On the Impossibility of Convex Inference in Human Computation

Nihar B. Shah
U.C. Berkeley
nihar@eecs.berkeley.edu

Dengyong Zhou
Microsoft Research
dengyong.zhou@microsoft.com

Abstract

Human computation or crowdsourcing involves joint inference of the ground-truth-answers and the worker-abilities by optimizing an objective function, for instance, by maximizing the data likelihood based on an assumed underlying model. A variety of methods have been proposed in the literature to address this inference problem. As far as we know, none of the objective functions in existing methods is convex. In machine learning and applied statistics, a convex function such as the objective function of support vector machines (SVMs) is generally preferred, since it can leverage the high-performance algorithms and rigorous guarantees established in the extensive literature on convex optimization. One may thus wonder if there exists a meaningful convex objective function for the inference problem in human computation. In this paper, we investigate this convexity issue for human computation. We take an axiomatic approach by formulating a set of axioms that impose two mild and natural assumptions on the objective function for the inference. Under these axioms, we show that it is unfortunately impossible to ensure convexity of the inference problem. On the other hand, we show that interestingly, in the absence of a requirement to model "spammers", one can construct reasonable objective functions for crowdsourcing that guarantee convex inference.

1 Introduction

Human computation (or crowdsourcing) involves humans performing tasks which are generally difficult for computers to perform. Since humans may not have perfect abilities and sometimes may not even have good intentions, machine learning and statistical inference algorithms are typically employed to post-process the data obtained from the human workers in order to infer the true answers (Dawid and Skene 1979; Whitehill et al. 2009; Welinder et al. 2010; Raykar et al. 2010; Karger, Oh, and Shah 2011; Wauthier and Jordan 2011; McCreadie, Macdonald, and Ounis 2011; Luon, Aperjis, and Huberman 2012; Zhou et al. 2012; Liu, Peng, and Ihler 2012; Shah et al. 2013; Bachrach et al. 2012; Kamar, Hacker, and Horvitz 2012; Vempaty, Varshney, and Varshney 2013; Salek, Bachrach, and Key 2013;

Does the picture show a Bengal tiger ?

Figure 1: An example of a binary-choice task.

Shah et al. 2014; Matsui et al. 2013; Piech et al. 2013; Chen et al. 2013). These algorithms infer the true solutions to the tasks and the worker abilities by minimizing some carefully designed function that captures these parameters and their dependence on the observed responses. We shall call this function the "objective function". For example, it is typical to assume a generative model and take a maximum likelihood estimation. In such a case, the objective function is the negative-likelihood or the negative-log-likelihood of the data obtained from the workers, and Expectation-Maximization (EM) style optimization procedures are often employed to minimize this function (Dawid and Skene 1979; Whitehill et al. 2009; Welinder et al. 2010; Raykar et al. 2010; Liu, Peng, and Ihler 2012; Zhou et al. 2012; Chen et al. 2013).

To the best our knowledge, none of the objective functions in the literature for the inference problem in human computation is convex, and there are no generic guarantees available regarding obtaining a global optimum. Inspired by the extremely successful machine learning algorithms which are developed with convex objective functions (such as support vector machines (SVMs) (Vapnik 1998)), we investigate if there exists a meaningful convex objective function for the inference problem in human computation. Convex functions have appealing properties, and convex problems have been studied extensively in the literature, with high performance algorithms and rigorous guarantees available for very generic settings. It is thus of interest to investigate 'reasonable' models for human computation that ensure convexity in inference, thereby providing the ability to take advantage of this vast body of literature.

In this paper, we investigate the problem of convexity in

human computation via an axiomatic approach. Our results show that unfortunately, under two mild and natural assumptions in crowdsourcing, no model can guarantee convexity in the inference procedure. We show subsequently that all known models for crowdsourcing satisfy the two proposed mild axioms. The takeaway from this result is that it is futile to construct human-computation models, for present setups, that attempt to gain tractability by ensuring convexity. Finally, we show that interestingly, if one can forgo the explicit modelling of "spammers", it is indeed possible to construct 'reasonable' models for human computation that guarantee convexity in the objective for inference.

2 Problem Setting

Consider a binary-choice setting, where the worker must select from two given options for every question (e.g., Figure 1).[1] Suppose there are k workers, each of whom is assumed to have some latent 'ability'. Denote the latent ability of worker $i \in [k]$ as $w_i \in [W_{\min}, W_{\max}]$ for some $W_{\min} < W_{\max} \in \mathbb{R}$, with a higher value of w_i representing a more able worker.[2] We will use the value "0" to represent the ability of a *spammer* and assume that $0 \in [W_{\min}, W_{\max})$. A spammer is a worker who answers randomly with no regard to the question being asked. Spammers are known to exist in plentiful numbers among the worker pools on crowdsourcing platforms (Bohannon 2011; Kazai et al. 2011; Vuurens, de Vries, and Eickhoff 2011; Wais et al. 2010). Throughout the paper, we will restrict our attention to the (convex) subset $[0, W_{\max}]$ of the parameter space of the worker ability, since a function that is not convex on this set $[0, W_{\max}]$ will not be convex on the superset $[W_{\min}, W_{\max}]$ either. For convenience of notation, we define $\boldsymbol{w} := [w_1 \cdots w_k]^T \in [0, W_{\max}]^k$.

