REAL-TIME DATA SELECTION AND ORDERING FOR COGNITIVE BIAS MITIGATION

Sara Mourad and Ahmed Tewfik

Department of Electrical and Computer Engineering
The University of Texas at Austin

***Abstract*—We consider the problem of selecting and ordering a subset of N' out of N observations to be presented to a human being in the context of a binary hypothesis testing problem. We restrict our attention to i.i.d. Gaussian observations. We propose an extension of the approximate subset sum algorithm, and show that it can be used to solve the problem with polynomial complexity. Furthermore, we show that the solution yields near optimal detection performance when compared to the case where all N observations are optimally processed.**

***Index* —** Cognitive biases, Subset selection, Neyman-Pearson, Likelihood, Subset sum.

I. INTRODUCTION

Statistical inference methods allow to optimally draw conclusions from data. For these methods, the larger the number of observations, the lower the probability of errors. However, humans do not process information optimally due to cognitive biases they are subject to. For example, humans might become insensitive to new data after some point. Furthermore, the order in which information is presented to a person can affect their decision. Hence with the increase in the number of data sources available to each individual, it is becoming critical to figure out what subset of information to emphasize to a human being, and in which order, to optimize their decision-making.

Sensor selection has been widely studied in the literature, with the aim of cutting the number of sensors and saving resources. In [10], a heuristic is proposed to solve the problem of selecting k out of m sensors based on convex relaxation. In [15], the sensor selection problem is cast as the maximization of a submodular function over uniform matroids, and solved using a greedy algorithm with performance guarantees. However, in our case, we would like to leverage all the available sensors, and select in real time an optimal subset of observations to show to a human subject.

The way humans process information has also been widely studied, and modeled based on statistical inference models. For hypothesis testing problems, the Neyman-Pearson (NP) test, which is a likelihood ratio test, is the most powerful test to maximize the probability of detection under some constraint on the probability of false alarm, when the number of observations is determined in advance [11, 16]. If the number of observations is not predetermined, the sequential probability ratio test (SPRT) is the optimal test for i.i.d. observations in terms of the expected number of observations needed to make a decision within the required probabilities of missed detection and of false alarm [11, 16].

These models have been adapted to humans in different ways in order to incorporate cognitive biases. Cognitive biases are heuristics that hinder humans from making rational decisions [4]. As an example, we cite the anchoring bias, modeled in [5] and [8], where humans are influenced by starting points or initial beliefs. Other biases are the confirmation bias modeled in [3] where humans tend to emphasize data confirming their beliefs, and disregard data contradicting their beliefs. In [9], Hogarth and Einhorn closely study the order effects in the update of belief, and mention two main biases: the primacy effect [2, 14] where humans emphasize the first set of information and the recency effect where humans emphasize the last set of information. They point out that as the number of pieces of information increase, subjects can get tired if asked to process all the data, and beliefs become less sensitive to new observations; this is the primacy effect [9]. Hence it is important to study which subset of the observations should be presented to humans, and in which order, in order to maximize their performance given these biases.

The problem of optimal ordering of observations for humans have been considered in [1], where the bias is modeled by modifying the values of the thresholds in the SPRT test. However, the ordering is done based on the statistics of the data and not on their real time values, and bias is modeled based on the SPRT test, differently than in this paper. In [13], a heuristic is proposed for optimally ordering data to humans using the bias model of this paper, and giving good performance, but not near optimal performance.

In this paper, we study the problem of selecting and ordering information to show to humans in real time in order to achieve near optimal performance. In Sec. II, the main problem is defined as well as the approach to solving it, when the observations are Gaussian i.i.d. In Sec. III, we propose and describe an algorithm for the subset selection and ordering based on the approximate subset sum algorithm. We then verify in Sec. IV the performance of the proposed algorithm through simulations in case of biases.

