Title: Hypothesis tests and threshold crossing probabilities in random walks

Introduction:

In this set of notes we discuss two hypothesis tests: the MAP test and the sequential test. We then find bounds on the probability of error in both tests. We also discuss the Chernoff bound inequality used in finding bounds on error probabilities and provide a proof of it. We conclude these notes by discussing the power of the MAP test and the advantage of the sequential test.

Section 1:

In this section, we describe the MAP test sequential test and state the main results about the bound on the probability of error in the MAP test and the sequential test.

Maximum Apriori Probability (MAP test)

Let H be a discrete random variable such that $H \in \{0,1\}$. We assign apriori probabilities P_0 and P_1 such that

$P_0 = P(H=0)$, $P_1 = P(H=1)$ and $P_0 + P_1 = 1$.

We want to now perform a hypothesis test $H=0$ v.s. $H=1$.

To do this, we make a series of observations $Y_1, Y_2, \ldots, Y_n$ of H and based on the observations choose between $H=0$ and $H=1$.

We assume that $Y_1, Y_2, \ldots, Y_n$ is a sequence of random variables that are iid conditional on $H=0$ and are iid conditional on $H=1$.

Let $\vec{y} = (y_1, y_2, \ldots, y_n)$. and let $f_{Y|H}(\cdot|\cdot)$ be the conditional density of Y given H.

$$\Longrightarrow f_{\vec{Y}|H}(\vec{y}\,|\,l) = \prod_{i=1}^{n} f_{Y|H}(y_i\,|\,l)$$

where $l = 0$ or $l = 1$.

Now from the Bayes theorem we have that
$$P(H=0|\vec{y}) = \frac{P_0\, f_{\vec{Y}|H}(\vec{y}\,|\,0)}{\sum_{i=0}^{1} P_i\, f_{\vec{Y}|H}(\vec{y}\,|\,i)}$$

and
$$P(H=1|\vec{y}) = \frac{P_1\, f_{\vec{Y}|H}(\vec{y}\,|\,1)}{\sum_{i=0}^{1} P_i\, f_{\vec{Y}|H}(\vec{y}\,|\,i)}$$

$$\Longrightarrow \frac{P(H=0|\vec{y})}{P(H=1|\vec{y})} = \frac{P_0\, f_{\vec{Y}|H}(\vec{y}\,|\,0)}{P_1\, f_{\vec{Y}|H}(\vec{y}\,|\,1)}$$

$$= \frac{P_0}{P_1} \prod_{i=1}^{n} \frac{f_{Y|H}(y_i\,|\,0)}{f_{Y|H}(y_i\,|\,1)}$$

The MAP test states that

$$\text{if } \frac{p_0}{p_1} \prod_{i=1}^{n} \frac{f_{Y|H}(y_i \mid 0)}{f_{Y|H}(y_i \mid 1)} \begin{cases} > 1, & \text{choose } H=0 \\ \leq 1, & \text{choose } H=1 \end{cases}$$

Now we define $\lambda(\vec{y}) := \prod_{i=1}^{n} \frac{f_{Y|H}(y_i \mid 0)}{f_{Y|H}(y_i \mid 1)}$

as the likelihood ratio.

$$\Rightarrow \text{if } \lambda(\vec{y}) \begin{cases} > \frac{p_1}{p_0}, & \text{choose } H=0 \\ \leq \frac{p_1}{p_0}, & \text{choose } H=1. \end{cases}$$

Let $Z = \ln \frac{f_{Y|H}(y \mid 0)}{f_{Y|H}(y \mid 1)}$.

Let $\gamma_1(r) = \ln g_{Z|H=1}(r)$

where $g_{Z|H=1}(r)$. is the mgf of Z, conditional on $H=1$.

$\gamma_1(r)$ is called the semi-invariant mgf of Z conditional on $H=1$

Now we state the main theorem about bound on the probability of error conditional on $H=0$ and conditional on $H=1$ in the MAP test.

Theorem 1: Let $a = \frac{1}{n} \ln \frac{p_1}{p_0}$. Let r_1 be such that $\gamma_1'(r_1) = a$. Then

i) $P(\text{error} \mid H=1) = P(\text{choose } H=0 \mid H=1)$
$\leq e^{n[\gamma_1(r_1) - r_1 a]}$

ii) $P(\text{error} \mid H=0) = P(\text{choose } H=1 \mid H=0)$
$\leq e^{n[\gamma_1(r_1) + (1-r_1)a]}$.

Now we define the sequential test. We take off from the MAP test. In the MAP test, we had a fixed number of observations (n) i.e $Y_1, Y_2, \ldots, Y_n$. We stopped the test after recording n observations.

In the sequential test however, there isn't a fixed number of observations.

Let $S_n = \ln(\lambda(\vec{y}))$

$$\Rightarrow S_n = \sum_{i=1}^{n} Z_i \quad \text{where } Z_i = \ln \frac{f_{Y|H}(y_i \mid 0)}{f_{Y|H}(y_i \mid 1)}$$

We continue recording observations $Y_1, Y_2, \ldots, Y_n, Y_{n+1} \ldots$ until $\{S_n\}$ doesn't cross an upper fixed threshold $\alpha > 0$ or a lower fixed threshold $\beta < 0$. Once $\{S_n\}_{n \geq 1}$ does cross α or β, we stop our test and make the decision.

