# DISTRIBUTED ESTIMATION VIA PAID CROWD WORK

*Song Jianhan, Vei Wang Isaac Phua*[*]

Chinese University of Hong Kong
Hong Kong

*Lav R. Varshney*

University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA

**ABSTRACT**

Consider a distributed estimation problem to be carried out by paid crowdworkers, where results are to be returned quickly and accurately. Estimation accuracy is a function of the number of workers completing the job and of the quality of the workers, both of which may be influenced by the payment offered. With limited budget, payment allocation should consider both effects to obtain best results. Since people are not deterministic, payment offers will lead to a random number of variable-quality workers, as governed by choice models. We consider average performance and focus on estimating a parameter from measurements through uniform noise. Since we have shown the optimality of the midrange estimator in specific settings of the general problem, we focus on the best linear unbiased estimator based on order statistics (BLUE-OS) under the mean-squared error (MSE) criterion. Best payment allocations are determined for single crowd platforms, joint population models and separated platform models. Illustrative numerical examples are provided.

***Index Terms***— choice models, crowdsourcing, distributed estimation, resource allocation

## 1. INTRODUCTION

Dating to Francis Galton accurately estimating the weight of a fat ox at the 1906 Plymouth country fair using guesses from both amateurs and experts (like farmers and butchers) competing for a prize [1], there has been interest in using the wisdom of the crowd for distributed estimation. In recent times, crowdsourcing through platforms like Amazon Mechanical Turk has become prevalent for estimation tasks ranging from participatory infectious disease and rural crop surveillance [2, 3], to real-time paid crowdsourcing to help the blind [4], where it is important to get results quickly and accurately. Unlike classical settings of distributed estimation like sensor networks [5], however, resource constraints are different in human crowdsourcing. Human workers are governed by choices and desires, which determine how much cognitive energy they will allocate to the task at hand [6]. We consider the classical parallel tree topology for distributed estimation, rather than requiring all agents to obtain final estimates [7].

Monetary payment is a standard way to incentivize workers [8] and changes in rewards have two effects: changing the *quality* of work produced [9], and changing the likelihood for the work to be completed [10, 11] and thereby the *quantity*. Drawing on much evidence from human experiments, it is now standard to model the choices of crowd workers by *discrete choice models* [10], where individuals act to maximize gain in utility. Workers may have different utilities depending on wage, ease, and expected time of tasks [11]. We assume a worker accepts a task when his or her utility exceeds that of other tasks [11]. If accepted, the quality of work may depend on the worker's expertise, abilities, motivation, etc. [8]. For a given payment scheme, we have a random number of variable-quality workers.

This paper characterizes and optimizes distributed estimation performance under these novel crowd-based resource constraints. For concreteness and since it provides a good model for estimating beliefs, we focus on workers making observations through uniform noise [12]. For a fixed number of agents (as in classical distributed estimation), it is well-known that the midrange estimator is optimal under mean-squared error (MSE) [12–15]. Proof omitted here for brevity, we have extended this optimality result for a *random* number of agents using the theory of order statistics for random number of variates [16–19]. This theory has been used in reliability engineering [20–22], but not statistical signal processing.

For settings with variable-quality workers, we also focus on estimators that are linear functions of order statistics, which include not just the midrange, but also the median, mean, and many other typical location parameter estimators. We find the best linear unbiased estimator based on order statistics (BLUE-OS) [23].

We also optimize payment allocations for several kinds of platforms: those with equal-quality workers, and those with variable-quality workers where amateurs and experts can either be specifically addressed or not (comparisons are also made between the two). Surprisingly, it is not always best to have as many experts as possible. Note that even though the number of workers participating will be a random response to the allocation, once a payment is shown to a worker, it is considered "spent" whether or not the task is completed.

Though there is prior work on budget allocation for crowdsourcing, it all deals with classification rather than estimation [24–27].