II. PROBLEM STATEMENT

In the context of binary hypothesis testing, consider 2 agents Alice and Bob, where Alice is unbiased and Bob is affected by cognitive biases. Alice has access to N observations, and can only show Bob N' out of the N observations. The problem is finding out which N' observations to present to Bob, and in which order, such that Bob makes an optimal decision, i.e., his probability of detection is maximized for any constraint on the probability of false alarm that he adopts. In other words, such that Bob's probability of detection for a given upper bound on the false alarm rate is as close as possible to Alice's optimal probability of detection for the same upper bound on the probability of false alarm. Alice and Bob use the NP test which is the optimal test for maximizing the probability of detection given an upper bound on the probability of false alarm. Alice has N i.i.d. Gaussian observations, and Bob considers that he has N' i.i.d. Gaussian observations. (Note that by selecting a subset of observations, the statistics of the subset change. However, Bob doesn't know that and treats the new observations as independent observations with the same initial statistics). We consider that the observations y_i seen by the agents are distributed as follows under both hypotheses:

$$\begin{aligned} H_0 &: Y_i = V_i,\ \forall\ 1 \leq i \leq N \\ H_1 &: Y_i = m + V_i,\ \forall\ 1 \leq i \leq N, \end{aligned} \tag{1}$$

where V_i are i.i.d. with a Gaussian distribution $\mathbf{N}(0, \sigma^2)$ and m is the difference in the means under the two hypotheses. Let l_i denote the log-likelihood ratio for observation y_i, $l_i = \log(\frac{f(y_i|H_1)}{f(y_i|H_0)}) = (2my_i - m^2)/2\sigma^2$. Note that, under the two hypotheses, l_i follows Gaussian distributions with variances $\text{Var}[l_i|H_0] = \text{Var}[l_i|H_1] = m^2/\sigma^2$. Moreover, we have $\mathbb{E}[l_i|H_0] = -m^2/(2\sigma^2)$ and $\mathbb{E}[l_i|H_1] = m^2/(2\sigma^2)$. Let L_i denote the cumulative log-likelihood ratio up to the ith observation y_i, i.e.,

$$L_i = \log(\prod_{k=1}^{i} \frac{f(y_k|H_1)}{f(y_k|H_0)}) = \sum_{k=1}^{i} \log(\frac{f(y_k|H_1)}{f(y_k|H_0)}) = \sum_{k=1}^{i} l_k.$$

For Alice, under the two hypotheses, L_N follows Gaussian distributions with variances $\text{Var}[L_N|H_0] = \text{Var}[L_N|H_1] = Nm^2/\sigma^2$ and means $\mathbb{E}[L_N|H_0] = -Nm^2/(2\sigma^2)$ and $\mathbb{E}[L_N|H_1] = Nm^2/(2\sigma^2)$. Similarly for Bob, $L_{N'}$ follows Gaussian distributions with variances $\text{Var}[L_{N'}|H_0] = \text{Var}[L_{N'}|H_1] = N'm^2/\sigma^2$ and means $\mathbb{E}[L_{N'}|H_0] = -N'm^2/(2\sigma^2)$ and $\mathbb{E}[L_{N'}|H_1] = N'm^2/(2\sigma^2)$. In an NP test, for a given upper bound α on the probability of false alarm, Alice and Bob use the thresholds $\lambda_A = \frac{\sqrt{N}m}{\sigma}Q^{-1}(\alpha) - \frac{Nm^2}{2\sigma^2}$ and $\lambda_B = \frac{\sqrt{N'}m}{\sigma}Q^{-1}(\alpha) - \frac{N'm^2}{2\sigma^2}$, respectively (where Q is the standard Q-function). We assume that α is small enough not to fall in the trivial case where Alice and Bob can use threshold of zero.

A. Biased information processing model

Bob may process the information in a biased way due to cognitive biases. In other words, he updates his cumulative log-likelihood ratio according to a model proposed in [13], which is based on the belief adjustment framework by Hogarth and Einhorn [9]. These biases are quantified by the adjustment weight for the new observation as such:

$$L_i = L_{i-1} + w_i l_i. \tag{2}$$

where w_i is the adjustment weight that the subject gives to the new observation due to biases. The model parameters are learned through an adaptive learning process, that may involve asking the subject to do specific tasks. However, the parameter estimation problem is beyond the scope of the paper, and is discussed elsewhere. See also [9], [7].

