# Information Elicitation Mechanisms for Statistical Estimation

Yuqing Kong, Peking University
Grant Schoenebeck, Biaoshuai Tao, Fang-Yi Yu, University of Michigan

## Motivation Question

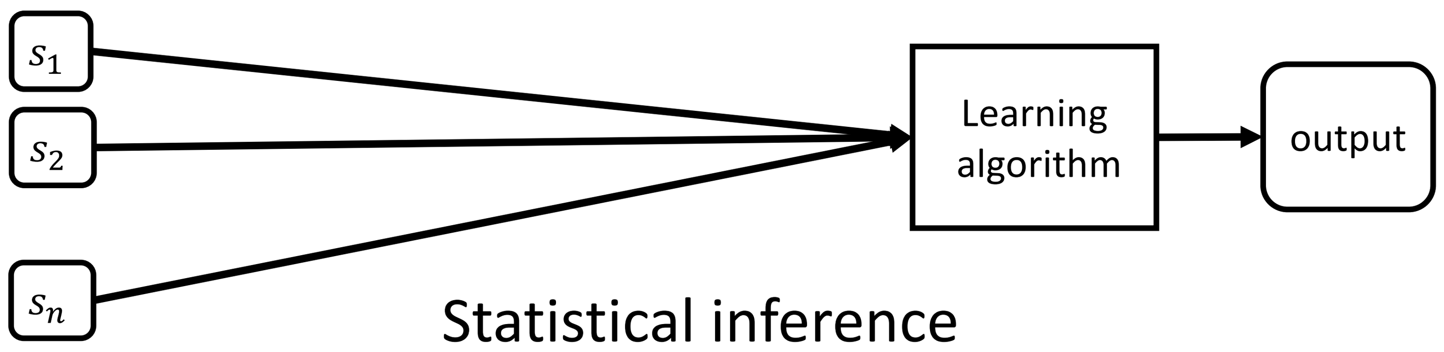
How can we do statistical estimation via crowdsourcing if workers are

- Selfish and want to maximize their revenue
- Bayesian with complicated signal structure

## Statistical inference and Crowdsourcing

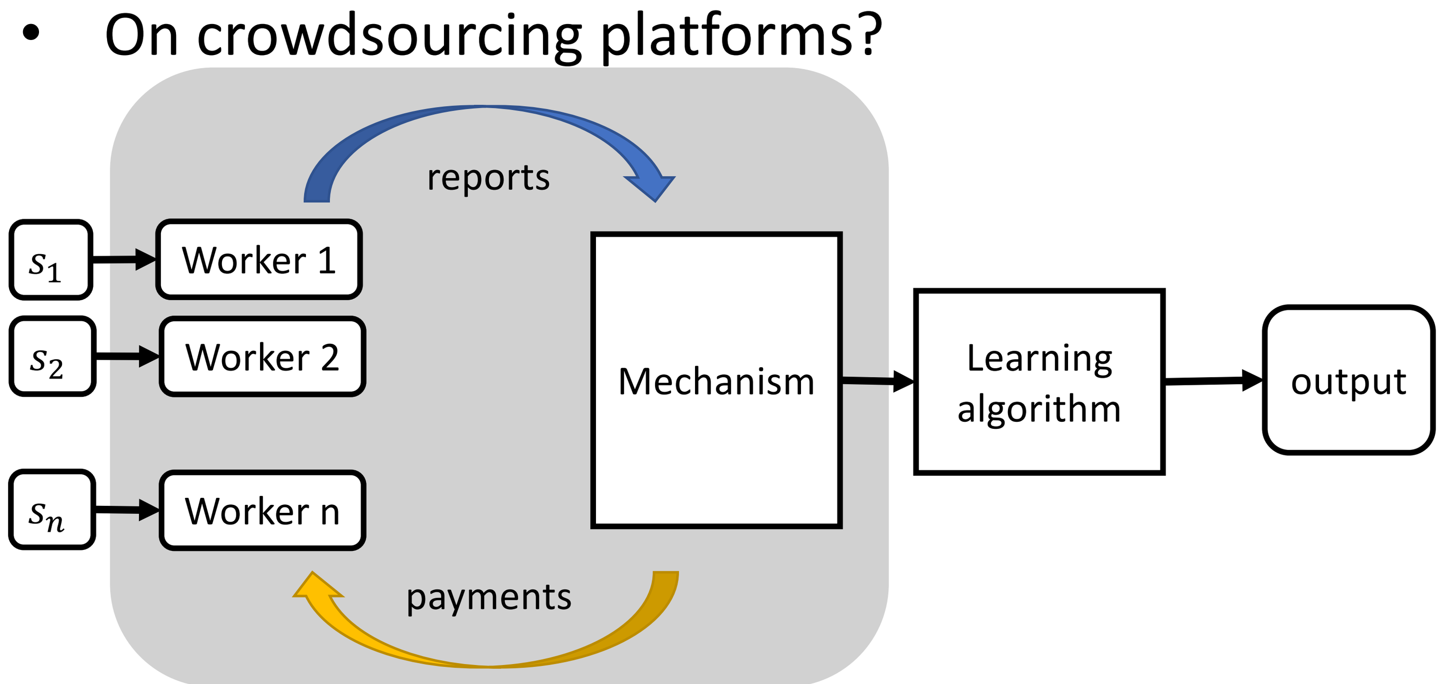
There is a device that can estimate gravitational acceleration with a small random errors.