## 2. RANDOM NUMBER OF EQUAL-QUALITY WORKERS

We are to estimate a parameter $\theta$ in a crowdsourcing platform where worker observations are uniformly distributed on $[\theta - \frac{\alpha}{2}, \theta + \frac{\alpha}{2}]$, for some constant $\alpha$. There are no specific communication constraints between workers and the platform. The platform uses the midrange estimator, the average of the minimum and the maximum of the sample. This is optimal for uniform sources, whether for a fixed or random number of variates.

**Theorem 1** *The best unbiased estimator of the location parameter from an i.i.d. sequence of uniform observations of fixed or random length is the sample midrange.*

Optimality of the sample midrange estimator for a fixed length sample is classically known, e.g. [12, 13]. The random case follows from independence, linearity, and the prior fixed-length results.

[*]Song Jianhan and V. W. I. Phua were visiting the University of Illinois at Urbana-Champaign when this work was completed, under the auspices of the Prof. Charles K. Kao Summer Research Exchange Scholarship program.

By selecting payments $\vec{c}$ (e.g. in cents) with elements in a discrete space $\mathcal{C}$ under a fixed budget $B$, we are to maximize system performance. The probability a worker will choose to do a specific task is derived from a *conditional logit model* of discrete choice. Assuming the utility of a task is linear in payment $c$, the task acceptance probability function follows a *multinomial logit distribution* [10,11]:

$$q(c) = \exp\left\{\frac{c}{s} - b\right\} / \exp\left\{\frac{c}{s} - b\right\} + M,$$

where the parameters $b, s, M$ are empirically determined.

The allocation of payments among $n$ workers can be expressed as an allocation vector $\vec{c} = \{c_0, c_1, \ldots, c_n\}$, where $c_k$ is the payment offered to the $k$th worker. We define the best allocation $\vec{c}^*$ as the allocation which results in the lowest MSE in estimation.

The MSE for a deterministic participation count $n$ is $\frac{\alpha^2}{2(n+1)(n+2)}$ for $n \geq 1$, from known results on the midrange [13]. Thus we set $\mathsf{mse}(n = 0) = \alpha^2/4$ for consistency, but this is not critical. The best allocation, given the constraint $\sum_{i=1}^{N_0} c_i = B$ for any allocation vector $\vec{c}$ and upper-bounding the population at $N_0$, satisfies:

$$\vec{c}^* = \arg\min_{\vec{c}} \sum_{n=0}^{N_0} \frac{\alpha^2}{2(n+1)(n+2)} \cdot p_N(n; \vec{c}),$$

where the distribution of the random participant count $N$:

$$p_N(n; \vec{c}) = \frac{1}{n!(N_0-n)!} \times$$
$$\text{per} \begin{bmatrix} q(c_1) & \cdots & q(c_1) & 1-q(c_1) & \cdots & 1-q(c_1) \\ q(c_2) & \cdots & q(c_2) & 1-q(c_2) & \cdots & 1-q(c_2) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \underbrace{q(c_{N_0}) \quad \cdots \quad q(c_{N_0})}_{n} & & & \underbrace{1-q(c_{N_0}) \quad \cdots \quad 1-q(c_{N_0})}_{N_0-n} & & \end{bmatrix}$$

is a generalization of the binomial distribution for counting successes among unequally-likely events, in terms of the matrix permanent.

**Theorem 2** *The best payment allocation is homogeneous.*

Proof of this result follows directly from the following.

**Theorem 3** *Among allocations of form $\{x, 2B' - x, c_3, \ldots, c_{N_0}\}$ with fixed $B'$ and $c_i$, $i = 3, \ldots, N_0$ and given $x$, $0 \leq x \leq 2B'$, the best allocation occurs only when either $x = 0$, $2B'$, or $B'$.*

To prove this, we note the $\mathsf{mse}$ function is continuous. The boundary points when $x = 0$ or $2B'$ can yield maximum or minimum. Further, we can note that $p_N(n; \vec{x})$ is symmetric about $x = B'$ since $p_N(n; \vec{B' - t}) = p_N(n; \vec{B' + t})$. Therefore, $x = B'$ is a local extremum which may also be the global extremum.