Thus let J be a stopping time such that J is the smallest $n \in \mathbb{N}$ such that $S_J \geq \alpha$ or $S_J \leq \beta$.

Sequential Test:

For each $n \in \mathbb{N}$ if

a) $S_n \in (\beta, \alpha)$, we continue the test by recording one more observation Y_{n+1}.

b) $S_n \geq \alpha$ i.e $S_J \geq \alpha$, we stop the test and choose $H=0$.

c) $S_n \leq \beta$ i.e $S_J \leq \beta$, we stop the test and choose $H=1$.

Now we state the main result about bounds or error probabilities conditional on $H=0$ and conditional on $H=1$ in the sequential test and conclude this section

Theorem 2: In the sequential test as defined above:

i) $P(\text{error} \mid H=1) = P(\text{choose } H=0 \mid H=1)$

$$\leq e^{-\alpha}$$

ii) $P(\text{error} \mid H=0) = P(\text{choosing } H=1 \mid H=0)$

$$\leq e^{\beta}$$

Section 2:

In this section, we prove the main results of Section 1 i.e Theorem 1 and Theorem 2. assuming three propositions needed for proving theorems 1 and 2.

We need two propositions as stated below for Theorem 1.

Proposition 1:

$S_n = \sum_{i=1}^{n} Z_i$ is a random walk conditional on $H=0$ and is a random walk conditional on $H=1$.

Proposition 2: (Chernoff Bound).

Let $\{S_n\}_{n\geq 1}$ s.t $S_n = X_1 + X_2, \dots + X_n$ s.t $X_1, X_2, \dots, X_n$ are iid, be a random walk. Let $g_X(r)$ be the mgf of X s.t it is finite over the interval (r_-, r_+) s.t $r_- < 0 < r_+$. Let $\gamma_X(r)$ be the semi-invariant moment generating function $\ln g_X(r)$.

Then we have that

$$P(S_n \geq na) \leq e^{n[\gamma(r) - ra]}.$$

If $\exists\, r_0 \in (0, r_+)$ such that $\gamma'(r_0) = a$, then

$$P(S_n \geq na) \leq e^{n[\gamma(r_0) - r_0 a]}$$

is the tightest bound on $P(S_n \geq na)$

Now we are ready to prove Theorem 1

Proof of Theorem 1:

$$S_n = \sum_{i=1}^{n} Z_i \quad , \quad Z_i = \ln \frac{f_{Y|H}(y_i|0)}{f_{Y|H}(y_i|1)}.$$

$a = \ln\left(\frac{P_1}{P_0}\right)$ and $\gamma_1(r)$ is the semi-invariant mgf of Z_i given $H=1$. r_1 is s.t $\gamma_1'(r_1) = a$.

We want to prove that

i) $P(\text{error} \mid H=1) \leq e^{n[\gamma_1(r_1) - r_1 a]}$

ii) $P(\text{error} \mid H=0) \leq e^{n[\gamma_1(r_1) + (1-r_1) a]}$.

For (i) we have that

$$P(\text{error} \mid H=1) = P\left(\lambda(\vec{y}) > \frac{p_1}{p_0} \,\middle|\, H=1\right)$$

$$= P\left(S_n > \ln\left(\frac{p_1}{p_0}\right) \,\middle|\, H=1\right).$$

$$= P(S_n > na \mid H=1).$$

From proposition 2 we have that

$$P(S_n > na \mid H=1) \leq e^{n[\gamma_1(r_1) - r_1 a]}.$$

This proves (i).

Now for (ii) we have

$$P(\text{error} \mid H=0) = P\left(\lambda(\vec{y}) \leq \frac{p_1}{p_0} \,\middle|\, H=0\right)$$

$$= P\left(S_n \leq \ln\left(\frac{p_1}{p_0}\right) \,\middle|\, H=0\right)$$

$$= P(S_n \leq na \mid H=0).$$

$\Rightarrow P(\text{error} \mid H=0) = P(S_n \leq na \mid H=0)$

$= P(-S_n \geq -na \mid H=0).$

Let us now define

$T_n = -S_n$ for all $n \in \mathbb{N}$.

such that $T_n = \sum_{i=1}^{n} X_i$ where

$X_i = -Z_i$.

$\Rightarrow$ Since $Z_i = \ln \frac{f_{Y|H}(y_i \mid 0)}{f_{Y|H}(y_i \mid 1)}$, we see

that $X_i = \ln \frac{f_{Y|H}(y_i \mid 1)}{f_{Y|H}(y_i \mid 0)}$.

Let $b = -a$. $\Rightarrow$

$P(S_n \leq na \mid H=0) = P(T_n \geq nb \mid H=0).$

Let $\gamma_0(r)$ be the semi-invariant mgf of X_i conditional on $H=0$.

$\Rightarrow \gamma_0 = \ln g_{X|H=0}(r)$ where

$g_{X|H=0}(r)$ is the mgf of X conditional on $H=0$.

Let r_0 be s.t $\gamma_0'(r_0) = b$.

