Comprehensive and Reliable Crowd Assessment Algorithms

Manas Joglekar, Hector Garcia-Molina, Aditya Parameswaran

Introduction

- **Crowdsourcing**: Using human workers to perform tasks that are hard for computers. E.g. Image tagging
- **Our Problem**: Measure worker accuracy, and compute *calibrated* confidence intervals.
- **General Setting**: Any number of workers. May or may not have attempted every task. Non-binary questions. No gold standard data
- **Need for confidence Intervals**: Worker one makes 1 error in 3 tasks, worker two makes 10 errors in 30 tasks
  - Equal expected error rates (1/3), but worker 2 is more likely to be very bad

	T1	T2	T3	T4	T5
W1	Y	Y	-	N	N
W2	N	-	Y	N	N
W3	-	Y	Y	N	Y
True	N	Y	Y	N	N

2/3 Agreements
$E[q_{12}] = 2/3$

Compute Agreements
- $q_{ij} \sim p_i p_j + (1-p_i)(1-p_j)$
- Solve to get p_i

Compute Variances
- $Cov(q_{ij}, q_{ik})$
- $Cov(q_{ij}, q_{ij})$

Compute Derivatives
- dp_i/dq_{ij}
- dp_i/dq_{jk}

Compute Intervals
- p_i mean
- p_i variance
- Confidence interval size

For N > 3 workers, form N^2 groups of 3 workers

Find group-wise estimates, variances. Inter-group covariances

Combine group estimates using optimized weights

Calibration Results :