The key is showing other points cannot also be minima. First we consider $N_0 = 2$ and the result for larger $N_0$ will follow from the base case. Without loss of generality, set $\alpha = 1$. We will use the following special property of $q(c)$. Suppose $K(B', p(B'))$ is an arbitrary point on the curve $q(c)$. Then $M(B' - t, p(B' - t))$ and $N(B' + t, p(B' + t))$ lie on either side of $K (t > 0)$. Thus,

$$\frac{1}{t \cdot k_{KM}} - \frac{1}{t \cdot k_{KN}} = \frac{1}{M} \exp\{\frac{B'}{a} - b\} - M \exp\{b - \frac{B'}{a}\}.$$

The difference explicitly driven by $t$ (the two slopes are also determined by $t$) is a constant depending on $B'$ without regard to $t$.

Recall when $N_0 = 2$:

$$\mathsf{mse}(\vec{c}) = \sum_{n=0}^{2} \frac{p_N(n; \vec{c})}{2(j+1)(j+2)}.$$

Suppose the budget is $2B'$ and let $\vec{c}_t$ be $[B' - t, B' + t]$, $0 \leq t \leq B'$. Then:

$$\begin{aligned} \Delta_{B'}(t) &= \mathsf{mse}(\vec{c}_t) - \mathsf{mse}(\vec{c}_0) \\ &= A_1(t\beta_1(t) - t\beta_2(t)) - A_2 t^2 \beta_1(t)\beta_2(t) \end{aligned}$$

where

$$\beta_1(t) = \frac{q(B') - q(B' - t)}{t}, \beta_2(t) = \frac{q(B' + t) - q(B')}{t}$$
$$A_1 = (\tfrac{1}{4} - \tfrac{1}{12})(1 - q(B')) + (\tfrac{1}{12} - \tfrac{1}{24})q(B'), A_2 = \tfrac{1}{4} - 2 \cdot \tfrac{1}{12} + \tfrac{1}{24}.$$

Note $\beta_1(t) > 0$ and $\beta_2(t) > 0$ for all $0 \leq t \leq B'$ since $q(c)$ is monotonically increasing. Therefore,

$$\begin{aligned} \Delta_{B'}(t) &= (t^2\beta_1(t)\beta_2(t)) \left[A_1\left(\tfrac{1}{t\beta_2(t)} - \tfrac{1}{t\beta_1(t)}\right) - A_2\right] \\ &= (t^2\beta_1(t)\beta_2(t)) \cdot \text{Const}(B'). \end{aligned}$$

When $\text{Const}(B') \geq 0$, $\Delta_{B'}(t) \geq 0$ for all $0 \leq t \leq B'$, and $\vec{c}_0$ is the best allocation.

When $\text{Const}(B') < 0$, we see $\vec{c}_{B'} = \{0, 2B'\}$ is the best allocation since

$$\begin{aligned} \Delta_{B'}(B') &= (B'^2\beta_1(B')\beta_2(B')) \cdot \text{Const}(B') \\ &= [q(B') - q(0)][q(2B') - q(B')] \cdot \text{Const}(B') \\ &\leq ([q(B') - q(B' - t)][q(B' + t) - q(B')] \cdot \text{Const}(B'), \\ &\qquad\qquad \text{for all } 0 \leq t \leq B' \\ &= (t^2\beta_1(t)\beta_2(t)) \cdot \text{Const}(B') = \Delta_{B'}(t). \end{aligned}$$

Hence we have

$$\mathsf{mse}(\vec{c}_{B'}) \leq \mathsf{mse}(\vec{c}_t) \text{ for all } 0 \leq t \leq B'.$$

Reducing the $N_0 > 2$ case to a summation of many $N_0 = 2$ cases extends the base case. Restricting to homogeneous allocations (without loss of optimality) simplifies determining best allocations.