Thus from Proposition 2, we have.

$$P(T_n \geq nb \mid H=0) \leq e^{n[\gamma_0(r_0) - r_0 b]}.$$

$\Rightarrow P(\text{error} \mid H=0) \leq e^{n[\gamma_0(r_0) - r_0 b]}$

Since $b = -a$,

$P(\text{error} \mid H=0) \leq e^{n[\gamma_0(r_0) + r_0 a]}$

We claim now that

$r_0 + r_1 = 1$ and $\gamma_0(r) = \gamma_1(1-r)$.

Hence it follows from our claim that $\gamma_0(r_0) = \gamma_1(1-r_0) = \gamma_1(r_1)$ and that $r_0 = 1 - r_1$. Substituting these into $P(\text{error} \mid H=0)$, we get that

$P(\text{error} \mid H=0) \leq e^{n[\gamma_1(r_1) + (1-r_1)a]}$

That proves (ii)

$\square$.

Proof of Claim:

$$\gamma_1(r) = \ln g_{Z_1|H=1}(r)$$

$$= \ln\left[E\left(e^{rZ_1} \mid H=1\right)\right].$$

$$= \ln\left[E\left[e^{r \ln \frac{f_{Y_1|H}(y_1|0)}{f_{Y_1|H}(y_1|1)}} \mid H=1\right]\right]$$

$$= \ln \int f_{Y_1|H}(y_1|1)\, e^{r \ln \frac{f_{Y|H}(y_1|0)}{f_{Y|H}(y_1|1)}}\, dy$$

$$= \ln \int \left[f_{Y_1|H}(y_1|1)\right]^{1-r} \left[f_{Y|H}(y_1|0)\right]^{r} dy.$$

$$\gamma_0(r) = \ln g_{X_1|H=0}(r)$$

$$= \ln\left[E\left(e^{rX_1} \mid H=0\right)\right]$$

$$= \ln\left[E\left[e^{r \ln \frac{f_{Y|H}(y|1)}{f_{Y|H}(y|0)}} \mid H=0\right]\right]$$

$$= \ln \int f_{Y|H}(y|0)\, e^{r \ln \frac{f_{Y|H}(y|1)}{f_{Y|H}(y|0)}}\, dy.$$

$$= \ln \int [f_{Y|H}(y|0)]^{1-r} [f_{Y|H}(y|1)]^{r}\, dy$$

$$= \ln \int [f_{Y|H}(y|1)]^{r} [f_{Y|H}(y|0)]^{1-r}\, dy.$$

$\Longrightarrow$

$\gamma_1(r) =$

$$\ln \int [f_{Y|H}(y|1)]^{1-r} [f_{Y|H}(y|0)]^{r}\, dy$$

$\gamma_0(r) =$

$$\ln \int [f_{Y|H}(y|1)]^{r} [f_{Y|H}(y|0)]^{1-r}\, dy.$$

$\Longrightarrow \gamma_0(r) = \gamma_1(1-r) \quad \forall r. \quad (1)$

$\gamma_0'(r_0) = b.$

$$\Longrightarrow \frac{d}{dr} \gamma_1(1-r) \Big|_{r=r_0} = b.$$

$\Rightarrow -\gamma_1'(1-r_0) = b.$

$\Rightarrow -\gamma_1'(1-r_0) = -a.$

$\Rightarrow \gamma_1'(1-r_0) = a.$ But $\gamma_1'(r_1) = a$

$\Rightarrow 1-r_0 = r_1 \Rightarrow r_0 + r_1 = 1$ — (2)

$\Rightarrow$ Both parts of our claim are verified and that proves our claim.

□.

Now we note some observations based on the function $\gamma_1(r)$ and $\gamma_0(r)$ and Theorem 1 that we have just proved.

$$\gamma_1(r) = \ln \int [f_{Y|H}(y|1)]^{1-r}[f_{Y|H}(y|0)]^r\, dy.$$

(I) $\gamma_1(0) = \ln \int [f_{Y|H}(y|1)]\, dy = \ln 1 = 0.$

$\Rightarrow \gamma_1(0) = 0.$ Similarly $\gamma_0(0) = 0$

(II) $\gamma_1(1) = \ln \int [f_{Y|H}(y|0)]\, dy = \ln 1 = 0$

$\Rightarrow \gamma_1(1) = 0.$ Similarly $\gamma_0(1) = 0$

(III) $\gamma_1(r)$ is convex. This is in fact true for any semi-invariant mgf. We shall see this in the next section.

$\Rightarrow \gamma_1''(r) \geq 0.$

(IV) From (ii) and (iii) it follows that $\gamma_1'(0) < 0$ and $\gamma_1'(1) > 0.$

(V) $\gamma_1'(0) = \left.\dfrac{g'_{Z|H=1}(r)}{g_{Z|H=1}(r)}\right|_{r=0} = E[Z|H=1]$.

(VI) Let $T(x)$ be the line passing through $(r_1, \gamma_1(r_1))$ with slope $\gamma_1'(r_1)$.

$\Rightarrow T(x) - \gamma_1(r_1) = \gamma_1'(r_1)[x - r_1]$.

$\Rightarrow T(x) = \gamma_1(r_1) + \gamma_1'(r_1)[x - r_1]$.

