proof as pointed out by Wolfowitz (1955). On the other hand, in Wald and Wolfowitz (1950) there is a measure-theoretic flaw in the first part of the proof. The union of these two papers contains a completely correct proof. Other proofs have been given by Lehmann (1959, Section 3.12, second part of the proof attributed to Lucien M. LeCam), Ghosh (1961), Burkholder and Wijsman (1963, with relaxation of the finite $E_{i}N$ condition), Matthes (1963), Chow and Robbins (1963), Chow, Robbins, and Siegmund (1971, Section 5.2(a)), and Lorden (1980).

Siegmund p. 21-22 show pf. if one could ignore overshoot very nice, based on Wald's LRI has additional ref. Ferguson.

When obs. not iid, optimality need not hold in fact, SPRT need not even terminate.

Eisenber & Ghosh, p. 58 G&S.
No overshoot

Among standard models, only testing situation where overshoot cannot occur:
Symmetric Bernoulli w. certain boundary values.

\[ X_i \sim \text{iid} \quad p(x=1) = \theta \quad p(x=0) = 1 - \theta \]

\[ H_0: \theta = \theta_0 \quad H_1: \theta = \theta_1 = 1 - \theta_0 \]

\[ z_i = \log \left[ \frac{f(x_i; \theta_1)}{f(x_i; \theta_0)} \right] = \pm d \quad d = \log \left[ \frac{1 - \theta_0}{\theta_0} \right] \]

If \( a + b \) are taken as integral multiples of \( d \), \( S_n \) = \( a \) or \( b \).

Wijnsma, p. 276 GdS.
Example W+6 p. 24-5

\[ X_i = \begin{cases} \leq 1 & \text{w. prob } \theta \\ > 1 & \text{w. prob } 1-\theta \end{cases} \]

\[ H_0: \theta = .25 \quad H_1: \theta = .75 \]

\[ \alpha = \beta = .001 \]

\[ b = \log \left(\frac{1-\beta}{\alpha}\right) = \log 999 = 6.907 \]

\[ a = \log \left(\frac{1-\alpha}{1-\beta}\right) = \log \frac{1}{999} = -6.907. \]

Likelihood ratio

\[ \frac{\Theta_1^x (1-\theta)^{1-x}}{\Theta_0^x (1-\theta)^{1-x}} \leq 3 \quad \text{if } x = 1 \]

\[ \frac{1}{3} \quad \text{if } x = 0 \]

\[ Z_i: \log \text{ likelihood ratio } = (2x-1) \log 3 \]

\[ \mu_0 = -5 \log 3 \quad \mu_1 = 5 \log 3. \]

\[ E \theta \equiv 36 \beta + (1-\alpha) a^3 / \mu_0 = 12.55 \equiv E_1 (\alpha) \text{ Symmetry} \]

Compare to FSS. Requires \( n = 33. \)

Reject \( H_0 \) if \( \sum X_i \geq 17. \) \( \alpha = \beta = 0.00095 \)

Discouraging data set:

Suppose first 17 obs. gave 9 1's \( \sum Z_i = \log 3 = 1.09 \)

next 16 \( 10 \) 1's \( \sum Z_i = 6 \log 3 = 6.59 \)

19 1's in 33 obs. If FSS reject \( H_0. \)

If SPRT, keep sampling.
SPRT: Composite one-sided hypothesis

\[ X_i \text{ iid w. } \text{pdf } f(x, \theta), \]
\[ H_0: \theta \leq \theta^* \quad H_1: \theta > \theta^* \]

more realistic than just 2 simple hypotheses, as heretofore.

Set up via 2 simple hypotheses

Specify \( \Theta_0 \) in \( \Theta_0 \) (i.e., \( \theta \leq \theta^* ) \) w. type I error \( \alpha \)
\[ \operatorname{Pr}(\text{reject } H_0) = \alpha \]

Specify \( \Theta_1 \) in \( \Theta_1 \) w. type II error \( \beta \)
\[ \operatorname{Pr}(\text{accept } H_1) = \beta \]

Do standard 2-state SPRT to test \( H_0: \theta = \theta_0 \) vs. \( H_1: \theta = \theta_1 \).

Let \( Z_i = \log \frac{f(x_i, \theta^*)}{f(x_i, \theta_0)} \)
\[ S_n = \sum_{i=1}^{n} Z_i \]

Let \( K = \inf \{ n \geq 1, S_n \in (a, b) \} \)
\[ a \& b \text{ chosen to give } \alpha \& \beta \]
\[ a \approx \log \frac{\beta}{1 - \alpha} \quad b \approx \log \frac{1 - \beta}{\alpha} \]

If \( S_K \leq a \), accept \( H_0 \); If \( S_K \geq b \), reject \( H_0 \).
Properties of composite SPRT

1) Finite termination. If $0 < a < b$, Stein's lemma holds for all $\Theta$.

2) Power: $\Pi(\Theta) = \Pr(\text{reject } H_0 | \Theta)$. Eisenberg (Chosh in 675 p. 54)

Suppose monotone likelihood ratio.
Whenever $\Theta_i < \Theta_j$, $f(x; \Theta_j) / f(x; \Theta_i)$ is non-decreasing in $x$.