## 3. FIXED NUMBER OF VARIABLE-QUALITY WORKERS

The quality of workers in the crowd may not be equal: some workers may provide better measurements than others, but this may be controlled through financial incentives [8]. Here we model the quality of work with uniform distributions of different scales and assume there are two levels: *expert* and *amateur*. Smaller scales yield better estimates. Contrary to the previous section, we assume fixed task completion. There are two ways to model platforms with variable-quality workers: joint population and separated platforms.

### 3.1. Joint Population Model

The measurements of amateurs and experts are modeled by two uniform distributions with support $[-1, 1]$ and $[-0.5, 0.5]$, respectively. For simplicity of notation, we take parameter $\theta = 0$. The platform combines reported measurements using the best linear unbiased estimator based on order statistics (BLUE-OS) [23], without knowing which workers are experts and which are amateurs. Since variates are not identically distributed, results of [23] muat be modified and so we start from first principles.

Let $n$ be the total number of workers and let $m$ be the number of amateurs among the $n$. The marginal pdf of the $r$th order statistic from the joint population is derived following [28, Ch. 5] and is given

$$
f_{(r)}(x) = \frac{1}{(r-1)!(n-r)!} \times \begin{array}{c} \qquad r-1 \text{ columns} \qquad\qquad\qquad\qquad n-r \text{ columns} \\ per \begin{bmatrix} F_1(x) & \dots & F_1(x) & f_1(x) & 1-F_1(x) & \dots & 1-F_1(x) \\ & \vdots & & (m\,rows) & & \vdots & \\ F_1(x) & \dots & F_1(x) & f_1(x) & 1-F_1(x) & \dots & 1-F_1(x) \\ F_2(x) & \dots & F_2(x) & f_2(x) & 1-F_2(x) & \dots & 1-F_2(x) \\ & \vdots & & (n-m\,rows) & & \vdots & \\ F_2(x) & \dots & F_2(x) & f_2(x) & 1-F_2(x) & \dots & 1-F_2(x) \end{bmatrix} \end{array} \tag{1}
$$

$$
f_{(r)(s)}(x,y) = \frac{1}{[(r-1)!(s-r-1)!(n-s)!]} \times \begin{array}{c} \qquad m \text{ columns} \qquad\qquad\qquad\qquad n-m \text{ columns} \\ per \begin{bmatrix} F_1(x) & \dots & F_1(x) & F_2(x) & \dots & F_2(x) \\ & \vdots & & (r-1\,rows) & \vdots & \\ F_1(x) & \dots & F_1(x) & F_2(x) & \dots & F_2(x) \\ f_1(x) & \dots & f_1(x) & f_2(x) & \dots & f_2(x) \\ F_1(y)-F_1(x) & \dots & F_1(y)-F_1(x) & F_2(y)-F_2(x) & \dots & F_2(y)-F_2(x) \\ & \vdots & & (s-r-1\,rows) & \vdots & \\ F_1(y)-F_1(x) & \dots & F_1(y)-F_1(x) & F_2(y)-F_2(x) & \dots & F_2(y)-F_2(x) \\ f_1(y) & \dots & f_1(y) & f_2(y) & \dots & f_2(y) \\ 1-F_1(y) & \dots & 1-F_1(y) & 1-F_2(y) & \dots & 1-F_2(y) \\ & \vdots & & (n-s\,rows) & \vdots & \\ 1-F_1(y) & \dots & 1-F_1(y) & 1-F_2(y) & \dots & 1-F_2(y) \end{bmatrix} \end{array} \tag{2}
$$

by (1). The joint pdf of the $r$th and $s$th order statistics is given in (2). We can use these distributions to calculate the expected values and the covariance matrix of the order statistics from the joint population, which in turn allows us to find the BLUE-OS by optimizing weights in the summation, and finally its MSE (following [23]). For brevity, this is all omitted, and we show numerical results.

