

# OPTIMAL ORDERING OF OBSERVATIONS FOR FAST SEQUENTIAL DETECTION

*Rahul Iyer and Ahmed H. Tewfik*

The University of Texas at Austin  
Electrical and Computer Engineering  
Austin, TX 78712

## ABSTRACT

The effect of the ordering of independent and non-identical observations on the average number of samples needed to make a decision in a sequential binary hypothesis test is analyzed in this paper. We show that among all permutations of ordering of the observations, the average sample number (ASN) is minimum for the order in which the area under the receiver operating characteristic (ROC) curve for each of the non-identically distributed observations is monotonically decreasing. The claim is verified by computing the ASN of a generalized sequential probability ratio test (GSPRT) for different orderings of observations, which are independent and non-identical Gaussian random variables, using a combination of analytical and numerical techniques.

**Index Terms**— Generalized SPRT, Non-i.i.d. observations, Ordering, ASN

## 1. INTRODUCTION

Humans collect information sequentially and make a decision at a certain point rather than wait till the amount of information collected exceeds some preset threshold and then make a decision. In sequential decision making, the order in which information is presented to a person affects the time taken to make a correct or incorrect decision. If important information regarding the problem posed is presented at an early stage as opposed to a later stage, a correct decision can be made sooner. An example of this is to consider the speed of response of a person who is asked a question in different ways. To demonstrate the effect of ordering of information on sequential decision making we use the framework of sequential analysis with non-identical observations.

The optimal sequential test for a binary hypothesis test with i.i.d. observations is the sequential probability ratio test (SPRT) [1]. The ASN is calculated using Wald's identity and is inversely proportional to the Kullback Leibler (KL) divergence between the two hypotheses. The literature in the domain of sequential tests for non-i.i.d. observations is less extensive as compared to that of sequential analysis assuming i.i.d. observations. It is shown in [2], [3] that the optimal sequential test for binary hypothesis testing with non-i.i.d. ob-

servations is the generalized SPRT where the thresholds are time varying. The computation of the varying thresholds and the ASN of the GSPRT is a difficult problem. The ASN is computed for dependent but identically distributed samples in [4] where the dependence between the observations is a first order Markov chain defined on a finite state space. It is shown in [5] that two tests which are multihypothesis versions of the SPRT (MSPRT's) with constant thresholds, are asymptotically optimal in minimizing the ASN and in addition all higher order moments of the stopping time for multihypothesis testing problems with i.i.d. or non-i.i.d. observations when the error probabilities are small. In [6] the authors prove that a generalized SPRT using a maximum-likelihood ratio statistic with fixed thresholds is asymptotically optimal in minimizing the ASN for i.i.d. observations and also for the quasi-i.i.d. case where the log-likelihood ratio is expressed as the sum of an i.i.d. sequence and a slowly changing random sequence, for certain class of multihypothesis testing problems.

In this paper we study and characterize the optimal ordering that leads to the smallest average sample size in a sequential binary hypothesis testing problem. To our best knowledge there has been no work done on this problem before. The paper is organized as follows. In section 2, the GSPRT procedure is described. The problem statement of ordering of observations based on ROC curve is described in section 3. In section 4, a closed form expression is given for the ROC curve when considering a binary hypothesis test with i.i.d. Gaussian samples. Approximate expressions for the time varying thresholds and the ASN is obtained in section 5 for the binary hypothesis test with non-identically distributed independent Gaussian random variables as observations. Numerical results and plots are provided in section 6 and we conclude the paper in section 7.

