THE DATA/IDENTITY TRADEOFF WITH CENSORED SENSORS

Zachariah Sutton, Peter Willett

ECE Dept., Univ. of Connecticut

Stefano Marano

DIEM, Univ. of Salerno

ABSTRACT

All practical sensing operations must work with quantized data. Along with measurements, each sensor is assumed to have some "label" value that is relevant to its stochastic measurement parameterization and must be communicated to the decision center. We are interested in cases that require very low communication cost, and thus require very "coarse" quantization of the measurements as well as the labels (2-bit values, for instance). Censoring is used to control the expected communication cost—each sensor decides locally whether or not to send its data to the decision center based on the value of its label as well as the value of its measurement. In this work we formalize the test statistic based on censored and quantized data.

Index Terms— Unlabeled, Sensor Network.

1. PROBLEM STATEMENT

Detection is framed as a binary hypothesis testing problem performed at a specific test "reference point" $\mathbf{x}^{\mathrm{t}}$ where the hypotheses being decided between are

$$\begin{aligned} &\mathcal{H}_0 : \text{There is no target object at point } \mathbf{x}^{\mathrm{t}} \\ &\mathcal{H}_1 : \text{There is a target object at point } \mathbf{x}^{\mathrm{t}}. \end{aligned} \tag{1}$$

For example, the target object could be a point target, and the reference point some point in a global Euclidean space that the tester wants to "look at". Alternatively, the target object could be a time varying signal, and the reference point a particular global timestamp that the tester is using as the reference time. The decision is made at a central decision center that receives a "batch" of data from distributed sensors that each communicate with the decision center, but are assumed to not communicate with each other.

This work focuses on a quantized communication, with each sensor's communicated bits being either of their measurement or of their identity; and here the "no-send" (censoring) option [2, 13] is now considered. The work follows from [15], and some initial study of the trade-off *without* censoring is in [16], absent the concept of censoring and its contribution to the implied information available to the decision-maker. We are not aware of others' work in this – as we consider, important – niche. However, there is a growing literature on the root problem of *unlabeled* data. Work in [9], [10] considers unlabeled detection in terms of making a decision based on shuffled sets of binary observations. Some specific applications include signal processing when received data has inaccurate time indexing [14], and making decisions with unlabeled "polling" data (yes/no answers from unidentified agents) as discussed in [19], [11]. The works [18, 4, 1, 3, 12] focus on signal reconstruction problems with unlabeled data, and identity-blind inference (as here) can be found in [6, 20, 21, 10, 7].

The sensors are assumed to inhabit a Poisson field. Each sensor sends its pair (z_i, s_i) – the quantized observation and the quantized identity – to the decision center. In a censoring scheme, each sensor decides locally whether or not to send its measurement to the decision center. In this model, the local censoring decision is made based on the value of a sensor's measurement Y_i, as well as the value of its label θ_i. Mathematical details are deferred to the longer paper [17].

THIS WORK WAS SUPPORTED BY AFOSR UNDER CONTRACT FA9500-18-1-0463.

2. STOCHASTIC MODEL

2.1. Ground Rules

Assume that the occurrence of sensors is a (possibly inhomogeneous) Poisson point process on the label space $\boldsymbol{\Theta}$ with arbitrary intensity function $\lambda(\vartheta)$. This is equivalent to both of the following statements being true [23]:

1. For any bounded interval $\Theta \subset \boldsymbol{\Theta}$, the number of sensors with label in Θ is Poisson random with mean value $\Lambda(\Theta) = \int_{\vartheta \in \Theta} \lambda(\vartheta)\, \mathrm{d}\vartheta$.
2. For any collection of disjoint intervals in $\boldsymbol{\Theta}$, the number of sensors in any one label interval is independent from the number in all others.

The intensity-density function $\mu_h(y, \vartheta)$ under hypothesis h defined on the joint measurement-label space $\boldsymbol{\mathcal{Y}} \times \boldsymbol{\Theta}$ is

$$\mu_h(y, \vartheta) = \lambda(\vartheta)\phi_h(y|\vartheta) \tag{2}$$

where $\phi_h(y|\vartheta)$ is the (uncensored) measurement probability density function under hypothesis h for a sensor with label ϑ.