- How can we estimate the gravitational acceleration $\mu$ at NYC if we have time and the device at hand?




Statistical inference

- On crowdsourcing platforms?




Crowdsourcing + Statistical inference

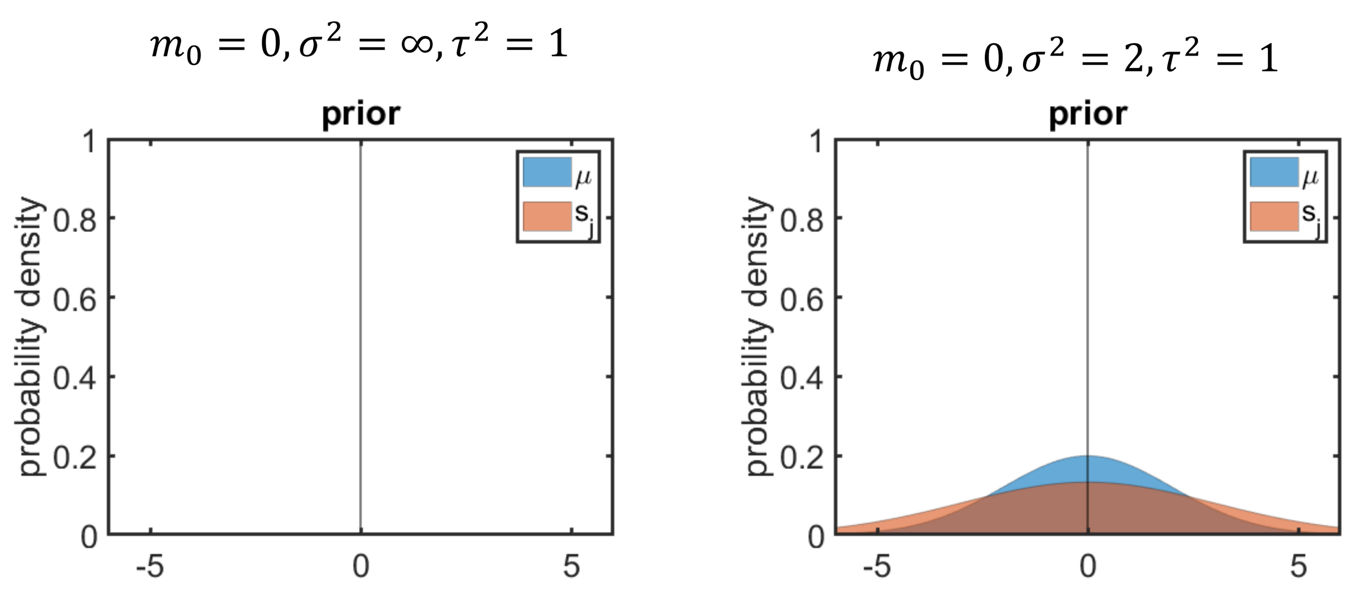
## Signal Structure

Two-step Gaussian distribution $(n, m_0, \sigma^2, \tau^2)$

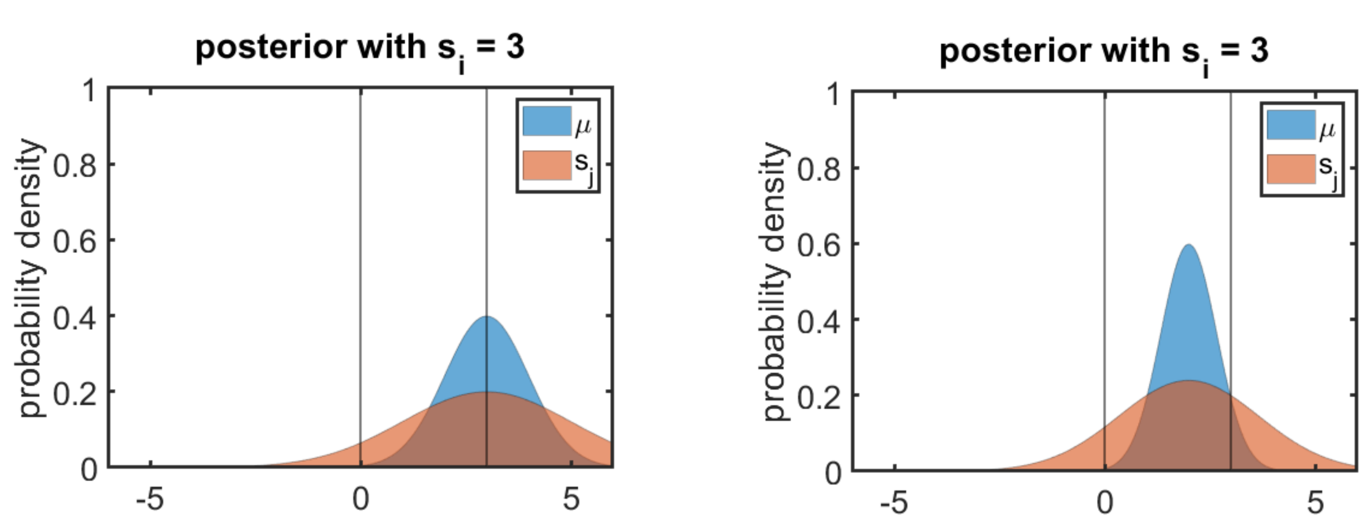
- Prior mean $m_0 \in \mathbb{R}^d$ where $d = 1$.
- Covariance matrices $\sigma^2, \tau^2 \in \mathbb{R}^{d \times d}$

Ground truth $\mu \sim \mathcal{N}(m_0, \sigma^2)$

Agent $i$'s private signal $s_i \sim \mathcal{N}(\mu, \tau^2)$ i.i.d. with $i \in [n]$




$$\mathbb{E}[s_j \mid s_i] = \frac{\tau^2 m_0 + \sigma^2 s_i}{\tau^2 + \sigma^2}$$




## Metric mechanism on Jeffery prior

1. Agents report $(\hat{s}_i)$ after observing signals $(s_i)$
2. For agent $i$
   1. Target: a random agent $j$
   2. Competitor: a random agent $k$
3. Pay agent $i$

$$\mathbb{1}[\|\hat{s}_j - \hat{s}_k\| > \|\hat{s}_j - \hat{s}_i\|].$$

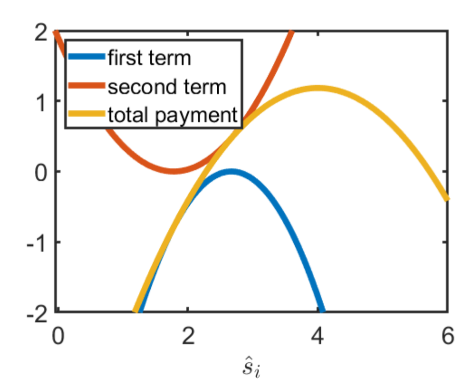
**Theorem (metric mechanism)** If $\sigma = \infty$ and $n \geq 4$, the metric mechanism is informed-truthful

- Truth-telling strategy profile ($\hat{s}_i = s_i$)
  - a Bayesian Nash equilibrium and