## 2. GENERALIZED SPRT

It is well known that the optimal sequential test for a binary hypothesis test with i.i.d. observations is the SPRT in the sense that it minimizes the expected number of samples to be taken to achieve error probabilities less than the prescribed probability of miss ( $P_m$ ) and probability of false alarm ( $P_f$ ). The SPRT involves comparing the likelihood ratio (LRT) of

all observations up to the current time instant and comparing it with a pair of thresholds. It is described as follows:

If  $\prod_{k=1}^n l(Y_k) \geq B$ , decide  $H_1$  is true

If  $\prod_{k=1}^n l(Y_k) \leq A$ , decide  $H_0$  is true

If  $A < \prod_{k=1}^n l(Y_k) < B$ , take another observation

where  $Y_1, Y_2, \dots, Y_n$  are the i.i.d. observations,  $A = \frac{P_m}{1 - P_f}$ ,

$B = \frac{1 - P_m}{P_f}$ ,  $l(Y_k) = \frac{f(Y_k|H_1)}{f(Y_k|H_0)}$  and  $f(Y_k|H_i)$  ( $i = 0, 1$ )

is the probability density function (pdf) of the observation conditioned on the hypothesis ( $i = 0, 1$ ).

When the observations are not i.i.d., the optimal sequential test which minimizes the expected number of samples is the generalized SPRT [2]. It is given by the following rule:

If  $\prod_{k=1}^n l_k(Y_k) \geq B_n$ , decide  $H_1$  is true

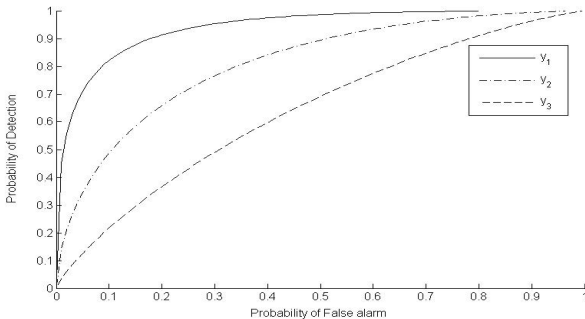
If  $\prod_{k=1}^n l_k(Y_k) \leq A_n$ , decide  $H_0$  is true

If  $A_n < \prod_{k=1}^n l_k(Y_k) < B_n$ , take another observation

The thresholds here are time varying instead of being fixed as in SPRT. The problem of finding the optimal time varying thresholds has not been solved yet. We propose an approximate method to calculate a sequence of varying thresholds in Section 5.

### 3. PROBLEM STATEMENT

Consider a binary hypothesis test with independent and non-identically distributed observations  $y_1, y_2, y_3, \dots$  and let the two hypotheses be denoted by  $H_0$  and  $H_1$ . The ROC curve of  $y_1, y_2$  and  $y_3$  is shown in the figure below. It can be seen that the ROC curve of  $y_1$  is higher than that of  $y_2$  which itself is higher than that of  $y_3$ .



**Fig. 1.** ROC curves for non-identically distributed observations  $y_1, y_2$  and  $y_3$

**Claim:** The ASN conditioned on either hypothesis corresponding to the ordering in which the area under the ROC

curve of each observation is monotonically decreasing, i.e.  $E[N|H_i]$ ,  $i = 0, 1$  is minimum among all possible orderings of observations and is thus optimal.

An immediate corollary of this is as follows: Let the samples be sequentially observed in the order  $y_2, y_1, y_3, \dots$  and let the corresponding ASN be  $E[N'|H_i]$  for  $i = 0, 1$ . Then  $E[N'|H_i] > E[N|H_i]$  for  $i = 0, 1$ .

### 4. RECEIVER OPERATING CHARACTERISTIC

The ROC curve of a test is a plot of the probability of detection  $P_d$  vs. the probability of false alarm  $P_f$  such that all pairs  $(P_d, P_f)$  below the curve are achievable and those above are not. Thus a test is deemed to be better than another if its ROC curve is higher, i.e. if the area under its ROC curve is larger.

Consider the following binary hypothesis test:

$$\begin{aligned} H_1 : Y_n &= A + w_n \\ H_0 : Y_n &= w_n \end{aligned}$$

where  $w_n \sim \mathcal{N}(0, \sigma^2)$  is an i.i.d. process and  $A$  denotes the strength of the signal. Let  $N$  be the number of i.i.d. observations taken to achieve a prescribed pair  $(P_d, P_f)$ . The ROC obtained by Neyman-Pearson test can be expressed in closed form as derived in [7] by the equation,

$$P_d = 1 - Q(d - Q^{-1}(P_f)) \quad (1)$$

where  $d^2 = \frac{NA^2}{\sigma^2}$  and  $Q(x)$  is the complementary cdf of the standard Gaussian distribution given by,

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt$$

Thus  $d$ , which is proportional to the signal to noise ratio (SNR) completely characterizes the ROC of this binary hypothesis test. It can be clearly seen that at a particular value of  $P_f$ , higher value of SNR results in larger  $P_d$ . Hence samples with a higher SNR will have an ROC curve that will be higher than those with lower SNR's.