Define the non-censored regions $D_{m,n}$ as

$$D_{m,n} = \{(y, \vartheta) : t_m \leq \vartheta < t_{m+1},\, q_{m,n} \leq y < q_{m,n+1}\}$$
$$m = 0, 1, \ldots, M^{\mathrm{s}} - 1, \quad n = 0, 1, \ldots, M^{\mathrm{z}} - 1 \tag{3}$$

Then define $\bar{\kappa}_{h,m,n}$ as the expected number of sensors with (measurement, label) pairs in $D_{m,n}$ under hypothesis h. Define $\bar{N}_{h,m}$ as the expected total number of sensors with label type $s = m$ (disregarding measurement type) under hypothesis h. Due to the Poisson properties,

$$\begin{aligned}\bar{N}_{h,m} &= \sum_{n=0}^{M^{\mathrm{z}}-1} \bar{\kappa}_{h,m,n} \\ &= \int_{t_m}^{t_{m+1}} \lambda(\vartheta) \left[\Phi_h(q_{m,M^{\mathrm{z}}}|\vartheta) - \Phi_h(q_{m,0}|\vartheta)\right] \mathrm{d}\vartheta \end{aligned} \tag{4}$$

where $\Phi_h(\cdot|\vartheta)$ is the evaluation of the measurement cumulative distribution function conditioned on the label. This reveals the special role of the lower and uppermost measurement quantization levels as censoring bounds—they have influence on the expected number of data of the corresponding label type that will be received by the decision center. Given the label type $s = m$ for a single sensor, the measurement type has a categorical distribution where the probability of type $z = n$ under hypothesis h is denoted $p_{h,m,n}$ and is given by

$$p_{h,m,n} = \frac{\bar{\kappa}_{h,m,n}}{\bar{N}_{h,m}}. \tag{5}$$

Let $\mathbf{p}_{h,m}$ be a vector of these probabilities over the measurement types.

2.2. Test Statistic

For any particular realization of the scheme, the decision center receives a batch $\mathbf{Z}$ of (measurement type, label type) pairs which can be equivalently represented as a series of "count vectors" denoted as $\boldsymbol{\kappa}_0, \ldots, \boldsymbol{\kappa}_m, \ldots, \boldsymbol{\kappa}_{M^{\mathrm{s}}-1}$, where element n of $\boldsymbol{\kappa}_m$ (denoted as $\kappa_{m,n}$) is the observed *number* of type $z = n$ measurements that were paired with type $s = m$ labels. Let N_m be the total number of type m labels received; that is $N_m = \sum_n \kappa_{m,n}$.

Since, for an observed batch, each N_m is a known constant, the observed count vectors are multinomial random. That is, under hypothesis h,

$$\boldsymbol{\kappa}_m \sim \mathrm{Mult}\left(N_m, \mathbf{p}_{h,m}\right) \tag{6}$$

where the symbol probabilities are defined in (5). Due to the definition of the probabilities $p_{h,m,n}$ given in (5), we can write

$$L(\mathbf{Z}) = \sum_{m=0}^{M^{\mathrm{s}}-1} \sum_{n=0}^{M^{\mathrm{z}}-1} \kappa_{m,n} \log\left(\frac{\bar{\kappa}_{1,m,n}}{\bar{\kappa}_{0,m,n}}\right) \tag{7}$$

where $\bar{\kappa}_{h,m,n}$ are the expected counts. For any given batch, the decision center will compute (7), and choose a hypothesis by comparing the result to a threshold τ as

$$L(\mathbf{Z}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau. \tag{8}$$

2.3. Optimization Metric

Several threshold require optimization, and as a criterion[1] we use Jeffrey's divergence [5]. That is

$$\begin{aligned} K = &D_{\mathrm{KL}}\left(f(\mathbf{Z}|\mathcal{H}_1); f(\mathbf{Z}|\mathcal{H}_0)\right) \\ &+ D_{\mathrm{KL}}\left(f(\mathbf{Z}|\mathcal{H}_0); f(\mathbf{Z}|\mathcal{H}_1)\right) \end{aligned} \tag{9}$$

where $D_{\mathrm{KL}}(\cdot\,;\cdot)$ denotes the Kullback-Leibler (KL) divergence. Via (7) this yields K explicitly as

$$K = \sum_{m=0}^{M^{\mathrm{s}}-1} \sum_{n=0}^{M^{\mathrm{z}}-1} (\bar{\kappa}_{1,m,n} - \bar{\kappa}_{0,m,n}) \log\left(\frac{\bar{\kappa}_{1,m,n}}{\bar{\kappa}_{0,m,n}}\right) \tag{10}$$

which is a function of only the quantization rules and any state-of-nature parameters. Thus one can seek a maximum in K over the quantization rules, and (if a maximum exists) treat the corresponding location in the parameter space as "optimal".