We show the BLUE-OS MSEs for several values of $n$ and $m$ in Fig. 1. Observe for sufficiently large $n$, there is a peak in the graph, such that maximizing the number of experts is not best. Though, the position of the peak shifts to the left, i.e. the best fraction of experts does increase, as $n$ increases. Moreover, the value of the peak decreases as $n$ increases (estimation accuracy improves with sample size), but the curves become flatter as $n$ increases, indicating the quality of the estimator relies less on experts.

The first point is non-intuitive: the estimator may improve if the number of amateurs in the sample exceeds a threshold $m_0$. In the $m \leq m_0$ regime, as the number (and influence) of the amateurs in the sample increases, the estimator degrades (as expected). When $m$ exceeds $m_0$, however, measurements from amateurs dominate the sample distribution, and the distribution become more uniform. Hence, the BLUE-OS estimates $\theta$ more closely, and the dispersion decreases. Nevertheless, the MSE when $m = n$ (all amateurs) is always larger than the MSE when $m = 0$ (all experts).

### 3.2. Separated Platforms Model

Now suppose we can maintain two separate platforms with experts and amateurs that can be distinguished from one another. Let the measurements from these two platforms both be uniform with scale parameters $\alpha_A < \alpha_B$, respectively. Let the budget assigned to the two platforms be $B_A$ and $B_B$, so $B_A + B_B = B$. We first characterize MSE for a given payment allocation.

To do so, we should first achieve best estimation given a fixed number of workers in platforms A and B, as in prior sections, and base a BLUE-OS solution on that.

**Theorem 4** *Let $\theta$ be a parameter to be estimated in $m$ independent platforms $P_1$ to $P_m$, $m \geq 1$. Suppose there are $N_i$ independent observations from $P_i$, $(X_1^i, X_2^i, \dots, X_{N_i}^i)$. Let $e_i$ be the BLUE-OS using the observations of $P_i, 1 \leq i \leq m$ respectively. Then, the BLUE-OS of $\theta$ from the $m$ platforms is a linear combination of $e_i, 1 \leq i \leq m$.*

For proof, we follow the method of Lloyd [23] to prove there exists a BLUE-OS estimator for $\theta$. A key is noting the independence of the different platforms, which leads to block-diagonal covariance.

Since the midrange is the BLUE-OS for a single uniform distribution, the BLUE-OS for two platforms A and B is a weighted average of two midrange estimators. Working through derivations, it turns out the BLUE-OS of $\hat{\theta}$ is

$$
\hat{\theta} = \frac{k_A}{2(k_A + k_B)}(Y_1^A + Y_{N_A}^A) + \frac{k_B}{2(k_A + k_B)}(Y_1^B + Y_{N_B}^B)
$$

with $k_A = \frac{(N_A+1)(N_A+2)}{\alpha_A^2}$ and $k_B = \frac{(N_B+1)(N_B+2)}{\alpha_B^2}$. The variance is the MSE, which we can write down and optimize under the budget constraint (omitted for brevity).

Fig. 2 plots the MSEs with different numbers of amateurs $m$ in $n$ participants (this can be computed continuously). Again, we observe the non-intuitive behavior that after a certain threshold $m_0$, increasing the number of amateurs lowers the MSE. For the separated model, we can take $\frac{d}{dm}\mathsf{mse}(n-m, m) = 0$ and easily obtain

$$
m_0 = \frac{\alpha_B^2(2n+3) - 3\alpha_A^2}{2(\alpha_A^2 + \alpha_B^2)}.
$$

Since the extremum exists only when $m_0 < n$, we have the non-intuitive behavior if $3(\alpha_B^2 - \alpha_A^2)/2\alpha_A^2 < n$.