Holds for 1 parameter exponential families.
Also $U(\Theta; \Theta_0, \Theta)$ non-central $t$ if

- Cauchy $f(x; \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}$
- Shifted $t$, add constant to central $t$ (non-central $t$)

If monotone likelihood ratio, $\Pi(\Theta)$ non-decreasing in $\Theta$.

$\Pi(\theta) \leq \alpha$ for $\theta \leq \theta_0$; $\Pi(\theta) \geq 1 - \beta$ for $\theta > \theta_1$.

Good control of error probs by controlling at 2 points.
However, unlike FSS case, this method of test construction does not give UMP test.

**UMP**: uniformly most powerful.

**Preliminary def'n**: Test \( \Theta \) of \( H_0: \Theta \in \Theta_0 \) vs. \( H_1: \Theta \in \Theta \), has size \( \alpha \) if

\[
\sup_{\Theta \in \Theta_0} P_{\Theta} (\Theta \text{ rejects } H_0) = \alpha.
\]

**Def'n**: Test \( \Theta \) is UMP of size \( \alpha \) for testing \( \Theta \in \Theta_0 \) vs. \( H_1: \Theta \in \Theta \); if \( \Theta \) is of size \( \alpha \)

and for any other test \( \Theta \) of size \( \leq \alpha \)

\[
P_{\Theta} (\Theta \text{ rejects } H_0) \geq P_{\Theta} (\Theta \text{ rejects } H_0)
\]

for each \( \Theta \in \Theta_1 \).

For FSS case, Neyman-Pearson Lemma (likelihood ratio test) typically takes form

Reject \( H_0 \) for \( X > k \)

where \( X \) is single obs. or sufficient statistic

and \( k \) is chosen to give type I error \( \alpha \).

For monotone likelihood ratio case, this gives UMP test because test does not involve \( \Theta \).
For sequential case, Eisenberg B & Ghosh, Bh, 1980, Annals of Stat., 8, 1123-31 (see p. 1126)

E & G consider special case $H_0: \Theta = \Theta_0$ vs. $H_1: \Theta > \Theta_0$

their def'n of UMP for sequential:

Test $(\gamma, S)$ is UMP if for all tests $(\gamma', S')$

with same stopping time $\gamma$, $\alpha' \leq \alpha$ implies

$P_{\Theta} (S$ rejects $) \geq P_{\Theta} (S'$ rejects $)$ for all $\Theta > \Theta_0$

SPRT is not UMP. The construction of the stopping time $\gamma$ in the construction of the decision rule depend on choice of $\Theta_1 \in \Theta_1$.

(c.f. normal example later)
Wald's Fundamental Identity of Sequential Analysis

Let $Z_1, Z_2, \ldots$ be non-degenerate iid r.v.'s with moment-generating function $\phi(\lambda) = E \exp(\lambda Z_I)$ finite for $\lambda$ in some open interval containing origin.

Let $S_n = \sum_{i=1}^{\infty} Z_i$.

Let $T$ be a stopping time wrt $Z_i$'s.

Then under regularity conditions

$$E \left[ \phi(\lambda)^{-T} \exp(\lambda S_T) \right] = 1$$

for any $\lambda$ s.t. $\phi(\lambda)$ is finite.

Minor comments:

1) Wald's original proof in Appendix 2.2. of his 1947 book also require $\phi(\lambda)^{-1}$. However, later authors don't need this. See, e.g. Wijsman, p. 71 in GoS.

2) The regularity conditions are met for Wald's SPRT.

This identity is useful for power function for SPRT for composite hypothesis.
Martingale Ref. of Fundamental Identity

Let $X_n = \phi(\lambda)^{-n} \exp(\lambda S_n)$

then $E X_n^3$ is a mart. wrt $E Z_n^3$ for any $\lambda$ with $\phi(\lambda)$ defined

a) $X_n \geq 0$, so $E|X_n| = EX_n = 1 \forall n$

b) Martingale property

$$X_n = \phi(\lambda)^{-1} \exp(\lambda Z_n) X_{n-1}$$

where $X_{n-1}$ is a function of $\{Z_1, Z_2, ..., Z_{n-1}\}$

$$E[X_n | Z_1, ..., Z_{n-1}] = E[\phi(\lambda)^{-1} \exp(\lambda Z_n) X_{n-1} | Z_1, ..., Z_{n-1}]$$

$$= X_{n-1} E[\phi(\lambda)^{-1} \exp(\lambda Z_n) | Z_1, ..., Z_{n-1}]$$

$$= X_{n-1} \phi(\lambda)^{-1} E[\exp(\lambda Z_n)] \text{ by indy}$$

$$= X_{n-1} \phi(\lambda)^{-1} \phi(\lambda)$$

$$= X_{n-1}.$$  

Need regularly conditions to apply Optional Stopping thm.

KoT, p.269 show how they hold for SPRT

Optional Stopping $\Rightarrow EX_T = EX_1 = 1$,

$$E[\phi(\lambda)^{-T} \exp(\lambda S_T)] = 1.$$