3. MEASUREMENT MODEL

Measurements are assumed to have an exponential distribution.

$$\phi_0(y|\theta_i) \quad = e^{-y}, \qquad y > 0 \tag{11}$$
$$\phi_1(y|\theta_i) \quad = a_i e^{-a_i y}, \quad y > 0, \;\; a_i \in (0, 1) \tag{12}$$

where the "label" is $\theta_i = ||\mathbf{x}^{\mathrm{t}} - \mathbf{x}_i||_2$ and

$$a_i \triangleq \frac{\theta_i^2}{\theta_i^2 + \rho^2}. \tag{13}$$

meaning that ρ quantifies the underlying signal to noise ratio. Sensors are modeled as arising from a homogeneous Poisson point process on $\mathbb{R}^3$ with constant intensity γ (Length^{-3}). Since the label is the Cartesian distance from the test point, any particular label value ϑ symbolizes a spherical shell centered on the test point with radius ϑ. The Possion point process on the label space thus has intensity function

$$\lambda(\vartheta) = \gamma 4\pi\vartheta^2 \tag{14}$$

[1] We could use many others.

—the constant volume intensity times the area of the spherical surface corresponding to ϑ.

It can be shown [17] that we have (see (3))

$$\bar{\kappa}_{0,m,n} = \gamma\frac{4}{3}\pi(t_{m+1}^3 - t_m^3)[e^{-q_{m,n}} - e^{-q_{m,n+1}}] \quad (15)$$

$$\bar{\kappa}_{1,m,n} = \int_{t_m}^{t_{m+1}} \gamma 4\pi\vartheta^2 [e^{\frac{-\vartheta^2 q_{m,n}}{\vartheta^2+\rho^2}} - e^{\frac{-\vartheta^2 q_{m,n+1}}{\vartheta^2+\rho^2}}]\, \mathrm{d}\vartheta \quad (16)$$

The integral in (16) has no known closed form.

To obtain the expected per-batch communication cost in bits, first note that each (measurement type, label type) pair can be encoded to a unique string of

$$b = \log_2(M^{\mathrm{z}} M^{\mathrm{s}}) \quad (17)$$

bits. That is, b is the per-sensor communication cost for sending a single (measurement type, label type) pair. Then, the expected per-batch communication cost (under $\mathcal{H}_0$) is given by multiplying b by the expected total number of reporting sensors [17] as

$$\begin{aligned} B &= b \sum_{m=0}^{M^{\mathrm{s}}-1} \sum_{n=0}^{M^{\mathrm{z}}-1} \bar{\kappa}_{0,m,n} \\ &= b\,\gamma\frac{4}{3}\pi \sum_{m=0}^{M^{\mathrm{s}}-1} (t_{m+1}^3 - t_m^3)(e^{-q_{m,0}} - e^{-q_{m,M^{\mathrm{z}}}}) \end{aligned} \quad (18)$$

The analysis in the next section is based on numeric maximization of K over the quantization parameters subject to (18) as an equality constraint.

4. NUMERIC ANALYSIS

For any particular fixed per-batch expected communication cost constraint B, and fixed state of nature ρ, constrained maximization of K is performed with the 'trust-constr' method of SciPy's optimize.minimize module [22]. Since both B and K are simply proportional to the other state of nature parameter γ, results are simply normalized by γ. For this particular model, some of the censoring parameters can be optimized trivially. Namely, the lower label censoring threshold can be fixed as $t_0 = 0$, and all the upper measurement censoring thresholds can be fixed as $q_{m,M^{\mathrm{z}}} = \infty,\ m = 0, ..., M^{\mathrm{z}}-1$. Thus the maximization is performed over the remaining free parameters—the upper label censoring threshold, the lower measurement censoring thresholds, and any intermediate quantization levels.