There are d questions, and each question has two choices, say, "0" and "1". This paper looks at procedures to infer, *via convex optimization*, the true answers $\boldsymbol{x}^* \in \{0, 1\}^d$ to these d questions and the worker abilities $\boldsymbol{w} \in [0, W_{\max}]^k$ from the responses received from the workers. However, an optimization problem over a discrete set is non-convex by definition. A convex optimization based procedure will thus relax the discrete domain of $\mathbf{x}^*$ to a continuous set. To this end, we associate every question $j \in [d]$ to a parameter $x_j \in [0, 1]$. The inference procedure operates on $\boldsymbol{x} := [x_1 \cdots x_d]^T \in [0, 1]^d$, and the inferred values may be subsequently quantized to obtain a solution in the discrete set $\{0, 1\}^d$.

Every question is asked to one or more of the k workers, and every worker is asked one or more of the d questions. We shall represent the workers' responses as a $\{0, 1, \infty\}$-valued matrix Y of size $(k \times d)$. For every $i \in [k]$ and

---

[1]Results on 'impossibility of convexity' for the binary-choice setting extend to more general settings (such as multiple-choice with more than two choices) by restricting attention to only two choices.

[2]We adopt the standard notation of representing the set $\{1, 2, \ldots, \alpha\}$ as $[\alpha]$ for any positive integer α, and representing the interval of the real line between (and including) β_1 and β_2 as $[\beta_1, \beta_2]$ for any two real numbers $\beta_1 < \beta_2$.

$j \in [d]$, we let $Y_{i,j}$ denote the $(i, j)^{\text{th}}$ element of the matrix Y. The value of $Y_{i,j}$ is set as ∞ if worker i was not asked question j, and is set as worker i's response to question j otherwise. The answers to the questions and the abilities of the workers are now inferred by minimizing some reasonable objective function L of $\boldsymbol{x}$ and $\boldsymbol{w}$ given the worker-responses Y:

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{w}}) = \underset{\boldsymbol{x} \in [0,1]^d, \boldsymbol{w} \in [0, W_{\max}]^k}{\arg\min} L(\boldsymbol{x}, \boldsymbol{w}; Y).$$

In the sequel, we will discuss the possible convexity of this optimization program under certain proposed axioms, and show that this optimization cannot be convex even after relaxing the domain of the answers from $\{0, 1\}^d$ to $[0, 1]^d$.

Upon completion of the inference procedure, an additional rounding step is often executed to convert $\hat{\boldsymbol{x}} \in [0, 1]^d$ to a discrete space $\hat{\boldsymbol{x}}^* \in \{0, 1\}^d$. This rounding may be performed via a deterministic approach, for instance, by quantizing x_j to 1 iff $\hat{x}_j > 0.5$, or via a probabilistic approach, for instance, by quantizing $\hat{x}_j$ to 1 with a probability $\hat{x}_j$, for every question $j \in [d]$. Many a times, however, the inferred vector $\hat{\boldsymbol{x}}$ is often left in the continuous set $[0, 1]^d$ as a "soft" output. In either case, the value $|\hat{x}_j - 0.5|$ is interpreted as the "confidence" associated with the inference of the answer to the jth question, with a higher value of $|\hat{x}_j - 0.5|$ implying a higher confidence in the answer inferred for question j. *A larger value of $\hat{x}_j$ indicates a greater confidence that x_j^* is 1. A higher value of $\hat{w}_i$ for any $i \in [k]$ is interpreted as a greater belief in the ability of worker i.*

3 Axiomatization of Objective Function

We take an axiomatic approach towards the design of the objective function L, and formulate two weak and natural axioms that the objective function L must satisfy. Recall that the inference procedure *minimizes* the objective function L, and hence the objective function should have a *lower* value when its arguments form a *better fit* to the observed data. While the axioms are stated in a general manner below, for an intuitive understanding one may think of a maximum likelihood approach with $L(\boldsymbol{x}, \boldsymbol{w}; Y)$ as the negative log likelihood of the observed data Y conditioned on $(\boldsymbol{x}, \boldsymbol{w})$.

First consider the case when there is only $d = 1$ question and $k = 1$ worker. The vectors $\boldsymbol{x}$ and $\boldsymbol{w}$ can be represented by scalars $x \in [0, 1]$ and $w \in [0, W_{\max}]$ respectively, and the matrix Y can be represented as a scalar in $\{0, 1\}$.

Axiom 1 (Distinguishing different worker abilities). *There exists $\epsilon > 0$ such that $L(x, w; 1) > L(x, 0; 1) \ \forall (x, w) \in (0, \epsilon) \times (0, \epsilon)$ and $L(x, w; 1) < L(x, 0; 1) \ \forall (x, w) \in (1 - \epsilon, 1) \times (0, \epsilon)$.*

Informally, Axiom 1 says that if the worker reports '1' when x is also close to 1, then the worker is likely to be more able. On the other hand, if the worker reports the opposite answer, then the worker is likely to be less able.