This is the tangent to $\gamma_1(r)$ at $(r_1, \gamma_1(r_1))$. Since $\gamma_1'(r_1) = a$,

$T(x) = \gamma_1(r_1) + a(x - r_1)$.

Then $T(0) = \gamma_1(r_1) - ar_1$

and $T(1) = \gamma_1(r_1) + a(1 - r_1)$.

And we know from Theorem 1 that

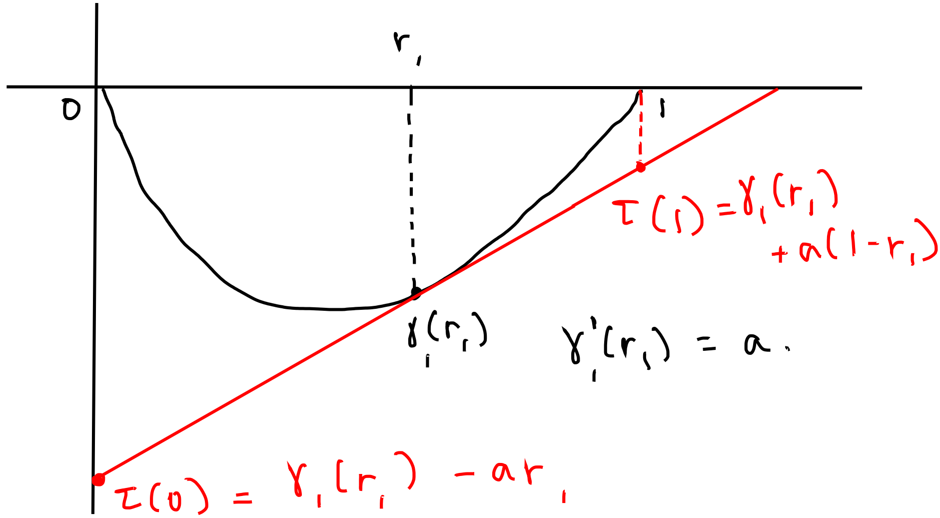
$P(\text{error} \mid H=1) \leq e^{n[\gamma_1(r_1) - ar_1]}$

$P(\text{error} \mid H=0) \leq e^{n[\gamma_1(r_1) + a(1-r_1)]}$.

$\Rightarrow P(\text{error} \mid H=1) \leq e^{nT(0)}$

$P(\text{error} \mid H=0) \leq e^{nT(1)}$

Judging from all the observations above, we have that



$a = \frac{1}{n} \ln\left(\frac{P_1}{P_0}\right).$ Thus note that as n varies the slope of the tangent varies. Thus as the slope increases/decreases, we see that $I(0)$ and $I(1)$ keep changing. Thus as $I(0)$ decreases, $I(1)$ increases and vice versa. Thus as $P(\text{error} \mid H=0)$ decreases, $P(\text{error} \mid H=1)$ increases and vice versa.

Now we move onto the proof of Theorem 2 of Section 1.

To prove Theorem 2, we need the following proposition as stated below. stated below:

Proposition 3: Let $\{X_i : i \geq 1\}$ be iid random variables and $\gamma(r) = \ln E[e^{rX}]$ be the semi invariant mgf of each X_i. Let $\gamma(r)$ be finite in the open interval (r_-, r_+) s.t $r_- < 0 < r_+$. Let $S_n = X_1 + \cdots + X_n$ and let $\alpha > 0$, $\beta < 0$, $\alpha, \beta \in \mathbb{R}$.

Let J be the smallest n for which $S_n \geq \alpha$ or $S_n \leq \beta$. Assume that $E[X] < 0$ and $\exists\, r^* > 0$ s.t $\gamma(r^*) = 0$. Then

$$P(S_J \geq \alpha) \leq e^{-r^* \alpha}.$$

Proof of Theorem 2

We want to prove for the sequential test that

(A) $P(\text{error} \mid H=1) \le e^{-\alpha}$

(B) $P(\text{error} \mid H=0) \le e^{\beta}$.

. . .

$$P(\text{error} \mid H=1) = P(\text{choosing } H=0 \mid H=1)$$

$$= P(S_J \ge \alpha \mid H=1), \quad \text{where}$$

$$S_n = \sum_{i=1}^{n} Z_i, \quad Z_i = \ln \frac{f_{Y|H}(y_i \mid 0)}{f_{Y|H}(y_i \mid 1)}$$

Thus, we apply proposition 3 here.
Z_i are iid conditional on $H=1$.
$\Rightarrow$ We take the semi-invariant mgf here to be $\gamma_1(r)$ which is the semi-invariant mgf of each Z given $H=1$.

From Proposition 3, we have then that

$P(S_J \geq \alpha \mid H=1) \leq e^{-r^* \alpha}$ where r^* is such that $\gamma_1(r^*) = 0$.

But in Observation (II), we saw that $\gamma_1(1) = 0$. $\longrightarrow r^* = 1$ here.

$\Longrightarrow P(S_J \geq \alpha \mid H=1) \leq e^{-\alpha}$.

$\Longrightarrow P(\text{error} \mid H=1) \leq e^{-\alpha}$.

That proves (A)

Now for (B).