  - the highest social welfare
- Oblivious strategy profile
  - a strictly smaller social welfare

## General Two-step Gaussian

1. Each agent reports a signal and a prediction of the posterior mean$(\hat{s}_i, \hat{t}_i)$.
2. For agent $i$,
   1. Target: a random agent $j$
   2. Competitor: all other agents
3. Compute $L, M$ and average signal $\hat{t}_{-i}$
4. Prediction score
$$-L(\hat{s}_j - \hat{t}_i)^2 + M(\hat{s}_j - \hat{t}_{-i})^2$$
5. Information score
$$-M(\hat{s}_i - \hat{t}_{-i})^2 + L(\hat{s}_i - \hat{t}_j)^2$$

**Theorem (proxy BTS)** If $n \to \infty$, the proxy BTS mechanism is informed-truthful




1. Each agent reports $(\hat{s}_i, \hat{t}_i)$
2. For agent $i \in G_0$, pick a target $j$ randomly
3. Compute $T$
4. Prediction score
$$-(\hat{s}_j - \hat{t}_i)^2$$
5. Information score
$$-\|(T\hat{s}_i - \hat{t}_i) - (T\hat{s}_j - \hat{t}_j)\|$$

**Theorem (disagreement mechanism)** If $n \geq 3d + 3$, the disagreement mechanism is informed-truthful

## Discussion and Conclusion

- Streamline agents' report requirement: signal, posterior mean, or posterior belief
- Go beyond finite and simple signal structure: exponential family, graphical model, …
- Elicit information through geometry of parameter space