Axiom 2 (Modeling spammers). *The objective function $L(x, w; 1)$ is independent of x when $w = 0$.*

A spammer is a worker who answers randomly, without any regard to the question being posed. Spammers are

highly abundant in today's crowdsourcing systems, and pose a major challenge to the data collection as well as the inference procedures. Axiom 2 necessitates an explicit incorporation of a spammer into the parameter space.

One may also define analogous axioms for the function $L(\cdot, \cdot; 0)$, but for the purposes of this paper, it will suffice to work with solely the function $L(\cdot, \cdot; 1)$. Observe that we have *not* made any other assumptions on the function L such as continuity, differentiability, or Lipchitzness.

In the next section, we will show that no objective function L for inference in human computation that satisfies the two simple requirements identified above can be convex. We will also show that all existing inference techniques for crowdsourcing (that we are aware of) fall into this class. We will also subsequently demonstrate, by means of a constructive example, the interesting fact that in the absence of the requirement of modelling a spammer, the objective function can indeed be convex.

The three axioms listed above for the setting of $d = k = 1$ are translated to the general setting of $d \geq 1$ questions and $k \geq 1$ workers in the following manner. Consider the convex subset of the domain of parameters $\boldsymbol{x}$ and $\boldsymbol{w}$ that is given by $\boldsymbol{x} = x\mathbf{1}$ and $\boldsymbol{w} = w\mathbf{1}$, where $x \in [0,1], w \in [0, W_{\max}]$. Furthermore, suppose the observed data is $Y_{i,j} = 1$ for every (i, j) for which worker i is asked question j. In this case the objective function reduces to being a function of only the scalars x and w, as $L(x, w; 1)$, and we can now call upon the two axioms listed above. Now, if an optimization problem is non-convex over a convex subset of the parameter space, then it is non-convex over the entire parameter space as well. Thus, it suffices to show non-convexity for the case of $d = k = 1$ and the result for $d \geq 1$, $k \geq 1$ follows.

4   Impossibility of Convexity

Theorem 1 below proves the impossibility of models guaranteeing convex inference in crowdsourcing.

Theorem 1. *No function L satisfying Axiom 1 and Axiom 2 can be convex.*

Non-convex problems are often converted to convex problems via transformations of the variables involved. Our result however says that for any such transformation of the variables at hand, as long as the semantic meanings of the variables are retained, it is reasonable to expect the two aforementioned axioms to be satisfied, rendering the result of Theorem 1 applicable.

5   Some Examples of Existing Models

To the best of our knowledge, all existing models for inference in crowdsourcing tasks satisfy our two axioms. We illustrate the same with a few examples in this section.

While the two axioms of Section 3 were constructed to identify the precise cause of non-convexity, in this section, we propose three general properties that we would like any objective function for crowdsourcing to satisfy. As we will see later, satisfying these three properties automatically implies adherence to the two axioms. We will also show that

all existing models satisfy these three properties. As before, while the properties are defined for generic human-computation models, for an easier understanding one may think of the objective function L as the negative log likelihood function.

P1 (Monotonicity in accuracy of the answer). *$L(x, w; 1)$ is non-increasing in x.*

Property P1 says that the likelihood does not decrease if x is brought closer to the observed response "1 " (recall that we are considering only the interval $[0, W_{\max}]$ for w, where the latent ability of the worker is no worse than randomly answering).

P2 (Monotonicity in worker ability). *There exists some non-decreasing function $g : [0, W_{\max}) \rightarrow [0.5, 1)$ such that $L(x, w; 1)$ increases with w when $x < g(w)$ and $L(x, w; 1)$ decreases with an increase in w when $x > g(w)$.*

Recall that for an inferred answer $\hat{x}$, the value $|\hat{x} - 0.5|$ represents the confidence associated to the inference. A higher confidence in the inference directly relates to a greater belief in the ability of a worker. Property P2 formalizes this relation, with function $g(w)$ capturing the confidence associated to the work of a worker with ability w. The likelihood is thus higher when the confidence associated to the work of the worker is closer to the corresponding confidence in the inferred answer.

P3 (Modeling spammers). *The objective function $L(x, w; 1)$ is independent of x when $w = 0$.*

Property P3 is identical to Axiom 2.

Proposition 2. *Any objective function L that satisfies Property P2 also satisfies Axiom 1.*

We now present examples of existing models for crowdsourcing and show that these models indeed satisfy the three properties listed above (and therefore the two axioms defined in Section 3). Throughout this section, we will let n denote the number of responses received from the workers, i.e, the number of $\{0, 1\}$-valued entries in the response matrix Y.