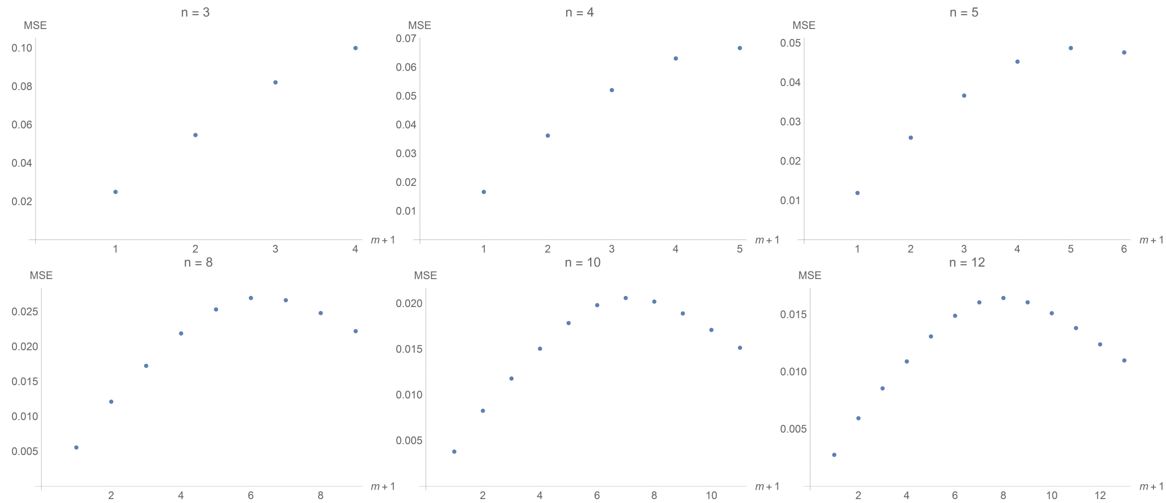


**Fig. 1**. Trends of BLUE-OS MSE under different values of $n$ and $m$.

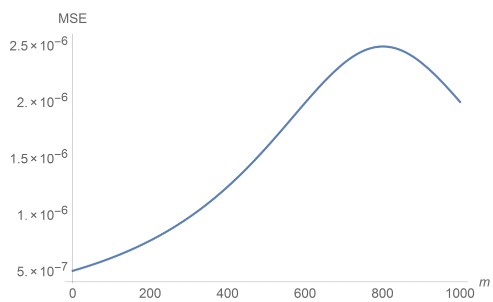


**Fig. 2**. BLUE-OS MSE as continuous function of $m$ ($n = 1000$).

## 4. OPTIMIZING QUALITY AND QUANTITY

Now we consider both quality and quantity aspects of payments in the separated model. As shown, the optimal reward allocation in a single platform with equal quality workers is homogeneous. It is clear the problem reduces to finding the optimal $[\frac{B_A}{N_A}, \frac{B_B}{N_B}]$, where $B_A$ and $B_B$ are the budgets allocated to the two platforms respectively, so MSE of the final estimate is minimal.

The algorithm to calculate the optimal budget allocation consists of several steps:

1. Choose $B_A$, $B_B$.
2. For the allocation, plot a graph of $(N_A, N_B, \mathsf{mse})$ and find the first minimum with reference to MSE.
3. Repeat 1-2 and get the allocation with the lowest MSE.

The process of trying all combinations of $B_A$ and $B_B$ is in linear time, but to speed up the calculation further, we notice a useful property of the model. We can transform previous results into:

$$\sum_{i=0}^{N_A} \alpha_B^2 p_N(i;A) \sum_{j=0}^{N_B} \frac{p_N(j;B)}{2\frac{\alpha_B^2}{\alpha_A^2}(i+1)(i+2)+2(j+1)(j+2)}$$

The MSE of one platform reduces to the following, when fixing the number of participants in the other platform:

$$f(M) = \sum_{j=0}^{M} \frac{1}{k+2(j+1)(j+2)} p_N(j).$$

When $k = 0$, $f(M)$ reduces to:

$$f_0(M) = \frac{1-(M+2)q(\frac{B_0}{M})^{t+1} - (1-q(\frac{B_0}{M}))^{t+2}}{2(t+1)(t+2)q(\frac{B_0}{M})^2}.$$