$P(\text{error} \mid H=0) = P(\text{choosing } H=1 \mid H=0)$
$= P(S_J \leq \beta \mid H=0)$

Like we had done in the proof of Theorem 1, let $T_n = -S_n$ and since $S_n = \sum_{i=1}^{n} Z_i$, $T_n = \sum_{i=1}^{n} X_i$

where $X_i = -Z_i$

Since $Z_i = \ln \dfrac{f_{Y|H}(y_i|0)}{f_{Y|H}(y_i|1)}$,

$$X_i = \ln \frac{f_{Y|H}(y_i|1)}{f_{Y|H}(y_i|0)}$$

$\Rightarrow P(S_J \leq \beta | H=0) = P(-S_J \geq -\beta \mid H=0)$

$-\beta > 0$, $-\alpha < 0$.

Since J is the smallest n such that S_n crosses α or β, it follows that J is the smallest n such that T_n crosses $-\beta$ or $-\alpha$.

$\Rightarrow P(S_J \leq \beta | H=0) = P(T_J \geq -\beta \mid H=0)$.

To use proposition 3 here, we must choose the semi-invariant mgf of X_i as $\gamma_0(r)$ as we had defined in the proof of Theorem 1.

$\Longrightarrow$ From Proposition 3 it follows that

$$P(T_J \geq \beta \mid H=0) \leq e^{-r^*(-\beta)}$$

such that $\gamma_0(r^*) = 0$.

$\Longrightarrow$ But from observation II, we know that $\gamma_0(1) = 0$.

$\Longrightarrow$ $r^* = 1$.

$\Longrightarrow$ $P(T_J \geq \beta \mid H=0) \leq e^{-(-\beta)} = e^{\beta}$.

$\Longrightarrow$ $P(S_J \leq \beta \mid H=0) \leq e^{\beta}$

$\Longrightarrow$ $P(\text{error} \mid H=0) \leq e^{\beta}$

which proves (B).

$\square$.

Section 3:

In this section, we provide proofs of Propositions 1, 2 and 3 as stated in Section 2.

Proof of Proposition 1:

We want to prove that $\{S_n\}_{n \geq 1}$ as defined in the MAP test is a random walk conditional on $H=0$ is a random walk conditional on $H=1$.

$$S_n = \sum_{i=1}^{n} Z_i \quad \text{where} \quad Z_i = \ln \frac{f_{Y|H}(y_i | 0)}{f_{Y|H}(y_i | 1)}$$

$$\text{Let } \Gamma(Y) = \ln \frac{f_{Y|H}(Y|0)}{f_{Y|H}(Y|0)}.$$

Then $Z_i = \Gamma(y_i)$ for $i \in \{1, 2, \ldots, n\}$.

But $Y_1, Y_2, \ldots, Y_n$ are iid conditional on $H=0$ and iid conditional on $H=1$.

$\Rightarrow \Gamma(Y_1), \Gamma(Y_2), \ldots, \Gamma(Y_n)$ are iid conditional on $H=0$ and iid conditional on $H=1$.

But $Z_i = \Gamma(Y_i)$ for $i \in \{1, 2, \ldots, n\}$.

$\Rightarrow$ Each Z_i is the same finite function of Y_i

$\Rightarrow Z_1, Z_2, \ldots, Z_n$ are iid conditional on $H=0$ and iid conditional on $H=1$.

$\Rightarrow S_n = Z_1 + Z_2 + \cdots + Z_n$ is a random walk conditional on $H=0$ and is a random walk conditional on $H=1$.

$\square$.

Now, we move onto the proof of Proposition 2 which is the Chernoff Bound for Random Walks.

Before proving Proposition 2, we will prove that for any random variable X, the semi-invariant mgf $= \gamma_X(r)$ $= \ln(g_X(r))$ finite in $r_- < 0 < r_+$, is a convex function.

<u>Theorem 3:</u> $\gamma_X(r)$ is a convex function and if $\gamma_X(r)$ is convex, then a local minimum of $\gamma_X(r)$ is a global minimum.

<u>Proof:</u> $\gamma_X(r) = \ln[g_X(r)]$

$= \ln[E(e^{rX})]$.

Holder's inequality states that

$$E[UV] \le [E|U|^p]^{\frac{1}{p}} [E|U|^q]^{\frac{1}{q}}$$

for any $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$

$U = e^{(1-\theta)r_1 X}$, $V = e^{\theta r_2 X}$ s.t

$p = \frac{1}{1-\theta}$, $q = \frac{1}{\theta}$.

$$\rightarrow E[UV] \le [E(e^{(1-\theta)r_1 X})^{\frac{1}{1-\theta}}]^{(1-\theta)} \cdot [E(e^{\theta r_2 X})^{\frac{1}{\theta}}]^{\theta}$$

$$= [E(e^{r_1 X})]^{(1-\theta)} [E(e^{r_2 X})]^{\theta}.$$

$\Rightarrow E[e^{((1-\theta)r_1 + \theta r_2)X}]$

$\leq [E[e^{r_1 X}]]^{1-\theta} [E(e^{r_2 X})]^{\theta}$

$\Rightarrow \ln E[e^{((1-\theta)r_1 + \theta r_2)X}]$

$\leq (1-\theta) \ln E[e^{r_1 X}] + \theta \ln E[e^{r_2 X}]$

$\Rightarrow \gamma_X((1-\theta)r_1 + \theta r_2)$

$\leq (1-\theta) \gamma_X(r_1) + \theta \gamma_X(r_2)$

$\Rightarrow \gamma_X(r)$ is a convex function.