Example 1 (Dawid-Skene model).
Model: The Dawid-Skene model is one of the most popular models for crowdsourcing (Dawid and Skene 1979; Ipeirotis, Provost, and Wang 2010; Gao and Zhou 2013; Zhang et al. 2014; Karger, Oh, and Shah 2011; Dalvi et al. 2013; Ghosh, Kale, and McAfee 2011). The model assumes that the ability of a worker represents the probability of her correctly answering any individual question, i.e., if worker $i \in [k]$ is asked question $j \in [d]$, then she will give the correct answer with a probability p_i and an incorrect answer with probability $(1 - p_i)$, for some parameter $p_i \in [0, 1]$ whose value is unknown. The response of any worker to any question is independent of all else. In order to ensure that the model is identifiable, it is typically also assumed that $p_i \in [0.5, 1] \,\forall i \in [k]$, or $\frac{1}{k} \sum_{i=1}^{k} p_i \geq 0.5$, or simply $p_1 \geq 0.5$. We shall restrict our attention to $p_i \in [0.5, 1] \,\forall\, i \in [k]$. Further, we will work with a shifted and scaled version of p_i's by defining $w_i = 2p_i - 1 \,\forall\, i \in [k]$. Under this transformation, we have $W_{\min} = 0$ and $W_{\max} = 1$.

Inference: Consider inferring $\boldsymbol{x}$ and $\boldsymbol{w}$ via maximum likelihood estimation. Observe that the likelihood of observation Y is

$$\prod_{(i,j):Y_{i,j}\neq\infty}\left(\frac{1+w_i}{2}\right)^{Y_{i,j}x_j+(1-Y_{i,j})(1-x_j)}$$
$$\times\left(\frac{1-w_i}{2}\right)^{Y_{i,j}(1-x_j)+(1-Y_{i,j})x_j}.$$

The negative log-likelihood is thus given by

$$L(\boldsymbol{x},\boldsymbol{w};Y):=-\sum_{\substack{(i,j):\\Y_{i,j}\neq\infty}}\Bigg((Y_{i,j}x_j+(1-Y_{i,j})(1-x_j))\log\left(\frac{1+w_i}{2}\right)$$
$$+(Y_{i,j}(1-x_j)+(1-Y_{i,j})x_j)\log\left(\frac{1-w_i}{2}\right)\Bigg),$$

and this function is minimized over $(\boldsymbol{x},\boldsymbol{w})\in[0,1]^{d+k}$.

Consider the following subspace of the arguments: $\boldsymbol{x}=x\mathbf{1}$ and $\boldsymbol{w}=w\mathbf{1}$ for some $x\in[0,1]$ and $w\in[0,1]$. Further, suppose $Y_{i,j}=1$ for every (i,j) for which worker i was asked question j. Under this restriction, the objective function reduces to

$$L(x,w;1)=-n\left(x\log\left(\frac{1+w}{2}\right)+(1-x)\log\left(\frac{1-w}{2}\right)\right)\ (1)$$

Properties: Let us understand what the three properties listed earlier mean in this context. Observe that

$$L(x,0;1)=-n(x\log 0.5+(1-x)\log 0.5)$$
$$=-n\log 0.5,$$

and hence $L(x,w;1)$ is independent of x when $w=0$. Thus $w=0$ models a spammer, and the function L obeys Property P3. For Property P1, observe that when $w\in[0,1]$,

$$\frac{\partial}{\partial x}L(x,w;1)=-n\left(\log\left(\frac{1+w}{2}\right)-\log\left(\frac{1-w}{2}\right)\right)$$
$$\leq 0.$$

Thus Property P1 is satisfied. Finally,

$$\frac{\partial}{\partial w}L(x,w;1)=-n\left(\frac{x}{1+w}-\frac{1-x}{1-w}\right),$$

which is positive when $2x-1<w$ and negative when $2x-1>w$. Property P2 is thus satisfied with $g(w)=\frac{1+w}{2}$.

The objective function (1) is plotted in Figure 2.

Example 2 (Two-coin Dawid-Skene model)**.**
The two-coin model associates the ability of every worker $i\in[k]$ to two latent variables $p_{i,0}\in[0.5,1]$ and $p_{i,1}\in[0.5,1]$. Under the two-coin model, the probability with which a worker answers a question correctly depends on the true answer to that question: if the true answer to a question is $x^*\in\{0,1\}$ then the worker (correctly) provides x^* as the answer with probability p_{i,x^*} and (incorrectly) provides $(1-x^*)$ as the answer otherwise, independent of all else. In order to connect to our theory, we simply restrict our attention to a (convex) subset of the parameters $\{p_{i,0},p_{i,1}\}_{i=1}^k$ obtained by setting $p_{i,0}=p_{i,1}\ \forall\ i\in[k]$. The resulting model is identical to the one-coin model discussed earlier.
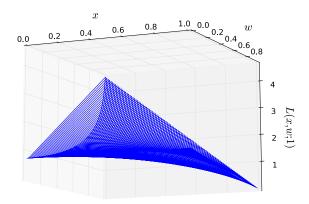

Figure 2: The objective function for maximum-likelihood inference under the Dawid-Skene model.

Example 3 (Additive Noise)**.**
Model: The additive noise model assumes that when worker $i\in[k]$ is asked question $j\in[d]$, the response $Y_{i,j}$ of the worker is given by

$$Y_{i,j}=\mathbb{1}\{w_i(x_j-0.5)+\epsilon_{i,j}>0\},$$

where $\{\epsilon_{i,j}\}_{i\in[d],j\in[k]}$ is a set of i.i.d. random variables with some (known) c.d.f. F. The function F is non-constant in the domain of interest $[-0.5W_{\max},0.5W_{\max}]$. The response is assumed to be independent of all other questions and all other workers. A common choice for F is the c.d.f. of the Gaussian distribution (Piech et al. 2013; Thurstone 1927; Welinder et al. 2010).