Numerical results show that for $k < k_0$, where $k_0$ is a constant depending on the choice model $q(c)$ and the budget for the platform $B_0$, the trend of $f(M)$ stays the same, i.e. the extremum occurs for the same $M$. The $k_0$ value can be obtained by testing different values of $k$. The allocation of one platform does not affect the best allocation of the other when $k < k_0$. Therefore, we obtain the optimal $M$ quickly by finding the minimum of $f_0(M)$ using derivatives.

## 5. CONCLUSION

As crowdsourcing becomes widespread, research challenges arise at the intersection of signal processing and social networks. Distributed estimation in sensor networks is limited by noise and bandwidth [5], but crowdsourcing is limited by human expertise and motivation. This paper discussed best payment allocations based on order statistics in a single platform with equal-quality workers and presented two modeling approaches for platforms with variable-quality workers. The separated platforms model performs better since worker identity information is lost in the joint population model. Algorithms to find optimal payment allocations in these models were developed.

Although models studied herein are based strongly on actual real-time crowd-based tasks, formulations can certainly be generalized and applied to other human computation scenarios. Experiments with human subjects, e.g. on Amazon Mechanical Turk, can also be carried out to verify our theoretical development.

We also note that transmission erasures (i.e. uncompleted work) in distributed estimation have been considered elsewhere [29] in contexts where multiple descriptions encoding is possible [30]. Here we just considered raw, uncoded crowd results, but the possibility of coding for crowdsourcing exists too [31]. As part of future work, we will investigate analog coding for crowd-based estimation.

## 6. REFERENCES


[1] J. Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday, 2004.

[2] O. P. Wójcik, J. S. Brownstein, R. Chunara, and M. A. Johansson, "Public health for the people: participatory infectious disease surveillance in the digital age," *Emerg. Themes Epidemiol.*, vol. 11, no. 7, Jun. 2014.

[3] J. Mohite, Y. Karale, P. Gupta, S. Kulkarni, B. Jagyasi, and A. Zape, "RuPS: Rural participatory sensing with rewarding mechanisms for crop monitoring," in *Proc. 2015 IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, Mar. 2015, pp. 378–383.

[4] J. P. Bigham, C. Jayant, A. Miller, B. White, and T. Yeh, "VizWiz::LocateIt - enabling blind people to locate objects in their environment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognition Workshops (CVPRW)*, Jun. 2010, pp. 65–72.

[5] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, Jun. 2005.

[6] L. R. Varshney, "Participation in crowd systems," in *Proc. 50th Annu. Allerton Conf. Commun. Control Comput.*, Oct. 2012, pp. 996–1001.

[7] C.-K. Yu, M. van der Schaar, and A. H. Sayed, "Information-sharing over adaptive networks with self-interested agents," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 1, no. 1, pp. 2–19, Mar. 2015.

[8] W. Mason and D. J. Watts, "Financial incentives and the "performance of crowds"," in *Proc. ACM SIGKDD Workshop Human Comput. (HCOMP'09)*, Jun. 2009, pp. 77–85.

[9] C.-J. Ho, A. Slivkins, S. Suri, and J. W. Vaughan, "Incentivizing high quality crowdwork," in *Proc. 24th Int. Conf. World Wide Web (WWW'15)*, May 2015, pp. 419–429.

[10] S. Faradani, B. Hartmann, and P. G. Ipeirotis, "Whats the right price? pricing tasks for finishing on time," in *Proc. AAAI Workshop Human Comput. (HCOMP'11)*, Aug. 2011, pp. 26–31.

[11] Y. Gao and A. Parameswaran, "Finish them!: Pricing algorithms for human computation," *Proc. VLDB Endowment*, vol. 7, no. 14, pp. 1965–1976, Oct. 2014.