Now let r_{loc} be a local minimum

and r_{glob} are a global minimum.

s.t $\gamma_X(r_{glob}) < \gamma_X(r_{loc})$

Since r_{loc} is a local minimum, $\exists$

$\varepsilon > 0$ for which $\gamma_X(r_{loc}) < \gamma_X(r)$

s.t $r \in (r_{loc} - \varepsilon, r_{loc} + \varepsilon)$.

Let θ be small enough such that

$r_\theta = \theta r_{loc} + (1-\theta) r_{glob}$

$\in (r_{loc} - \varepsilon, r_{loc} + \varepsilon)$.

$\Rightarrow \gamma_X(r_{loc}) \leq \gamma_X(r_\theta)$

$\gamma_X(r_\theta) \leq \theta \gamma_X(r_{loc}) + (1-\theta)\gamma_X(r_{glob})$
[Since γ_X is convex].

But $\theta \gamma_X(r_{loc}) + (1-\theta)\gamma_X(r_{glob})$
$< \gamma_X(r_{loc})$ since
$\gamma_X(r_{loc}) > \gamma_X(r_{glob})$

$\Rightarrow \gamma_X(r_{loc}) < \gamma_X(r_{loc})$
which is a contradiction.

$\Rightarrow$ A local minima must be a global minima. □.

Now we are ready to prove proposition 2.

Proof of Proposition 2
(Chernoff Bound).

We want to prove that if $S_n = X_1 + \cdots + X_n$ is a random walk such that $X_1, X_2, \ldots, X_n$ are iid random variables having semi variant mgf $\gamma_X(r) = \ln g_X(r)$, that is finite and existent over the interval $r_- < 0 < r_+$, then

$P(S_n \geq na) \leq e^{n[\gamma_X(r) - ra]}$

for any $r \in (0, r_+)$. We also want to prove that if $\exists\ r_0 \in (0, r_+)$ s.t $\gamma_X'(r_0) = a$, then

$P(S_n \geq na) \leq e^{n[\gamma_X(r_0) - r_0 a]}$

is the tightest bound.

For any random variable Y, we have that $P(Y \geq y) \leq \frac{E[Y]}{y}$.
from Markov Inequality.

For any given random variable Z, let $I(Z)$ be the interval over which its mgf $g_Z(r)$ exists and is finite.

Mgf $g_Z(r) = E[e^{rZ}]$.

For $Y = e^{rZ}$, for any $r \in I(Z)$,

$$P(e^{rZ}) \leq \frac{g_Z(r)}{y}$$

Let $y = e^{rb}$.

$Z \geq b \Longleftrightarrow e^{rZ} \geq e^{rb}$ if $r > 0$.

$\Rightarrow P(Z \geq b) \leq g_Z(r) e^{-rb}$ if $0 < r \in I(Z)$

If $S_n = X_1 + X_2 + \cdots + X_n$,
X_i, iid, then

$g_{S_n}(r) = [g_X(r)]^n$.

$\Rightarrow P(S_n \geq na) \leq [g_X(r)]^n e^{-rna}$
if $0 < r \in I(Z)$.

$\gamma_X(r) = \ln[g_X(r)]$ is finite and exists on (r_-, r_+).

$\Rightarrow P(S_n \geq na) \leq e^{n[\gamma_X(r) - ra]}$
for $r \in (0, r_+)$.

Now we want to tighten the bound and minimise $\gamma_X(r) - ra$.

We take the derivative of $\gamma_X(r) - ra$ i.e $\gamma_X'(r) - a$ and equate it to 0.

Let r_0 be such that $r_0 \in (0, r_+)$ and $\gamma_X'(r_0) - a = 0$.

Further taking the derivative, we get $\gamma_X''(r)$.

But $\gamma_X(r)$ is convex from Theorem 3.

$\Rightarrow \gamma_X''(r) \geq 0$. But $\gamma_X''(r) > 0$ if X is not deterministic.

$\Rightarrow$ Since X is not deterministic,

$\gamma_X''(r) > 0$.

$\Rightarrow \gamma_X''(r_0) > 0$

$\Rightarrow \gamma_X'(r_0) = 0, \ \gamma_X''(r_0) > 0$.

$\Rightarrow r_0$ is a local minimum.

But since γ_X is convex, we get from theorem 3 that r_0 is a global minimum for $\gamma_X(r) - ra$.

$\Rightarrow$ If $\exists\ r_0 \in (0, r_+)$ s.t

$\gamma_X'(r_0) = a$, then $\gamma_X(r) - ra$ is smallest for $r = r_0$.

$\Rightarrow P(S_n \geq na) \leq e^{n[\gamma_X(r_0) - r_0 a]}$

is the tightest bound on $P(S_n \geq na)$

Which is what we wanted to prove.

$\square$.

We finally move onto the proof of Proposition 3 before concluding this section.