Inference: The inference is usually performed by minimizing the negative log likelihood of the observed data Y:

$$\underset{\boldsymbol{x},\boldsymbol{w}}{\arg\min}\ -\sum_{(i,j):Y_{i,j}\neq\infty}(Y_{i,j}\log(1-F(-(x_j-0.5)w_i))$$
$$+(1-Y_{i,j})\log(F(-(x_j-0.5)w_i)).$$

Properties: Let us now relate this model to the three properties enumerated earlier. Let us restrict our attention to the (convex) subset of the parameters where $\boldsymbol{x}=x\mathbf{1}$ and $\boldsymbol{w}=w\mathbf{1}$ for some $x\in[0,1]$ and $w\in[0,W_{\max}]$. Suppose $Y_{i,j}=1$ for every (i,j) for which worker i was asked question j. The objective function then reduces to

$$L(x,w;1)=-n\log(1-F(-(x-0.5)w)).$$

One can verify that this function is non-increasing in x (whenever $w\geq 0$), thereby satisfying Property P1. Furthermore, setting $g(w)=0.5$ satisfies Property P2. A spammer is modeled by the parameter value $w=0$, in which case, the function L ceases to be dependent on x.

Example 4 (Minimax Entropy Model)**.**
Model: The minimax entropy model (Zhou et al. 2012) hypothesizes that when worker $i\in[k]$ answers question $j\in[d]$, she provides "1" as the answer with a probability $\pi_{i,j}$ and "0" otherwise, independent of all else, for some unknown value $\pi_{i,j}\in[0,1]$. Under the 'minimax entropy principle' proposed therein, the set $\{\pi_{i,j}\}_{i\in[k],j\in[d]}$ has the maximum entropy under the constraints imposed by the set

of true answers and the observed data, and the true answers minimize this value of maximum entropy:

$$\min_{\boldsymbol{x}} \max_{\{\pi_{i,j}\}_{i \in [k], j \in [d]}} -\sum_{i=1}^{k}\sum_{j=1}^{d} \pi_{i,j} \ln \pi_{i,j}$$

$$\text{s.t.} \sum_{i=1}^{k} \pi_{i,j} = \sum_{i=1}^{k} Y_{i,j} \ \forall j \in [k]$$

$$\sum_{j=1}^{d} x_j \pi_{i,j} = \sum_{j=1}^{d} x_j Y_{i,j} \ \forall i \in [k]$$

$$\sum_{j=1}^{d}(1 - x_j)\pi_{i,j} = \sum_{j=1}^{d}(1 - x_j)Y_{i,j} \ \forall i \in [k]$$

$$0 \le \pi_{i,j}, x_j \le 1 \ \forall i \in [k], \ j \in [d].$$

Inference: The authors show that the values $\pi_{i,j}$ must necessarily be of the form

$$\pi_{i,j} = \frac{\exp((1-x_j)(\tau_{j,1}+\sigma_{i,1,0}) + x_j(\tau_{j,1}+\sigma_{i,1,1}))}{\sum_{\ell=0}^{1}\exp((1-x_j)(\tau_{j,\ell}+\sigma_{i,\ell,0}) + x_j(\tau_{j,\ell}+\sigma_{i,\ell,1}))}$$

for some parameters $\{\sigma_{i,\ell_1,\ell_2}\}_{i \in [k], \ell_1 \in \{0,1\}, \ell_2 \in \{0,1\}}$ and $\{\tau_{j,\ell}\}_{j \in [d], \ell \in \{0,1\}}$. The authors then propose minimization (with respect to variables $\{x_i, \sigma_{i,\ell_1,\ell_2}, \tau_{j,\ell}\}$) of the dual of the aforementioned program, which they derive to be of the form

$$-\sum_{\substack{(i,j): \\ Y_{i,j} \ne \infty}} \frac{\exp\big((1-x_j)(\tau_{j,Y_{i,j}}+\sigma_{i,Y_{i,j},0}) + x_j(\tau_{j,Y_{i,j}}+\sigma_{i,Y_{i,j},1})\big)}{\sum_{\ell=0}^{1}\exp((1-x_j)(\tau_{j,\ell}+\sigma_{i,\ell,0}) + x_j(\tau_{j,\ell}+\sigma_{i,\ell,1}))}$$

Properties: Consider the following (convex) subset of the parameter space: $\tau_{j,\ell} = 0 \ \forall j \in [d], \ell \in \{0,1\}$, $\sigma_{i,0,0} = \sigma_{i,1,1} = 1 - \sigma_{i,1,0} = 1 - \sigma_{i,0,1} := 0.5 + w_i \in [0,1] \ \forall i \in [k]$. The minimization program can now be rewritten as

$$\arg\min_{\boldsymbol{x},\boldsymbol{w} \in [0,1]^{d+k}} \sum L(x_i, w_j; Y_{i,j})$$

where

$$L(x,w;1) :=$$
$$\frac{e^{(1-x)(0.5-w)+x(0.5+w)}}{e^{(1-x)(0.5-w)+x(0.5+w)} + e^{x(0.5-w)+(1-x)(0.5+w)}},$$

and $L(x,w;0) = L(1-x,w;1)$. With some algebraic manipulations, one can verify that this function $L(\cdot,\cdot;1)$ satisfies Property P1, Property P2 (with $g(w) = 0.5$) and Property P3 (with $w = 0$ representing a spammer).