[12] A. Vempaty and L. R. Varshney, "The non-regular CEO problem," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2764–2775, May 2015.

[13] J. Neyman and E. S. Pearson, "On the use and interpretation of certain test criteria for purposes of statistical inference: Part I," *Biometrika*, vol. 20A, no. 1/2, pp. 175–240, Jul. 1928.

[14] P. R. Rider, "The midrange of a sample as an estimator of the population midrange," *J. Am. Stat. Assoc.*, vol. 52, no. 280, pp. 537–542, Dec. 1957.

[15] G. R. Arce and S. A. Fontana, "On the midrange estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 6, pp. 920–922, Jun. 1988.

[16] K. Raghunandanan and S. A. Patil, "On order statistics for random sample size," *Stat. Neerl.*, vol. 26, no. 4, pp. 121–126, Dec. 1972.

[17] J. M. Buhrman, "On order statistics when the sample size has a binomial distribution," *Stat. Neerl.*, vol. 27, no. 3, pp. 125–126, Sep. 1973.

[18] V. K. Rohatgi, "Distribution of order statistics with random sample size," *Commun. Stat. Theory Methods*, vol. 16, no. 12, pp. 3739–3743, 1987.

[19] A. K. Nanda, N. Misra, P. Paul, and H. Singh, "Some properties of order statistics when the sample size is random," *Commun. Stat. Theory Methods*, vol. 34, no. 11, pp. 2105–2113, 2005.

[20] G. S. Lingappaiah, "Bayes prediction in exponential life-testing when sample size is a random variable," *IEEE Trans. Rel.*, vol. R-35, no. 1, pp. 106–110, Apr. 1986.

[21] M. Shaked and T. Wong, "Stochastic comparisons of random minima and maxima," *J. Appl. Probab.*, vol. 34, no. 2, pp. 420–425, Jun. 1997.

[22] T. Nakagawa and X. Zhao, "Optimization problems of a parallel system with a random number of units," *IEEE Trans. Rel.*, vol. 61, no. 2, pp. 543–548, Jun. 2012.

[23] E. H. Lloyd, "Least-squares estimation of location and scale parameters using order statistics," *Biometrika*, vol. 39, no. 1/2, pp. 88–95, Apr. 1952.

[24] A. G. Parameswaran, H. Garcia-Molina, H. Park, N. Polyzotis, A. Ramesh, and J. Widom, "CrowdScreen: Algorithms for filtering data with humans," in *Proc. 2012 ACM SIGMOD Int. Conf. Manage. Data (SIGMOD '12)*, May 2012, pp. 361–372.

[25] L. Tran-Thanh, M. Venanzi, A. Rogers, and N. R. Jennings, "Efficient budget allocation with accuracy guarantees for crowdsourcing classification tasks," in *Proc. 12th Int. Conf. Autonomous Agents Multiagent Syst. (AAMAS '13)*, May 2013, pp. 901–908.

[26] D. R. Karger, S. Oh, and D. Shah, "Budget-optimal task allocation for reliable crowdsourcing systems," *Oper. Res.*, vol. 62, no. 1, pp. 1–24, Jan.-Feb. 2014.

[27] X. Liu, H. He, and J. S. Baras, "Trust-aware optimal crowdsourcing with budget constraint," in *Proc. IEEE Int. Conf. Commun. (ICC 2015)*, Jun. 2015, pp. 1176–1181.

[28] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, NJ: Wiley-Interscience, 2003.

[29] S. S. Pradhan, R. Puri, and K. Ramchandran, "$n$-channel symmetric multiple descriptions—part I: $(n, k)$ source–channel erasure codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 47–61, Jan. 2004.

[30] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[31] A. Vempaty, L. R. Varshney, and P. K. Varshney, "Reliable crowdsourcing for multi-class labeling using coding theory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 4, pp. 667–679, Aug. 2014.