The trade-off of quantization fineness between data and measurement was explored in [16]. Here, for the sake of demonstration, we choose fixed $M^{\mathrm{s}} = 2$ and $M^{\mathrm{z}} = 2$ (i.e., both binary). That is, each label type s takes on one of two values, and each measurement type z is also one of two values, meaning the per-sensor communication cost is $b = 2$ bits. Let the result of maximization be

$$\hat{K}(B) = \max_{t_1,t_2,q_{0,0},q_{0,1},q_{1,0},q_{1,1}} K \quad (19)$$

subject to the constraints

$$t_1 < t_2 \quad (20)$$

$$q_{0,0} < q_{0,1} \quad (21)$$

$$q_{1,0} < q_{1,1} \quad (22)$$

$$b\,\gamma\frac{4}{3}\pi\sum_{m=0}^{1}(t_{m+1}^3 - t_m^3)e^{-q_{m,0}} = B \quad (23)$$

Figure 1 plots the maximized value of K versus the communication constraint B for various values of ρ. Interestingly, we see that a peak in optimized performance occurs over B. That is, there is some per-batch communication cost above which performance is actually lost—it is possible to listen to too many sensors. This is because an increase in the expected number of reporting sensors, while resulting in more data, also corresponds to an increase in label and measurement uncertainty. There is a tradeoff between amount and quality of information received by the decision maker.

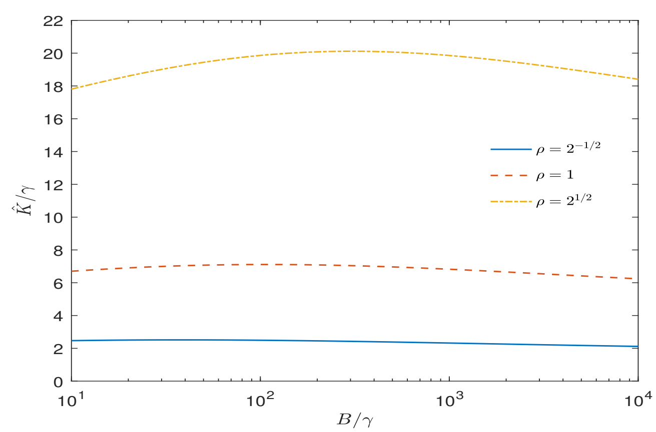

Fig. 1. The maximized value of K versus the total per-batch communication cost (in bits) for various choices of ρ, and $M^{\mathrm{s}} = 2, M^{\mathrm{z}} = 2$.

Figure 2 plots the optimal label quantization rule, including the upper label censor threshold. Figure 3 gives the optimal measurement censoring and quantization for sensors in the "inner" label bin $s = 0$, and Figure 4 gives the optimal rule for sensors in the "outer" label bin $s = 1$. Note that the upper and lower thresholds correspond to censoring by label (i.e., distance from the test point) and datum, respectively. Both smoothly change for more information flow as greater flow is permitted.

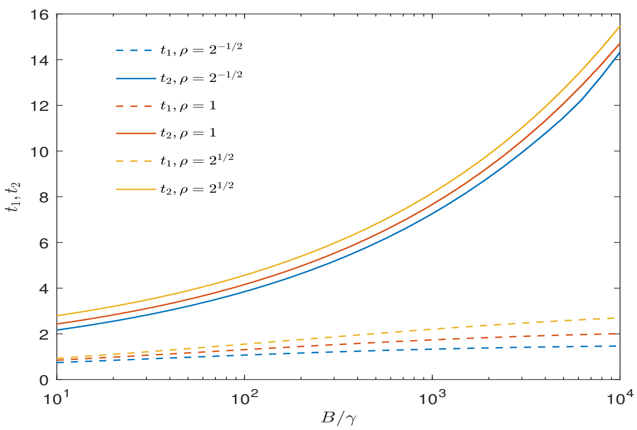

Fig. 2. The optimal label quantization levels versus the total per-batch communication cost (in bits) for various choices of ρ, and $M^{\mathrm{s}} = 2, M^{\mathrm{z}} = 2$. Note that t_2 is the upper label censoring threshold. The lower label censoring threshold is fixed as $t_0 = 0$.