Example 5 (GLAD model).
Model and inference: The GLAD model was introduced in (Whitehill et al. 2009). We will restrict attention to the subspace of the parameter set which has, in the notation of (Whitehill et al. 2009), $\beta_j = 0 \ \forall \ j$. Using the notation of the present paper for the rest of the parameters, the objective function (the negative log likelihood) is

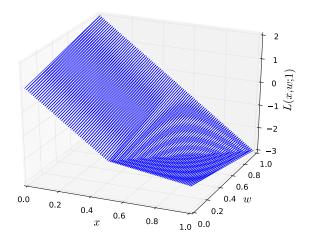$$L(x,w;1) = nx \log\left(1 + e^{-w}\right) + n(1-x)\log\left(1 + e^{w}\right)$$

Figure 3: An objective function that satisfies properties P1 and P2, is convex, but does not incorporate the modeling of a spammer. This function is defined in (2).

under the convex subset $\boldsymbol{x} = x\mathbf{1}$ and $\boldsymbol{w} = w\mathbf{1}$ of the parameters, wth $x \in [0,1], w \in [0,1]$.

Properties: The derivative of $L(x,w;1)$ with respect to x is non-positive (for $w \ge 0$), thus satisfying Property P1. One can also verify that Property P2 is satisfied with $g(w) = \frac{1}{1+e^{-w}}$. Finally, we have $L(x,0;1) = n\log 2$, thereby satisfying Property P3.

6 If Not Modelling Spammers

We now discuss the role of modeling a spammer, i.e., Axiom 2. We show that ignoring this axiom indeed allows for convex objective functions for crowdsourcing, not only satisfying Axiom 1, but also satisfying the stronger properties P1 and P2.

Theorem 3. *Let $W_{\min} = 0$ and $W_{\max} = 1$. The function $L : [0,1] \times [0,1] \to \mathbb{R}$ defined as*

$$L(x,w;1) = \begin{cases} -w - x - 1 & \text{if } w \le 2x - 1 \\ w - 5x + 1 & \text{if } w \ge 2x - 1 \end{cases} \quad (2)$$

satisfies properties P1 and P2 (and hence Axiom 1) and is (jointly) convex in its two arguments.

For the general setting of multiple workers and multiple questions, letting

$$L(x,w;0) = L(1-x,w;1),$$

and

$$L(\boldsymbol{x},\boldsymbol{w};Y) = \sum_{(i,j):Y_{i,j}\ne\infty} L(x_i, w_j; Y_{ij})$$

ensures that the function $L(\cdot,\cdot;Y)$ satisfies Axiom 1 and is convex in its arguments $(\boldsymbol{x}, \boldsymbol{w})$. The objective function $L(\cdot,\cdot;1)$ constructed in Theorem 3 is plotted in Figure 3.

Note that we do not intend to claim the proposed function (2) as a "good" objective function to use for crowdsourcing. Instead, the takeaway from this section is that if one forgoes the inclusion of spammers in the objective, then one may indeed be able to design a crowdsourcing model that is reasonable and permits convex inference.

7 Discussion

It is important to be aware of the limitations of the framework of this paper. Throughout the paper we assumed no prior knowledge or complexity controls on the parameter space. One may alternatively consider the inference problem in a Bayesian setting with non-uniform priors, or impose some convex regularization. In fact, if the regularizer is strictly convex, then giving it a sufficiently large weight can make the objective function convex, albeit perhaps at the expense of the model not capturing certain essential features of the problem. However, as long as the objective function continues to satisfy the two axioms presented here, the conclusions drawn in this paper continue to apply.

While convexity is certainly desirable, absence of convexity certainly does not mean the complete absence of guarantees; indeed, there is a line of recent works (Loh and Wainwright 2013; Netrapalli, Jain, and Sanghavi 2013; Zhang et al. 2014) which provide guarantees for non-convex problems as well. In particular, although existing models for human computation are not convex, there exist theoretical guarantees on inference under the popular Dawid-Skene model to a certain extent (Ghosh, Kale, and McAfee 2011; Karger, Oh, and Shah 2011; Dalvi et al. 2013; Gao and Zhou 2013; Zhang et al. 2014). For instance, (Zhang et al. 2014) show that the EM algorithm for the Dawid-Skene model can achieve a minimax rate up to a logarithmic factor when it is appropriately initialized by spectral methods. However, these results need certain conditions which may not hold in real scenarios. Moreover, algorithms that are minimax optimal may not always work well in practice, for instance, see the experimental results in (Liu, Peng, and Ihler 2012). Most importantly, guarantees for non-convex problems are constructed to-date on a case-by-case basis (for example, all known theoretical guarantees for crowdsourcing are for the Dawid-Skene model alone). These guarantees do not allow for a convenient application of the theory to any new model. On the other hand, although the theory of convex optimization is highly generic and extensive, the results of this paper imply that unfortunately one cannot readily exploit this theory in the context of human computation.