### 5. PROBLEM FORMULATION

The binary hypothesis test for which we prove the claim made in section 3 is set up as follows:

$$\begin{aligned} H_1 : Y_n &= m_n + w_n \\ H_0 : Y_n &= w_n \end{aligned}$$

where  $w_n \sim \mathcal{N}(0, \sigma^2)$  and  $\{m_n\}_{n=1}^\infty$  is a deterministic sequence such that  $m_n > 0$ ,  $\forall n \geq 1$  and  $m_i \neq m_j$  when  $i \neq j$ ,  $\forall i, j \geq 1$ . The noise process  $w_n$  is assumed to be i.i.d. and thus the observations are independent but non-identically distributed under  $H_1$ . The ordering of the ROC curves will thus depend on the ordering of the means of the observations conditioned on  $H_1$ . Hence according to the claim made in

section 3, we would expect to make a decision sooner under  $H_1$  if we observe samples with higher means in the beginning as opposed to later.

Let  $l_i = \log\left(\frac{f_i(Y_i|H_1)}{f_i(Y_i|H_0)}\right)$  be the log-likelihood ratio for observation  $Y_i$ . Since the observations  $Y_i$  are independent, the log-likelihood ratios  $l_i$  at each instant are independent of each other as well. Let the cumulative log-likelihood ratio up to time instant  $n$  be denoted by  $L_n$ .

$$\text{Hence we have } L_n = \log\left(\prod_{i=1}^n \frac{f_i(Y_i|H_1)}{f_i(Y_i|H_0)}\right) = \sum_{i=1}^n l_i$$

Now,  $l_i = \frac{2m_i Y_i - m_i^2}{2\sigma^2}$ . Hence we have  $E[l_i|H_1] = \frac{m_i^2}{2\sigma^2}$ ,  $E[l_i|H_0] = -\frac{m_i^2}{2\sigma^2}$  and  $\text{Var}[l_i|H_1] = \text{Var}[l_i|H_0] = \frac{m_i^2}{\sigma^2}$ . Thus the distribution of  $L_n$  under both hypotheses is given by,  $L_n|H_1 \sim \mathcal{N}\left(\frac{\sum_{i=1}^n m_i^2}{2\sigma^2}, \frac{\sum_{i=1}^n m_i^2}{\sigma^2}\right)$  and  $L_n|H_0 \sim \mathcal{N}\left(-\frac{\sum_{i=1}^n m_i^2}{2\sigma^2}, \frac{\sum_{i=1}^n m_i^2}{\sigma^2}\right)$ . Note that  $E[l_i|H_1]$  is the KL divergence between hypotheses  $H_1$  and  $H_0$  of the  $i^{\text{th}}$  sample.

Let  $a_n = \log(A_n)$  and  $b_n = \log(B_n)$  where  $A_n$  and  $B_n$  are the time varying thresholds of the GSPRT. Let the prescribed probability of false alarm and miss for the test be denoted by  $P_f$  and  $P_m$  respectively. We will denote the probability of making an error conditioned on  $H_0$  and  $H_1$  at stopping time  $n$  by  $P_f^{(n)}$  and  $P_m^{(n)}$  respectively. Hence we have  $P_f^{(n)} = P(a_1 < L_1 < b_1, \dots, a_{n-1} < L_{n-1} < b_{n-1}, L_n \geq b_n|H_0)$  and  $P_m^{(n)} = P(a_1 < L_1 < b_1, \dots, a_{n-1} < L_{n-1} < b_{n-1}, L_n \leq a_n|H_1)$ . We thus have the following inequalities,