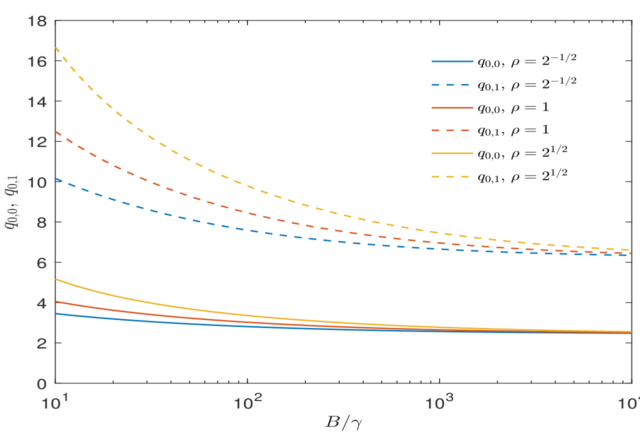

Fig. 3. The optimal measurement quantization levels for sensors in the inner label bin $s = 0$ versus the total per-batch communication cost (in bits) for various choices of ρ, and $M^{\mathrm{s}} = 2, M^{\mathrm{z}} = 2$. Note that $q_{0,0}$ serves a special role as the lower measurement censoring threshold. The upper label censoring threshold is fixed as $q_{0,2} = \infty$.

5. SIMULATION RESULTS

MC simulations are performed under both hypotheses for each of the peak locations in Figure 1. For instance, the data series corresponding to $\rho = \sqrt{2}$ has a peak at $B = 300$ bits. Sensors are simulated according to the Poisson point process on the censored label space, then batches of measurements are sampled under each hypothesis. The optimal rules are read from Figure 2 through Figure 4. The resulting ROC curves are plotted in Figure 5. The formulation of K is also checked: for each of the 3 simulation settings the difference between the empirical and theoretical values of K is below 0.05.

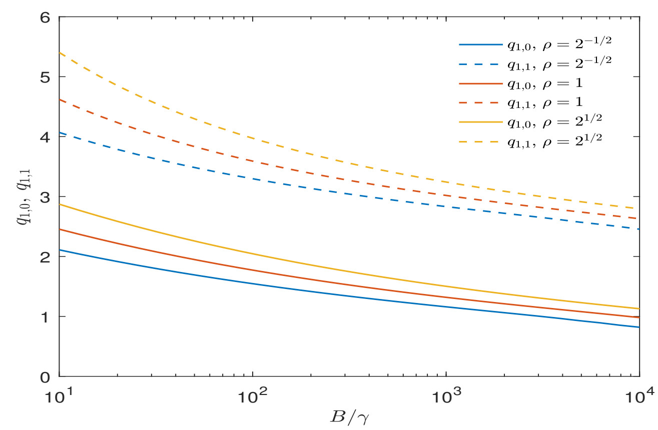

Fig. 4. The optimal measurement quantization levels for sensors in the outer label bin $s = 1$ versus the total per-batch communication cost (in bits) for various choices of ρ, and $M^{\mathrm{s}} = 2, M^{\mathrm{z}} = 2$. Note that $q_{1,0}$ serves a special role as the lower measurement censoring threshold. The upper label censoring threshold is fixed as $q_{1,2} = \infty$.

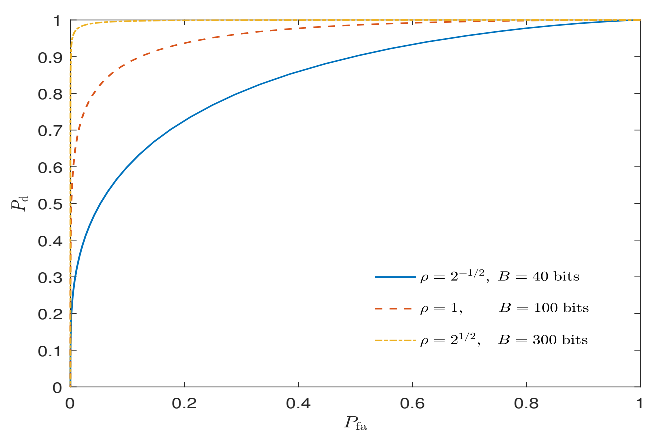

Fig. 5. The ROC curves for various choices of ρ operating at the corresponding optimal batch cost B. $\gamma = 1$, $M^{\mathrm{s}} = 2$ and $M^{\mathrm{z}} = 2$. There were 10^6 Monte Carlo simulations.

6. SUMMARY

A sensor network in which the bandwidth budget is split between the observation and the identity of the sensor is considered. Here, the "identity" refers to expected signal strength; but it might quantify (anonymized) knowledge level in a crowd-sourcing application. Here we explicitly model the process of "censoring" by which there is no transmission if such transmission is insufficiently suggestive. We find, in this exploratory work, that both data types are affected similarly by relaxation of communication constraints; and that there seems to be a "sweet spot" of allowable bandwidth beyond which added transmission simply roils the picture.