This paper also shows that a willingness to forgo the explicit incorporation of spammers into the crowdsourcing model indeed allows for reasonable objective functions guaranteeing convex inference. Successful deterrence of spammers in crowdsourcing systems, for instance, by designing suitable reward mechanisms, may thus expand the scope of model-design for human computation. In conclusion, we would like to enumerate, partially in jest, the problems resulting from spammers in crowdsourcing systems: (a) low-quality work, (b) depletion of the monetary budget, and now, (c) prevention of models guaranteeing convex inference.

Appendix: Proofs

Proof of Theorem 1: The proof employs a contradiction-based argument. Suppose there exists some function $L(\cdot, \cdot; 1)$ that satisfies the two axioms and is convex in its two arguments in the set $[0, 1] \times [0, W_{\max}]$. Without loss of generality assume $\epsilon \in (0, \min\{0.5, W_{\max}\})$. Axiom 1 mandates

$$L(\epsilon - \epsilon^2/2, \epsilon^2/2; 1) > L(\epsilon - \epsilon^2/2, 0; 1) \qquad \text{and} \qquad (3)$$
$$L(1 - \epsilon/2, \epsilon/2; 1) < L(1 - \epsilon/2, 0; 1). \qquad (4)$$

The assumed convexity of function L implies

$$(1-\epsilon)L(0,0;1)+\epsilon L(1-\epsilon/2,\epsilon/2;1) \geq L(\epsilon-\epsilon^2/2,\epsilon^2/2;1).$$

Substituting (3) and (4) in this inequality gives

$$(1 - \epsilon)L(0, 0; 1) + \epsilon L(1 - \epsilon/2, 0; 1) > L(\epsilon - \epsilon^2/2, 0; 1).$$

Now, calling upon Axiom 2 gives

$$(1 - \epsilon)L(0, 0; 1) + \epsilon L(0, 0; 1) > L(0, 0; 1),$$

yielding the desired contradiction. $\qquad \square$

Proof of Proposition 2: Pick a value δ arbitrarily in $(0, \min\{1, W_{\max}\}/2)$. Set

$$\epsilon = \min\{\delta, 1 - g(\delta)\}.$$

Observe that due to Property P2, $\forall\, x < 0.5$ (which includes all $x < \epsilon$), $L(x, w; 1)$ increases with an increase in w, and hence $L(x, w; 1) > L(x, 0; 1)$ when $w > 0$. Also, for our chosen ϵ, Property P2 implies that for any $x \in (1 - \epsilon, 1)$ and $w < \epsilon$, the function $L(x, w; 1)$ decreases with an increase in w. Thus we have $L(x, w; 1) < L(x, 0; 1)\ \forall\ (x, w) \in (1 - \epsilon, 1) \times (0, \epsilon)$. Axiom 1 is thus satisfied. $\qquad \square$

Proof of Theorem 3: Let us first verify that the proposed function $L(\cdot, \cdot; 1)$ satisfies the two properties. First, observe from the definition of L in (2) that L is always (strictly) decreasing in its first argument, and hence satisfies Property P1. Towards Property P2, set

$$g(w) = (1 + w)/2.$$

By definition, $L(x, w; 1)$ (strictly) decreases with an increase in w when $x > g(w)$, and (strictly) increases with an increase in w when $x < g(w)$. Proposition 2 now guarantees that the function also satisfies Axiom 1.

Let us now investigate the convexity of this function. Consider two hyperplanes H_0 and H_1 defined as

$$H_0(x, w) = -w - x - 1$$
$$H_1(x, w) = w - 5x + 1.$$

Observe that for any (x, w),

$$H_0(x, w) - H_1(x, w) = 2(-w + 2x - 1).$$

Thus, $H_0(x, w) \geq H_1(x, w)$ if $x \geq (1 + w)/2$ and $H_0(x, w) \leq H_1(x, w)$ if $x \leq (1 + w)/2$. It follows that

$$L(x, w; 1) = \max\{H_0(x, w), H_1(x, w)\}$$

meaning that L is the maximum of two linear functions. Hence L is convex. $\qquad \square$

References

Bachrach, Y.; Graepel, T.; Minka, T.; and Guiver, J. 2012. How to grade a test without knowing the answers—a Bayesian graphical model for adaptive crowdsourcing and aptitude testing. In *ICML*.

Bohannon, J. 2011. Social science for pennies. *Science* 334(6054):307–307.

Chen, X.; Bennett, P. N.; Collins-Thompson, K.; and Horvitz, E. 2013. Pairwise ranking aggregation in a crowdsourced setting. In *ACM international conference on Web search and data mining*, 193–202.

Dalvi, N.; Dasgupta, A.; Kumar, R.; and Rastogi, V. 2013. Aggregating crowdsourced binary ratings. In *International conference on World Wide Web*, 285–294.

Dawid, A. P., and Skene, A. M. 1979. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Applied statistics* 20–28.

Gao, C., and Zhou, D. 2013. Minimax optimal convergence rates for estimating ground truth from crowdsourced labels. *arXiv preprint arXiv:1310.5764*.