This model allows representing many cognitive biases that humans could be subject to, as we will show in the following sections. From now on, $L_{N'}$ denotes the cumulative log-likelihood ratio of Bob, resulting from a biased updating as described in (2). Since Alice represents the optimal performance we are trying to match, her cumulative log-likelihood ratio L_N is unbiased.

B. Problem formulation

For the NP test thresholds λ_A and λ_B, let $P_d(\lambda_A)$ ($P_d(\lambda_B)$) and $P_f(\lambda_A)$ ($P_f(\lambda_B)$) denote the probabilities of detection and false alarm of Alice (Bob), respectively. Note that for same upper bound α on $P_f(\lambda_A)$ and $P_f(\lambda_B)$, the best Bob can hope for is $P_d(\lambda_B) = P_d(\lambda_A)$. Thus, the problem boils down to the following optimization problem:

$$\underset{\mathcal{K}\subseteq[N]:|\mathcal{K}|=N'}{\text{argmin}} \int_{L_N=\lambda_A}^{+\infty} p(L_N|H_1) - \int_{L_{N'}=\lambda_B}^{+\infty} \hat{p}(L_{N'}|H_1)$$
$$\text{subject to} \quad \int_{L_{N'}=\lambda_B}^{+\infty} \hat{p}(L_{N'}|H_0) \leq \alpha.$$

Here, $\hat{p}$ denotes the altered statistics of the sum of the subset $\mathcal{K}$ of N' observations after selecting them from the complete set of N observations. We note that even if Bob uses λ_B such that his probability of false alarm is less than α (he considers the observations to be i.i.d.), this doesn't ensure that the actual probability of false alarm, based on the true statistics $\hat{p}$, is less than α. Hence the added constraint.

Writing out the problem relative to the performance metrics of Alice would allow us to tackle the problem without figuring out $\hat{p}$. In order to make $P_d(\lambda_B)$ as close as possible to $P_d(\lambda_A)$ while keeping $P_f(\lambda_B) \leq \alpha$, we can make $P_d(\lambda_B)$ as close as possible to $P_d(\lambda_A)$, and make $P_f(\lambda_B)$ as close as possible to $P_f(\lambda_A)$, which is guaranteed to be less than α. This will be satisfied by making the event $\{L_{N'} \geq \lambda_B|H_1\}$ and the event $\{L_N \geq \lambda_A|H_1\}$ statistically equivalent, and by also making the event $\{L_{N'} \geq \lambda_B|H_0\}$ and the event $\{L_N \geq \lambda_A|H_0\}$ statistically equivalent (Two events are equivalent if the occurrence of one event implies the occurrence of the other and vice versa).

Let $\tilde{\lambda}_A = \lambda_A - \mathbb{E}[L_N|H_0] = \lambda_A + \frac{Nm^2}{(2\sigma^2)} = \frac{\sqrt{N}m}{(\sigma)}Q^{-1}(\alpha)$ and $\tilde{\lambda}_B = \lambda_B - \mathbb{E}[L_{N'}|H_0] = \lambda_B + \frac{N'm^2}{(2\sigma^2)} = \frac{\sqrt{N'}m}{(\sigma)}Q^{-1}(\alpha)$. We notice that $\tilde{\lambda}_B = \sqrt{\frac{N'}{N}}\tilde{\lambda}_A$.