$$P_f^{(n)} \leq P(L_n \geq b_n|H_0), \quad \forall n \geq 1 \quad (2)$$

$$P_m^{(n)} \leq P(L_n \leq a_n|H_1), \quad \forall n \geq 1 \quad (3)$$

Equality is achieved in (2) and (3) for  $n = 1$  and for all  $n > 1$ , we have a strict inequality. The error probabilities at each stopping time can be made lesser than the prescribed values  $P_f$  and  $P_m$  by setting

$$P_f = \int_{b_n}^{\infty} g(L_i|H_0)dL_i, \quad \forall n \geq 1 \quad (4)$$

$$P_m = \int_{-\infty}^{a_n} g(L_i|H_1)dL_i, \quad \forall n \geq 1 \quad (5)$$

where  $g(L_i|H_j)$  is the pdf of the gaussian random variable  $L_i$  under hypothesis  $H_j$  for  $j = 0, 1$ . Using (4) and (5), the thresholds can be found using the following equations:

$$b_n = \frac{Q^{-1}(P_f)\sqrt{\sum_{i=1}^n m_i^2} - \sum_{i=1}^n m_i^2}{2\sigma^2} \quad (6)$$

$$a_n = \frac{Q^{-1}(1 - P_m)\sqrt{\sum_{i=1}^n m_i^2} + \sum_{i=1}^n m_i^2}{2\sigma^2} \quad (7)$$

As  $a_n < 0 < b_n$ ,  $\forall n \in \mathbb{Z}^+$ , the series  $\sum_{n=1}^{\infty} m_n^2$  should be convergent else as  $n \rightarrow \infty$ ,  $b_n \rightarrow -\infty$  and  $a_n \rightarrow \infty$ . Though these time varying thresholds reduce the ASN (while ensuring that the error probabilities are below  $P_m$  and  $P_f$ ) as compared to that using Wald's constant thresholds, we don't necessarily know if these thresholds are optimal, i.e. we may be able to find another sequence of varying thresholds which reduce the ASN even further.

Let  $h_i(t|H_j)$ , for  $j = 0, 1$  and  $i \geq 1$  denote the pdf of  $l_i$ . Let  $P_n(t|H_j) = P(L_n < t|a_1 < L_1 < b_1, \dots, a_{n-1} < L_{n-1} < b_{n-1}, H_j)$  be the cumulative distribution function (cdf) of  $L_n$  conditioned on the past observations up to  $n$  and hypothesis  $H_j$  for  $j = 0, 1$ . Let  $p_n(t|H_j)$  denote the pdf of  $L_n$  conditioned on the past which is obtained by taking the derivative of the cdf. To calculate  $p_n(t|H_j)$  for each  $n > 1$ , we can use the following recursion, for  $j = 0, 1$ ,

$$p_1(t|H_j) = h_1(t|H_j) \quad (8)$$

$$p_n(t|H_j) = \int_{a_{n-1}}^{b_{n-1}} p_{n-1}(u|H_j)h_n(t-u|H_j)du \quad (9)$$

Using (8) and (9), the expected number of samples  $E[N|H_j]$  for  $j = 0, 1$  can be calculated using the following equation:

$$E[N|H_j] = \sum_{n=1}^{\infty} n \left( \int_{-\infty}^{a_n} p_n(t|H_j)dt + \int_{b_n}^{\infty} p_n(t|H_j)dt \right) \quad (10)$$

As given in [8], the stopping time  $n$  is finite if  $P(n < \infty|H_j) = 1, j = 0, 1$ , i.e.

$$\lim_{n \rightarrow \infty} P(a_n < L_n < b_n|H_j) = 0, \quad j = 0, 1 \quad (11)$$