7. REFERENCES

[1] A. Abid, A. Poon, and J. Zou. (2017, May 4) Linear regression with shuffled labels. [Online]. Available: http://arxiv.org/abs/1705.01342

[2] S. Appadwedula, V. V. Veeravalli, and D. Jones, "Decentralized detection with censoring sensors," *IEEE Trans. Signal Processing*, vol. 56, pp. 1362-1373, Apr. 2008.

[3] G. Elhami, A. Scholefield, B. B. Haro, and M. Vetterli, "Unlabeled sensing: Reconstruction algorithm and theoretical guarantees," in *ICASSP*, New Orleans, March 5-9, 2017.

[4] S. Haghighatshoar and G. Caire, "Signal recovery from unlabeled samples," *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1242–1257, March 2018.

[5] H. Jeffreys, *Theory of Probability.* Clarendon Press, Oxford, 2nd edition, 1948.

[6] Z. Liu and J. Zhu, "Signal detection from unlabeled ordered samples," *IEEE Communications Letters*, vol. 22, no. 12, pp. 2431–2434, Dec. 2018.

[7] S. Marano and P. Willett, "Algorithms and fundamental limits for unlabeled detection using types," *IEEE Transactions on Signal Processing*, vol. 67, no. 8, pp. 2022–2035, Apr. 2019.

[8] S. Marano and P. Willett, "Algorithms and Fundamental Limits for Unlabeled Detection Using Types", *IEEE Transactions on Signal Processing*, vol: 67, no. 8, April 2019.

[9] S. Marano and P. Willett, "Making decisions with shuffled bits," *Proc. of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, 12-17 May, 2019.

[10] S. Marano, V. Matta, P. Willett, P. Braca, and R. Blum, "Hypothesis testing in the presence of Maxwell's daemon: Signal detection by unlabeled observations," *ICASSP*, New Orleans, March 2017.

[11] S. Marano, V. Matta, and P. Willett, "The importance of being earnest: Social sensing with unknown agent quality," *IEEE Trans. Signal and Info. Process. over Networks*, vol. 2, no. 3, pp. 306-320, Sep. 2016.

[12] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Linear regression with an unknown permutation: Statistical and computational limits," in *Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2016, pp. 417–424.

[13] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: a low communication rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554-568, April 1996.

[14] L. Schenato, "Optimal estimation in networked control systems subject to random delay and packet drop," *IEEE Transactions on Automatic Control*, vol. 53, no. 5, pp. 1311-1317, June 2008.

[15] Z. Sutton, P. Willett and S. Marano, "Sensor Network Target Detection with Unlabeled Observations," *IEEE Aerospace Conference*, Big Sky MT, March 2020.

[16] Z. Sutton, P. Willett and S. Marano, "Target Detection from Distributed Passive Sensors: Semi-Labeled Data Quantization," *ICASSP*, Toronto, Canada, June 2021.

[17] Z. Sutton, P. Willett and S. Marano, "Optimal Quantization and Censoring for Detection with Measurement and Label Uncertainty," *submitted.*

[18] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Linear regression with shuffled data: Statistical and computational limits of permutation recovery," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3286–3300, May 2018.

[19] D. Wang, L. Kaplan, and T. Abdelzaher, "Maximum likelihood analysis of conflicting observations in social sensing," *ACM Trans. on Sensor Networks*, vol. 10, no. 2, Jan. 2014.

[20] G. Wang, J. Zhu, R. S. Blum, P. Willett, S. Marano, V. Matta, and P. Braca, "Signal amplitude estimation and detection from unlabeled binary quantized samples," *IEEE Transactions on Signal Processing*, vol. 66, no. 16, pp. 4291–4303, Aug. 2018.

[21] J. Zhu, H. Cao, C. Song, and Z. Xu, "Parameter estimation via unlabeled sensing using distributed sensors," *IEEE Communications Letters*, vol. 21, no. 10, pp. 2130–2133, Oct 2017.

[22] scipy.optimize.minimize, 2021. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html. [Accessed: Sept 1 2021].

[23] "Poisson Point Processes," [Online]. Available: https://en.wikipedia.org/wiki/Poisson_point_process.