<u>Proof of Proposition 3</u>

For the proof of proposition 3, we assume Wald's identity which states in context of the conditions of Proposition 3 that for any $r \in (r_-, r_+)$, we have that $E[e^{rS_J - J\gamma(r)}] = 1$ where $\gamma(r)$ is the semi invariant mgf of X_i where $S_n = X_1 + \cdots + X_n$ and J is the smallest n for which $S_n \geq \alpha$ or $S_n \leq \beta$ for $\alpha, \beta \in \mathbb{R}$, $\alpha > 0$, $\beta < 0$.

We want to prove that if $\exists\ r^* > 0$ s.t. $\gamma(r^*) = 0$, then $P(S_J \geq \alpha) \leq e^{-r^* \alpha}$.

From Wald's Theorem, we have that

$E[e^{r^* S_J - J\gamma(r^*)}] = 1$.

But $\gamma(r^*) = 0 \Rightarrow E[e^{r^* S_J}] = 1$.

$\Rightarrow P(S_J \geq \alpha)\ E[e^{r^* S_J} \mid S_J \geq \alpha]$

$+ P(S_J \leq \beta)\ E[e^{r^* S_J} \mid S_J \leq \beta]$

$= 1$.

But $P(S_J \leq \beta)\ E[e^{r^* S_J} \mid S_J \leq \beta]$.

is non-negative.

$\Rightarrow P(S_J \geq \alpha)\ E[e^{r^* S_J} \mid S_J \geq \alpha] \leq 1$.

Now given $S_J \geq \alpha$, we have that

$e^{r^* S_J} \geq e^{r^* \alpha}$.

$\Rightarrow E[e^{r^* S_J} \mid S_J \geq \alpha] \geq E[e^{r^* \alpha}]$

$= e^{r^* \alpha}$.

$\Rightarrow P(S_J \geq \alpha)\ e^{r^* \alpha} \leq 1$.

$\Rightarrow P(S_J \geq \alpha) \leq e^{-r^* \alpha}$

$\square$.

Section 4:

In this section we discuss about the power of the MAP test and prove the Neyman Pearson Lemma which states that the MAP test has maximum power for a fixed level of significance. Following that we discuss why the Sequential Test is more advantageous than a fixed length MAP test.

First we discuss the power of a test and what makes the MAP test most powerful.

The MAP test states that

$$\text{If } \lambda(\vec{y}) = \frac{f_{\vec{Y}|H}(\vec{y}\,|0)}{f_{\vec{Y}|H}(\vec{y}\,|1)} \begin{cases} > \frac{P_1}{P_0}, & \text{choose } H=0 \\ \leq \frac{P_1}{P_0}, & \text{choose } H=1 \end{cases}$$

The MAP test is an example of a threshold test with threshold $\eta = \frac{p_1}{p_0}$.

From here on we shall refer to the MAP test as a threshold test of threshold η.

Thus note that for the threshold test of threshold η, we have the test as

$$\text{If } \lambda(\vec{y}) \begin{cases} > \eta, & \text{choose } H=0 \\ \leq \eta, & \text{choose } H=1. \end{cases}$$

For any test designed to select between $H=0$ and $H=1$, we have that there exists a rejection region R for s.t if $\vec{y} \in R$, then we reject our null hypothesis i.e $H=0$. Thus for a threshold test of threshold η, we have that

$$R = \{\vec{y} : \lambda(\vec{y}) \leq \eta\}.$$

For any test, let

$\alpha = P(\vec{y} \in R \mid H=0)$. This is the probability of rejecting $H=0$ when in truth H is infact 0. (Type I error).

Let $1-\beta = P(\vec{y} \in R \mid H=1)$. This is the probabilit of rejecting $H=0$ when H is in fact 1. (Good stuff. We want to maximise this).

$1-\beta$ is called the power of the test.

$\beta = P(\vec{y} \in R^c \mid H=1)$ = Probability of accepting $H=0$ when in fact $H=1$. (Type II error). We want to minimise this.

Ideally we want to reduce both α and β. But as α decreases, β increases and vice versa. Hence we just take α to be a fixed value and reduce β instead.

Theorem 4: (Neyman Pearson Lemma)

Consider the hypothesis $H=0$ vs. $H=1$.
Let R be the rejection region for the test of threshold η such that $\alpha = P(\vec{y} \in R \mid H=0)$. Then, for any other test of rejection region R^* s.t $\alpha^* = P(\vec{y} \in R^* \mid H=0) \leq \alpha$, we have that $P(\vec{y} \in R^* \mid H=1) \leq P(\vec{y} \in R \mid H=1)$.

Proof: Let I_A be the indicator function for any set A. Thus there is a 1-1 correspondence b/w a rejection region and indicator.

$$R \Longleftrightarrow I_R, \quad R^* \Longleftrightarrow I_{R^*}.$$

$\Rightarrow$ If $I_R(\vec{y}) = 1$,
then $\eta f(\vec{y} \mid H=1) - f(\vec{y} \mid H=0) \geq 0$.

If $I_R(\vec{y}) = 0$,
then $\eta f(\vec{y} \mid H=1) - f(\vec{y} \mid H=0) < 0$.