The GSPRT is truncated at a large value of  $n$  and thus we try to achieve (11) for a large but finite value of  $n$ . For the thresholds calculated using (6) and (7),

$$P(a_n < L_n < b_n|H_1) = 1 - P_m - Q\left(Q^{-1}(P_f) - \frac{\sqrt{c_n}}{\sigma}\right) \quad (12)$$

where  $c_n = \sum_{i=1}^n m_i^2$ . Hence to satisfy condition (11), for large  $n$  we need to set,

$$\frac{\sqrt{c_n}}{\sigma} = Q^{-1}(P_f) - Q^{-1}(1 - P_m) \quad (13)$$

From (6), (7) and (13), we get,  $b_n = a_n = \frac{(p^2 - q^2)}{2}$  as  $n$  becomes large, where  $p = Q^{-1}(P_f)$  and  $q = Q^{-1}(1 - P_m)$ . But the thresholds should satisfy  $b_n > 0$  and  $a_n < 0, \forall n \geq 1$ . Thus for a small  $\epsilon > 0$ , we need the following condition to hold true for large but finite  $n$

$$\frac{\sqrt{c_n}}{\sigma} = Q^{-1}(P_f) - Q^{-1}(1 - P_m) - \epsilon \quad (14)$$

The value of  $\epsilon$  should be chosen such that the upper and lower thresholds are positive and negative respectively for all  $n$  and the probability that the cumulative LLR up to large but finite  $n$  stays between the thresholds is as close to 0 as possible.

For SPRT with fixed thresholds  $b = \ln(B)$  and  $a = \ln(A)$  we have,

$$P(a < L_n < b|H_1) = Q\left(\frac{a - \mu_n}{\lambda_n}\right) - Q\left(\frac{b - \mu_n}{\lambda_n}\right) \quad (15)$$

where  $\mu_n = \frac{c_n}{2\sigma^2}$  and  $\lambda_n = \sqrt{\frac{c_n}{\sigma^2}}$ . Thus to ensure that the stopping time is finite with probability close to 1, we see from (11) and (15) that  $\mu_n \gg b$  should be satisfied as  $n \rightarrow \infty$ . Now if the series  $\sum_{n=1}^{\infty} m_n^2$  is divergent for a particular ordering, it will be divergent for all orderings since each term in the sequence  $\{m_n^2\}_{n=1}^{\infty}$  is positive. Hence for all orderings we have,

$$\begin{aligned} \lim_{n \rightarrow \infty} P(a < L_n < b|H_1) &= \lim_{n \rightarrow \infty} Q\left(\frac{a\sigma}{\sqrt{c_n}} - \frac{\sqrt{c_n}}{2\sigma}\right) \\ &\quad - \lim_{n \rightarrow \infty} Q\left(\frac{b\sigma}{\sqrt{c_n}} - \frac{\sqrt{c_n}}{2\sigma}\right) \\ &= Q(-\infty) - Q(-\infty) \\ &= 0 \end{aligned} \quad (16)$$

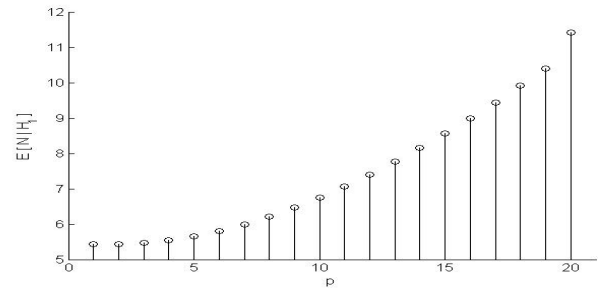
Hence we have finiteness of stopping time with probability 1 for all orderings. An example of such a series is  $m_n = \sqrt{d + s_n}$ ,  $\forall n \geq 1$ , where  $d > 0$  is a constant,  $s_n > 0$ ,  $\forall n \geq 1$  and  $\sum_{n=1}^{\infty} s_n$  is a convergent series. Note that Lyapunov's condition is satisfied for such a series for  $\delta = 2$  and thus by Lyapunov's variant of the central limit theorem (CLT) [9], the cumulative LLR is asymptotically normal. If the series is not convergent as in the case of deriving varying thresholds, the cumulative LLR is not asymptotically normal.

The ordering of observations based on their ROC curves is same as that based on the KL divergence between hypotheses  $H_1$  and  $H_0$  since they both are proportional to the SNR of each observation. In general it is more feasible to get closed form expressions for the KL divergence rather than the ROC curves. In the future we will derive conditions under which the claim can be proved for the case in which the observations are drawn from a non-Gaussian probability distribution. In this generalized case, the orderings will be defined as orderings of the KL divergence between the two hypotheses of each observation.