Ghosh, A.; Kale, S.; and McAfee, P. 2011. Who moderates the moderators?: Crowdsourcing abuse detection in user-generated content. In *ACM conference on Electronic commerce*, 167–176. ACM.

Ipeirotis, P. G.; Provost, F.; and Wang, J. 2010. Quality management on Amazon Mechanical Turk. In *ACM SIGKDD workshop on human computation*, 64–67. ACM.

Kamar, E.; Hacker, S.; and Horvitz, E. 2012. Combining human and machine intelligence in large-scale crowdsourcing. In *International Conference on Autonomous Agents and Multiagent Systems*, 467–474.

Karger, D. R.; Oh, S.; and Shah, D. 2011. Budget-optimal crowdsourcing using low-rank matrix approximations. In *Allerton Conference on Communication, Control, and Computing*, 284–291.

Kazai, G.; Kamps, J.; Koolen, M.; and Milic-Frayling, N. 2011. Crowdsourcing for book search evaluation: impact of HIT design on comparative system ranking. In *ACM SIGIR conference on Research and development in Information Retrieval*, 205–214.

Liu, Q.; Peng, J.; and Ihler, A. T. 2012. Variational inference for crowdsourcing. In *NIPS*, 701–709.

Loh, P.-L., and Wainwright, M. J. 2013. Regularized m-estimators with nonconvexity: Statistical and algorithmic theory for local optima. In *Advances in Neural Information Processing Systems*, 476–484.

Luon, Y.; Aperjis, C.; and Huberman, B. A. 2012. Rankr: A mobile system for crowdsourcing opinions. In *Mobile Computing, Applications, and Services*. Springer. 20–31.

Matsui, T.; Baba, Y.; Kamishima, T.; and Kashima, H. 2013. Crowdsourcing quality control for item ordering tasks. In *First AAAI Conference on Human Computation and Crowdsourcing*.

McCreadie, R.; Macdonald, C.; and Ounis, I. 2011. Crowdsourcing blog track top news judgments at trec. In *Workshop on crowdsourcing for search and data mining at the ACM international conference on web search and data mining (WSDM)*, 23–26.

Netrapalli, P.; Jain, P.; and Sanghavi, S. 2013. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems*, 2796–2804.

Piech, C.; Huang, J.; Chen, Z.; Do, C.; Ng, A.; and Koller, D. 2013. Tuned models of peer assessment in MOOCs. *arXiv preprint arXiv:1307.2579*.

Raykar, V. C.; Yu, S.; Zhao, L. H.; Valadez, G. H.; Florin, C.; Bogoni, L.; and Moy, L. 2010. Learning from crowds. *Journal of Machine Learning Research* 11:1297–1322.

Salek, M.; Bachrach, Y.; and Key, P. 2013. Hotspotting–a probabilistic graphical model for image object localization through crowdsourcing. In *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Shah, N. B.; Bradley, J. K.; Parekh, A.; Wainwright, M.; and Ramchandran, K. 2013. A case for ordinal peer-evaluation in MOOCs. In *NIPS Workshop on Data Driven Education*.

Shah, N. B.; Balakrishnan, S.; Bradley, J.; Parekh, A.; Ramchandran, K.; and Wainwright, M. 2014. When is it better to compare than to score? *arXiv preprint arXiv:1406.6618*.

Thurstone, L. 1927. A law of comparative judgment. *Psychological review* 34(4):273.

Vapnik, V. N. 1998. *Statistical learning theory*. Wiley, NY.

Vempaty, A.; Varshney, L. R.; and Varshney, P. K. 2013. Reliable classification by unreliable crowds. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5558–5562.

Vuurens, J.; de Vries, A. P.; and Eickhoff, C. 2011. How much spam can you take? An analysis of crowdsourcing results to increase accuracy. In *ACM SIGIR Workshop on Crowdsourcing for Information Retrieval*, 21–26.

Wais, P.; Lingamneni, S.; Cook, D.; Fennell, J.; Goldenberg, B.; Lubarov, D.; Marin, D.; and Simons, H. 2010. Towards building a high-quality workforce with Mechanical Turk. *NIPS workshop on computational social science and the wisdom of crowds*.

Wauthier, F. L., and Jordan, M. 2011. Bayesian bias mitigation for crowdsourcing. In *NIPS*, 1800–1808.

Welinder, P.; Branson, S.; Perona, P.; and Belongie, S. J. 2010. The multidimensional wisdom of crowds. In *Advances in neural information processing systems*, 2424–2432.

Whitehill, J.; Ruvolo, P.; Wu, T.-f.; Bergsma, J.; and Movellan, J. 2009. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in neural information processing systems*, 2035–2043.

Zhang, Y.; Chen, X.; Zhou, D.; and Jordan, M. I. 2014. Spectral methods meet EM: A provably optimal algorithm for crowdsourcing. In *Advances in neural information processing systems*.

Zhou, D.; Platt, J.; Basu, S.; and Mao, Y. 2012. Learning from the wisdom of crowds by minimax entropy. In *Advances in Neural Information Processing Systems 25*, 2204–2212.