Now if the subset $\mathcal{K}$ of N' elements is chosen such that $L_{N'} - \mathbb{E}[L_{N'}|H_0] = \frac{\sqrt{N'}(L_N - \mathbb{E}[L_N|H_0])}{\sqrt{N}}$ (i.e. $L_{N'} = \frac{\sqrt{N'}(L_N - \mathbb{E}[L_N|H_0])}{\sqrt{N}} + \mathbb{E}[L_{N'}|H_0]$), then the two events $\{L_N - \mathbb{E}[L_N|H_0] \geq \tilde{\lambda}_A|H_i\}$ and $\{L_{N'} - \mathbb{E}[L_{N'}|H_0] \geq \tilde{\lambda}_B|H_i\}$ become equivalent for any value of $\tilde{\lambda}_A$ since $\tilde{\lambda}_B = \sqrt{\frac{N'}{N}}\tilde{\lambda}_A$. Therefore, the two events $\{L_N \geq \lambda_A|H_i\}$ and $\{L_{N'} \geq \lambda_B|H_i\}$ are now equivalent, and thus $P_d(\lambda_A) = P_d(\lambda_B)$ and $P_f(\lambda_A) = P_f(\lambda_B)$. Hence we are interested in finding a subset $\mathcal{K}$ of N' observations such that $L_{N'}$ sums up to the target $T = \frac{\sqrt{N'}(L_N - \mathbb{E}[L_N|H_0])}{\sqrt{N}} + \mathbb{E}[L_{N'}|H_0]$. If there is no set $\mathcal{K}$ such that $L_{N'}$ equals T exactly, then the closer $L_{N'}$ to T, the more the two events coincide, and the smaller the difference between $P_d(\lambda_A)$ and $P_d(\lambda_B)$, as well as between $P_f(\lambda_A)$ and $P_f(\lambda_B)$.

III. APPROXIMATE SOLUTION

A. Approximate solution performance guarantee

Given that finding a subset $\mathcal{K}$ of N' out of the N observations that give a cumulative log-likelihood ratio $L_{N'}$ equals to the target T is not always feasible, we need to find an approximate solution resulting in $L_{N'}$ as close to T as possible.

Lemma .1. *Let $L_{N'}$ be such that $T - \frac{\delta}{2} \leq L_{N'} \leq T + \frac{\delta}{2}$, then $Q(\frac{\lambda_A - \mathbb{E}[L_N|H_1] + \frac{\sqrt{N}\frac{\delta}{2}}{\sqrt{N'}}}{\text{Var}(L_N)}) \leq P_d(\lambda_B)$ and $P_f(\lambda_B) \leq$*

$Q(\frac{\lambda_A-\mathbb{E}[L_N|H_0]-\frac{\sqrt{N}\frac{\delta}{2}}{\sqrt{N'}}}{\text{Var}(L_N)})$

Proof.

$$T-\frac{\delta}{2} \leq L_{N'} \implies P_d(\lambda_B) \geq \mathbb{P}(T-\frac{\delta}{2} \geq \lambda_B|H_1) \quad (3)$$

$$\implies P_d(\lambda_B) \geq \mathbb{P}(L_N \geq \lambda_A + \frac{\sqrt{N}\frac{\delta}{2}}{\sqrt{N'}}|H_1) \quad (4)$$

$$\implies P_d(\lambda_B) \geq Q(\frac{\lambda_A - \mathbb{E}[L_N|H_1] + \frac{\sqrt{N}\frac{\delta}{2}}{\sqrt{N'}}}{\text{Var}(L_N)}) \quad (5)$$

where (3) follows since $P_d(\lambda_B) = P(L_{N'} \geq \lambda_B|H_1)$, and (4) follows since $\lambda_B - E(L_{N'}|H_0) = (\lambda_A - E(L_N|H_0))(\frac{\sqrt{N'}}{\sqrt{N}})$ and by replacing T with its value. We prove similarly the upper bound of $P_f(\lambda_B)$. □

B. Approximate subset sum algorithm

In order to find the set $\mathcal{K}$ of N' observations which cumulative log-likelihood ratio $L_{N'}$ is the closest to the target T, an exhaustive solution would be to try all the N' combinations of observations, and chose the closest cumulative log-likelihood ratio to the target T. However, this solution is exponential in complexity if we let N' and N scale.

Another solution would be to use an algorithm based on the approximate-subset-sum algorithm which is a FPTAS (fully polynomial-time approximation scheme) [12] for the subset sum problem, where the set of elements x_i's corresponds to the set of log-likelihood ratios l_i's from which we are trying to select a subset.