## 6. RESULTS AND DISCUSSION

In this section we present numerical results for the problem formulated in the previous section. Consider the sequence  $\{m_n\}_{n=1}^{\infty}$  to be a monotonically decreasing sequence with a geometric decay, i.e.  $m_n = r^{n-1}$ ,  $n \geq 1$ , where  $0 < r < 1$  is the rate of decrease of the sequence. For convenience we

call this sequence as the *reference* sequence. Thus the different orderings of observations can be seen as permutations of elements of the reference sequence. The pdf of the stopping time and the ASN for different orderings is calculated using (8), (9) and (10) using MATLAB. In particular the integrals in (9) and (10) are computed using the trapezoidal rule with 1000 points in each of the intervals  $(-\infty, a_n]$ ,  $(a_n, b_n)$  and  $[b_n, \infty)$  for all  $n \geq 1$ . To satisfy condition (14), noise power  $\sigma^2 = 0.4$  and  $r = 0.96$  is chosen for  $P_m = P_f = 10^{-3}$ . As there are millions of orderings possible we cannot possibly compute the ASN for all of them. In addition it turns out that the ASN does not change much for small offsets in the orderings. Thus we consider the interesting case of  $p$ -reverse orderings for  $p \geq 1$ . A  $p$ -reverse ordering  $\{m'_n\}_{n=1}^{\infty}$  is defined such that  $m'_n = m_{p-n+1}$  for  $1 \leq n \leq p$  and  $m'_n = m_n$ ,  $\forall n > p$ . Note that the 1-reverse sequence is same as the *reference* sequence.

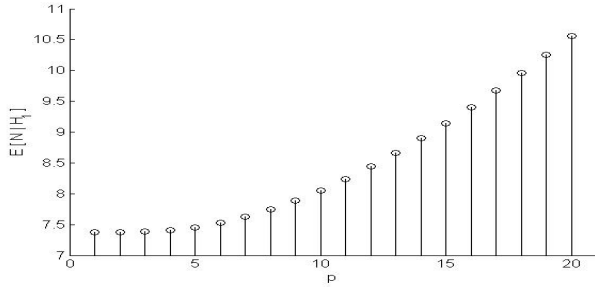


**Fig. 2.** Average sample size of GSPRT under hypothesis  $H_1$  for different  $p$ -reverse orderings

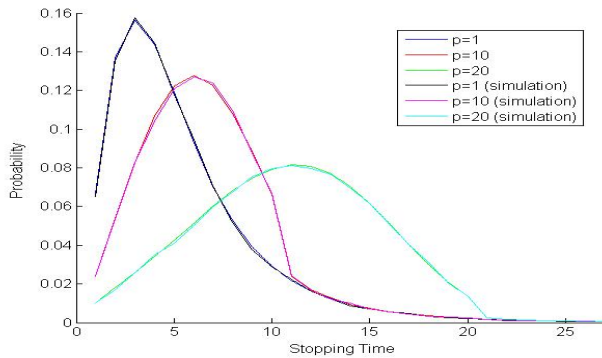
It can be seen from figure 2 that the ASN for the reference sequence is minimum and is equal to 5.4277. The second least ASN (equal to 5.4376) among all orders is obtained for 2-reverse ordering. The ASN increases monotonically when  $p$  increases. But a significant increase i.e. at least 2 samples more than the minimum ASN, only occurs when  $p > 10$ . The rate of change of ASN among orderings for smaller values of  $p$  is less due to the slow rate of change of the sequence  $m_n$  since  $r$  is chosen to be close to 1. If the rate of decrease of the reference sequence is increased, i.e. if  $r$  is decreased to enlarge the difference in the ASN among orderings, the probability of termination of the truncated test will decrease and thus there is a trade-off between increasing probability of termination of the GSPRT and increasing the change in ASN for different orderings.

