an annotator $i$, the probability of its agreeing (or disagreeing) with hypotheses $c_t$ and $h$ (having error rate $\epsilon$) is given by a probability tree shown in Figure 1. From the tree, it can be seen that $\mathbb{P}(m_1, \ldots, m_n)[E_t(h)]$ is same as the probability that the number of samples that fall under leaf B in the probability tree is at most the number of samples that fall under leaf A.

To compute the above quantity, first we compute the conditional probability, $\mathbb{P}(k_1, \ldots, k_n)(L_e(h) \leq L_e(c_t))$, defined as: if $k_i$ examples, $(0 \leq k_i \leq m_i)$, from annotator $i (i = 1, \ldots, n)$ come from the set $(c_t \Delta h)$, then the probability that empirical error of $h$ (given by $L_e(h)$) is less than or equal to empirical error of $c_t$ (given by $L_e(c_t)$).

Consider the random variable $Z_i^j$, $i = 1, \ldots, n$ $j = 1, \ldots, k_i$, which is the indicator of whether $j^{th}$ sample from $i^{th}$ annotator is from leaf node B, given that all the data points are from $c_t \Delta h$ region. Thus, $\mathbb{P}(Z_i^j = 1) = (1 - \eta_i)$ and $\mathbb{P}(Z_i^j = 0) = \eta_i$. Let $Z = \sum_{i=1}^{n} \sum_{j=1}^{k_i} Z_i^j$. Then the event $L_e(h) \leq L_e(c_t)$ is same as the event $Z \leq \sum_{i=1}^{n} k_i / 2$. Hence, we are interested in finding an upper bound on $\mathbb{P}(Z \leq \sum_{i=1}^{n} k_i / 2)$. We can use the multiplicative form of Chernoff bound (see e.g. Theorem 4.2 in [Motwani and Raghavan, 1995]), which says $\mathbb{P}(Z \leq (1 - \nu) \mu) \leq \exp(-\mu \nu^2 / 2)$, where $\mu = \mathbb{E}[Z] = \sum_{i=1}^{n} (1 - \eta_i) k_i$. Hence, by letting $\nu = \sum_{i=1}^{n} k_i (1 - 2 \eta_i) / 2 \sum_{i=1}^{n} k_i (1 - \eta_i)$, we get the following bound:

$$\mathbb{P}(k_1, \ldots, k_n)(L_e(h) \leq L_e(c_t)) \leq e^{- \frac{\sum_{i=1}^{n} k_i (1 - 2 \eta_i)}{2 \sum_{i=1}^{n} k_i (1 - \eta_i)}}$$

Simplifying this bound, for $0 \leq \eta_i \leq 1/3$ we get:

$$\mathbb{P}(k_1, \ldots, k_n)(L_e(h) \leq L_e(c_t)) \leq e^{- \frac{\sum_{i=1}^{n} k_i (1 - 3 \eta_i)}{8 \sum_{i=1}^{n} k_i (1 - \eta_i)}} \quad (5)$$

Summing up the above conditional probability bound over all possible values of $k_i$, the total probability $\mathbb{P}(m_1, \ldots, m_n)[E_t(h)]$ becomes:

$$\sum_{k_1=0}^{m_1} \cdots \sum_{k_n=0}^{m_n} \prod_{i=1}^{n} \bigg( \binom{m_i}{k_i} e^{k_i (1 - \epsilon)^{m_i - k_i}} \bigg)^{\mathbb{P}(k_1, \ldots, k_n)(L_e(h) \leq L_e(c_t))} \quad (6)$$

Using the bound in (5), we get the following upper bound on $\mathbb{P}(m_1, \ldots, m_n)[E_t(h)]$:

$$\prod_{i=1}^{n} \left( \sum_{k_i=0}^{m_i} \binom{m_i}{k_i} e^{k_i (1 - \epsilon)^{m_i - k_i}} \exp \left( -k_i (1 - 3 \eta_i) / 8 \right) \right)$$

Using the moment generating function of the Binomial distribution, the bound becomes

$$\prod_{i=1}^{n} \left[ 1 - \epsilon (1 - \exp(-(1 - 3 \eta_i) / 8)) \right]^{m_i}$$

In above bound, $h$ has an error rate exactly equal to $\epsilon$. However, this bound is valid for an $\epsilon$-bad hypothesis also because the expression decrease as $\epsilon$ increases. Substituting this upper bound on the LHS of (4), we get the desired claim. Q.E.D.

Note that the Theorem 1 is valid only for the range of $0 < \eta_i < 1/3$. However, we can extend the definition of $\psi(\cdot)$ to the boundary points in a manner that the same relation (3) holds true. For this, observe that minimum number of examples required from a single non-noisy annotator would be $m_0 = \log(N/\delta) / \log(1 - \epsilon)^{-1}$. This is because in such a case, we have $\mathbb{P}(L_e(h) \leq L_e(c_t) \mid x \in c_t \Delta h) = 0$ and $\mathbb{P}(x \notin c_t \Delta h) = (1 - \epsilon)$. Hence, we can let $\psi(0) = \log(1 - \epsilon)^{-1}$. Also, we let $\psi(1/3) = \log(1 - \epsilon (1 - \exp(-(1/18))))^{-1}$ and $m_{1/3} = \log(N/\delta) / \psi(1/3)$ from [Laird, 1988].

3 Cost Optimal Mechanism Design for PAC Learning

We consider the problem of procuring a feasible annotation plan when the learner needs to pay annotators for their efforts, under known and unknown noise rate scenarios. In the unknown noise rate scenario, we propose an auction model and present an optimal auction mechanism.

We assume that each annotator $i$ (with noise rate $\eta_i$) incurs an internal cost $c(\eta_i)$ of annotation for labeling one data point; note that the cost is dependent on the noise rate, and the cost function is same for all the annotators. The cost function is assumed to be a bounded, continuously differentiable, strictly decreasing function in $0 \leq \eta_i < 1/2 \forall i = 1, \ldots, n$. If an annotator is more competent (i.e. less noisy) then he can make more money by selling his services and time to somewhere else in the market, which translates to saying that his internal cost of annotation is high.

Consider a simplistic scenario of complete information where the learner knows noise rates of all the annotators. In such a case, the goal of the learner is purchase an annotation plan $m =$