In the classical subset sum problem $(\mathcal{S}, t)$ [6], $\mathcal{S} = \{x_1, x_2, ..., x_n\}$ denotes a set of positive integers and a positive integer t represents the target. The decision problem asks for a subset of $\mathcal{S}$ whose sum is as large as possible, but not larger than the target t. This problem is NP-complete. The approximate subset sum algorithm $(\mathcal{S}, t, \epsilon)$ [6], for $0 < \epsilon < 1$, is a FPTAS for the subset sum problem that approximates the optimal solution to within a ratio bound of $1 + \epsilon$, and works as follows:

Algorithm 1 Approximate subset algorithm $(\mathcal{S}, t, \epsilon)$

1: $n \leftarrow |\mathcal{S}|$
2: $R_0 \leftarrow \{0\}$
3: **for** $i \leftarrow 1$ to n **do**
4: $R_i \leftarrow$ MergeLists(R_{i-1}, R_{i-1} + x_i)
5: $R_i \leftarrow$ Trim(R_i , $\epsilon/2n$)
6: Remove from R_i every element greater than t
7: **end for**
8: **return** the largest element in R_n

Trimming a list R_i by $\beta = \epsilon/2n$ means removing as many elements as possible such that every element that is removed is approximated by some remaining element in the list. More precisely, given $0 < \beta < 1$, element z approximates element y if $\frac{y}{1+\beta} \leq z \leq y$, $\beta = \epsilon/2n$ in this algorithm.

The solution returned is within a factor $1+\epsilon$ of the optimal solution. The running time is polynomial in both N and $1/\epsilon$.

C. Proposed algorithm based on the approximate subset sum algorithm

When the adjustment weight w_i depends for example on the position of l_i (as for the anchoring bias) or on the cumulative log-likelihood ratio L_{i-1}, then permutations of the elements in the same subset lead to different sums, and thus the ordering of observations matters. The subset sum algorithm would need to be modified to account for all permutations of the same subset (Notice line 3 of algorithm 1 which restricts an order of appearance of the elements in the subset sum). A solution to the limitation of the approximate subset algorithm would be to redefine the set of initial elements $\mathcal{S}$ as $\mathcal{S}_1$ where $\mathcal{S}_1$ is a multiset consisting of the set $\mathcal{S}$ repeated N' times: $\mathcal{S}_1 = \{\mathcal{S}, \mathcal{S}, ..., \mathcal{S}\}$ with $|\mathcal{S}_1| = NN'$. Now each element in $\mathcal{S}_1$ is tagged by its index in the initial set $\mathcal{S}$ to avoid picking an element twice. The modified approximate subset sum algorithm designed in this section is executed on $\mathcal{S}_1$, but now, at each time a new element of $\mathcal{S}_1$ is added to list of subset sums (in line 4 of algorithm 2), we make sure that the existing subset sum value does not include an element with the same index as the new element.

This problem is thus approximately solved by modifying the approximate subset sum algorithm used in Sec. III-B as described below.

Algorithm 2 Modified approximate subset algorithm ($\mathcal{S}_1$,T,ϵ)

1: $n \leftarrow |\mathcal{S}_1|$
2: $R_0 \leftarrow \{0\}$
3: $G_0 \leftarrow \{0\}$
4: **for** $i \leftarrow 1$ to n **do**
5: $R_i \leftarrow$ MergeLists(R_{i-1}, R_{i-1} + $w_j l_i$)
6: $G_i \leftarrow$ MergeLists(G_{i-1}, G_{i-1} + 1)
7: (R_i,G_i) $\leftarrow$ Trim(R_i, G_i, N', $\epsilon/2n$)
8: **end for**
9: **return** the closest element in R_n to T with size in G_n less than or equal to N'

The list G_i keeps track of the corresponding sizes of the elements in R_i, and its use is explained in the next section. The modification for accounting for biases is in the Merge list step, where when adding the log-likelihood ratio l_i, we incorporate its corresponding adjustment weight w_j. We change the subscript from i to j in w_j to emphasize that the adjustment weight at step i need not only depend on l_i.

D. Taking into account other variations from the approximate subset sum algorithm

There are several differences of the proposed algorithm from the usual approximate subset sum algorithm described earlier in III-B. First, the values of x_i (corresponding to l_i's) as well as the target T could be positive or negative.