$\Rightarrow$ It follows that $\forall \vec{y}$,

$$I_{R^*}(\vec{y})\left(\eta f(\vec{y} \mid H=1) - f(\vec{y} \mid H=0)\right)$$
$$\leq I_R(\vec{y})\left(\eta f(\vec{y} \mid H=1) - f(\vec{y} \mid H=0)\right)$$

$$\int I_{R^*}(\vec{y})\left[\eta f(\vec{y} \mid H=1) - f(\vec{y} \mid H=0)\right] d\vec{y}$$
$$= \eta P(\vec{y} \in R^* \mid H=1) - P(\vec{y} \in R^* \mid H=0)$$

$\Rightarrow$ Integrating on both sides, we have:

$$\eta P(\vec{y} \in R^* \mid H=1) - P(\vec{y} \in R^* \mid H=0)$$
$$\leq \eta P(\vec{y} \in R \mid H=1) - P(\vec{y} \in R \mid H=0).$$

$$\Rightarrow P(\vec{y} \in R \mid H=0) - P(\vec{y} \in R^* \mid H=0)$$
$$\leq \eta\left[P(\vec{y} \in R \mid H=1) - P(\vec{y} \in R^* \mid H=1)\right].$$

But $P(\vec{y} \in R \mid H=0) - P(\vec{y} \in R^* \mid H=0) \geq 0$.

$\Rightarrow$ RHS ≥ 0.

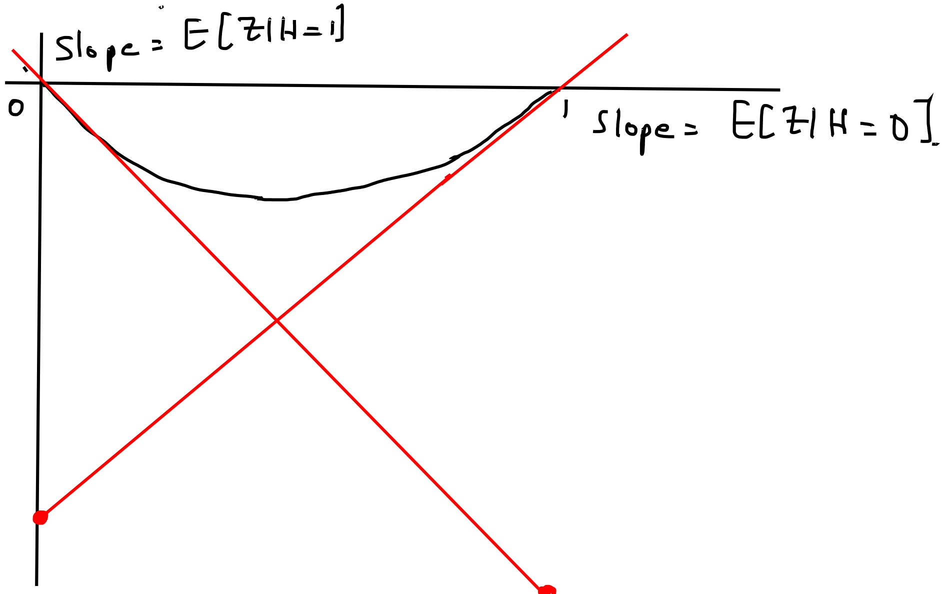
$\Rightarrow P(\vec{y} \in R \mid H=1) \geq P(\vec{y} \in R^* \mid H=0)$

$\Rightarrow$ The threshold test is more powerful than any other test for the same level of significance.

$\square$.

$\Rightarrow$ The MAP test at a threshold of $\frac{p_1}{p_0}$ is more powerful than any other test for the same level of significance.

Now we finally discuss why the Sequential Test is more advantageous than a MAP test.



Now say we want to find an exponentially tight bound on the P(error | H=1) with fixed length MAP test.

Then, we assume that $n = \dfrac{\alpha}{E[Z|H=0]}$

We take $r_i = 1$, $\gamma_i'(1) = $ slope

Then we know that

$$P(\text{error} \mid H=1) \le e^{n[\gamma_i(r_i) - r_i a]}$$

$$= e^{\frac{\alpha}{E[Z|H=0]}[\gamma_i(1) - E[Z|H=0]]}$$

Since $\gamma_i'(1) = E[Z|H=0]$.

$\rightarrow$ But $\gamma_i(1) = 0$.

$\rightarrow P(\text{error} \mid H=1) \le e^{-\alpha}$.

But the seesaw that marks of error $-\alpha$ for $f(0)$ goes to 0 for $f(1)$.

$\Rightarrow P(\text{error} \mid H=0) \le e^{0} = 1$

(Which is not useful)

Similarly if we find a tight bound for $P(\text{error} \mid H=0)$, we'll see that

$P(\text{error} \mid H=1) \le e^{0} = 1$.

(Not useful).

But in the sequential test, we simultaneously get the error exponent for $H=1$ that a fixed length test would give if we gave up on trying to find an error exponent for $H=0$ and vice versa.

We get both exponentially tight bounds simultaneously i.e

$P(\text{error} \mid H=1) \le e^{-\alpha}$

$P(\text{error} \mid H=0) \le e^{\beta}$.

That is the reason the sequential test is preferable as compared to the fixed length MAP test.

References:

1) Discrete Stochastic Processes (Sections 7.3 - 7.6) - Robert Gallager (MIT)

2) Convexity of log of Moment Generating function - Math Stack Exchange

3) Convex Optimization - Fall 2017 (CIMS - NYU).