In order to ensure that the stopping time is finite with probability almost equal to 1 while using SPRT, we set  $r = 0.98$  and retain the same values for other parameters. Figure 3 shows similar trend in the increase of ASN with  $p$  using SPRT as compared to that using GSPRT. It can be seen that even for a larger value of  $r$ , the ASN using SPRT is more than that using GSPRT for all orderings.

Figure 4 shows the pdf of the stopping time for the refer-



**Fig. 3.** Average sample size of SPRT under hypothesis  $H_1$  for different  $p$  – reverse orderings



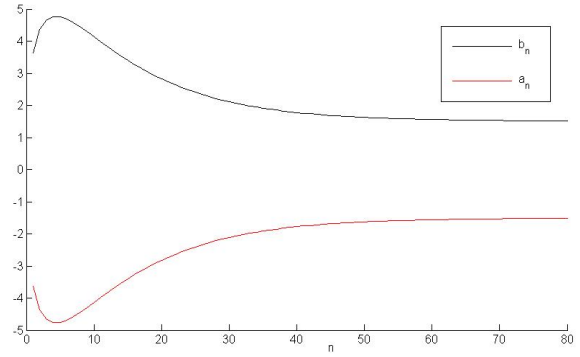
**Fig. 4.** Probability density function of stopping time for 3 different orderings obtained by analytical calculations and simulations

ence sequence and  $p$  – reverse orderings when  $p = 10$  and  $p = 20$  using GSPRT. It can be observed that the pdf for all 3 orderings obtained by computing the recursive integrals given by (10) and (11) is almost the same as that obtained by performing Monte Carlo simulations with 50,000 iterations. For higher values of  $p$ , the pdf is more spread out due to increased variance and its concentration around the larger values of stopping time is also more.

A plot of the lower ( $a_n$ ) and upper ( $b_n$ ) thresholds for the reference sequence is shown in figure 5. The thresholds will be different for different orderings but for every ordering they will converge to the same value at a large value of  $n$ .

## 7. CONCLUSION AND FUTURE WORK

In this paper we provide an analysis of the influence of ordering of non-identical and independent observations on the expected sample size in a sequential binary hypothesis testing problem. When the observations are independent and non-identically distributed Gaussian random variables, we showed that among all orderings, ASN is minimized when the samples are observed in a order such that the area under the ROC curve of each sample is monotonically decreasing. As part



**Fig. 5.** Upper and Lower thresholds for the reference sequence

of future work we will consider the case in which the observations are non-Gaussian random variables and derive conditions under which our claim remains true. We also want to extend these results to sequential decision making by human beings when cognitive biases interfere with the decision procedure.

## 8. REFERENCES

- [1] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.
- [2] J. Cochlar and I. Vrana, “On the optimum sequential test of two hypotheses for statistically dependent observations,” *Kybernetika*, vol. 14, no.1, pp. 57–69, 1978.
- [3] Y. Liu and S. D. Blostein, “Optimality of the Sequential Probability Ratio Test for nonstationary observations,” *IEEE Trans. Info. Theory*, vol. 38, no.1, pp. 177–182, 1992.
- [4] R. M. Phatarfod, “Sequential analysis of dependent observations. I,” *Biometrika*, vol. 52, no.1/2, pp. 157–165, 1965.
- [5] V. P. Dragalin, A. G. Tartakovsky and V. V. Veeravalli, “Multihypothesis Sequential Probability Ratio Tests. Part I: Asymptotic Optimality,” *IEEE Trans. Info. Theory*, vol. 45, no.7, pp. 2448–2461, 1999.
- [6] A. G. Tartakovsky, X. R. Li and G. Yaralov, “Sequential detection of targets in multichannel systems,” *IEEE Trans. Info. Theory*, vol. 49, no.2, pp. 425–445, 2003.
- [7] B. C. Levy, *Principles of Signal Detection and parameter estimation*. New York, NY: Springer, 2008.
- [8] B. K. Ghosh, *Sequential Tests of Statistical Hypotheses*. Cambridge, MA: Addison-Wesley, 1970.
- [9] P. Billingsley, *Probability and Measure*. New York, NY: John Wiley and Sons, 1986.