Second, the subset size is restricted to N' or less elements. To deal with this, the algorithm keeps track of the size of every element in the list R_i in another list G_i. By size of an entry we mean the number of observations which sum resulted in this entry. We then modify the trimming function so that it keeps the elements with the smallest size in the trimmed list. This is briefly explained as follows. In the traditional trimming function, we add gradually to the output list elements z, such

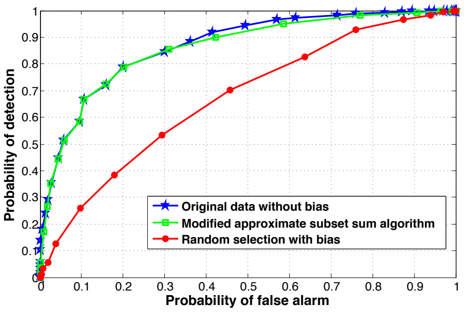

Figure 1: ROC curves under different selection of Gaussian i.i.d. observations for anchoring bias with small F/N'

that an element z_i is added whenever it is not approximated by the latest element in the output list z_{i-1}. In modified trimming function, we keep generating the list $\{z_1, z_2, ...\}$ throughout the algorithm. However, instead of adding to the output list the element z_i, we add the element u_i, which is the element with the smallest size among the elements approximated by z_i (u_i could be z_i only if z_i has the smallest size). Keeping the subset sum with the smallest size ensures that, if an element in the final list is disregarded because its size exceeds N', then there was not another subset sum approximated by this element with N' or fewer elements that was trimmed during the process.
Finally, the returned subset sum by the algorithm is not restricted to be less than T and can exceed T, but it should be the closest element in R_n to the target T with a size less than or equal to N'. If the returned subset sum has strictly less than N' observations, the $\hat{N}$ missing observations should be replaced by dummy observations (with log-likelihood ratio very close to 0). If this is not desired, then the trimming algorithm should be modified to trim subset sums that don't correspond to exactly N' observations; The rest of the algorithm as well as its analysis remain the same.

E. Results from the analysis of the new algorithm

The analysis of the performance of the algorithm is done but not included in this paper. Let's denote by u^* the value returned by the algorithm, and by y^* be the optimal solution, i.e., the closest biased subset sum of length less than or equal to N' to the target T. Since $|T| - \delta/2 \leq |y^*| \leq |T| + \delta/2$, then the analysis done concluded that $(1-\epsilon)(|T| - \delta/2) \leq |u^*| \leq (1+\epsilon)(|T| + \delta/2)$ and thus that the solution provided by the approximation algorithm is guaranteed to be ϵ away from the target $+/-\delta/2$. As for the running time, an analysis was also done and it concluded that the time complexity is $\mathcal{O}(NN' \log N')$. The NN' factor is due to the fact that $\mathcal{S}_1$ is of size NN', and the polynomial complexity is due the fact that the values that the subset sums can take are statistically bounded. Also, the algorithm is polynomial in $1/\epsilon$.

IV. **RESULTS**

A. Anchoring bias and results

In the anchoring bias, the first F observations are given the highest weight, and the subsequent observations are given small adjustment weights. So w_j depends on the size of the subset sum that x_i is being added to. In Fig. 1, we plot the ROC curve (blue) corresponding the unbiased performance of Alice given the total set of N observations. We plot the corresponding ROC of Bob (green) under the anchoring bias such that for each threshold used by Alice, we compute the corresponding threshold and probability of detection of Bob when using the approximate subset sum algorithm. The results show that for the anchoring bias model, the algorithm described allows obtaining the ROC (green) nearly overlapping the optimal unbiased ROC of Alice in the region of interest (low P_f and high P_d). In the simulations, the value of N used is 10, and the value of N' is 4. We note that the lower the fraction F/N' of unaffected observations F to the number of selected observations N', the worse the performance of the subset sum approximation algorithm under bias, and thus the corresponding ROC won't completely overlap with the original ROC. This is illustrated in Fig. 1 where we chose N' to be much larger than F. We notice that this is mainly notable for low probability of detection and for large probability of false alarm. This is because in those cases, it is more crucial to have the subset sum match exactly the target. And the smaller the fraction F/N', the farther the subset sum returned by the algorithm from the target.

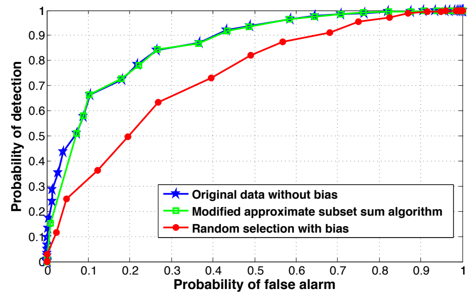

Figure 2: ROC curves under different selection of Gaussian i.i.d. observations with confirmation bias

B. Confirmation bias and results

The confirmation bias happens when humans emphasize data supporting one hypothesis, and neglect disconfirming data. Thus in the model, the adjustment weight w_j will depend on the value of the current log-likelihood ratio l_i; w_j will be hence denoted by w_i. We model the bias by assigning small adjustment weights w_i to l_i contradicting a given hypothesis, and w_i close to 1 to l_i supporting this hypothesis. Since w_i only depends on the current observation, different permutations of the same set of observations would lead to the same sum, and hence the approximate subset sum algorithm adapted to the bias can be used with the values l_i drawn from the set $\mathcal{S}$ and not $\mathcal{S}_1$ (And thus the running time is $\mathcal{O}(N \log(N'))$). We simulate the performance of the algorithm whenever the hypothesis supported by the subject is H_0, and as shown in Fig. 2, the algorithm gives an ROC (green) nearly overlapping the original ROC (blue). However, for low probability of detection, the green ROC is not completely matching the blue ROC. This is because whenever hypothesis H_1 is true, the bias leads to larger gap between the closest subset sum to the target and the target, and when the probability of detection is low, it is more important for the subset sum to be as close as possible to the target as discussed earlier.

REFERENCES

[1] N. Akl and A. Tewfik. Optimal information ordering in sequential detection problems with cognitive biases. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 1876–1880, May 2014.

[2] E. M. Altmann. Memory in chains: Modeling primacy and recency effects in memory for order, 2000.

[3] E. Bamber, R. J. Ramsay, and R. M. Tubbs. An examination of the descriptive validity of the belief-adjustment model and alternative attitudes to evidence in auditing. *Accounting, Organizations and Society*, 22(34):249 – 268, 1997.

[4] M. H. Birnbaum and B. A. Mellers. Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, 45(4):792, 1983.

[5] Y.-L. Chien, C. J. Huang, and D. Shaw. A general model of starting point bias in double-bounded dichotomous contingent valuation surveys. *Journal of Environmental Economics and Management*, 50(2):362–377, 2005.

[6] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms (3. ed.)*. MIT Press, 2009.

[7] G. Friesen and P. A. Weller. Quantifying cognitive biases in analyst earnings forecasts. *Journal of Financial Markets*, 9(4):333 – 365, 2006.

[8] J. A. Herriges and J. F. Shogren. Starting point bias in dichotomous choice valuation with follow-up questioning. *Journal of Environmental Economics and Management*, 30(1):112 – 131, 1996.

[9] R. M. Hogarth and H. J. Einhorn. Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24(1):1 – 55, 1992.

[10] S. Joshi and S. Boyd. Sensor selection via convex optimization. *Signal Processing, IEEE Transactions on*, 57(2):451–462, Feb 2009.

[11] B. C. Levy. *Principles of Signal Detection and Parameter Estimation*. Springer US, 2008.

[12] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, New York, NY, USA, 2005.

[13] S. Mourad and A. H. Tewfik. Cognitive biases in bayesian updating and optimal information sequencing. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2015, South Brisbane, Queensland, Australia, April 19-24, 2015*, pages 4095–4099. IEEE, 2015.

[14] R. Nisbett and L. Ross. *Human inference: strategies and shortcomings of social judgment*. Century psychology series. Prentice-Hall, 1980.

[15] M. Shamaiah, S. Banerjee, and H. Vikalo. Greedy sensor selection: Leveraging submodularity. In *Proceedings of the 49th IEEE Conference on Decision and Control, CDC 2010, December 15-17, 2010, Atlanta, Georgia, USA*, pages 2572–2577. IEEE, 2010.

[16] A. Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 06 1945.