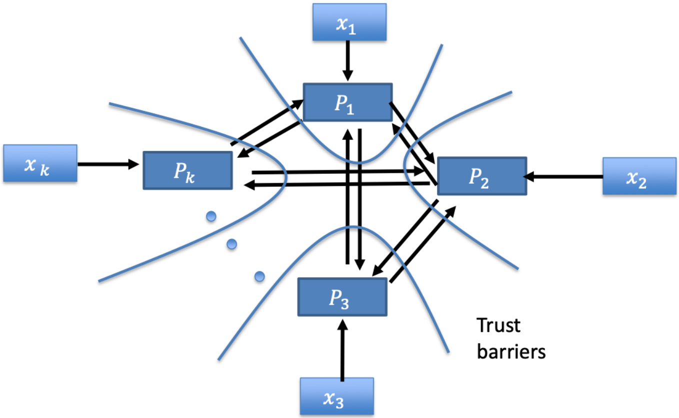
CS208: Applied Privacy for Data Science
End-to-end privacy

School of Engineering & Applied Sciences
Harvard University

April 12, 2022

Secure Multiparty Computation



Requirement: At end of protocol, each party P_i learns $f_i(x_1, \dots, x_n)$ and nothing else!

DP vs. Crypto

Model	Utility	Privacy	Who Holds Data?
Centralized Differential Privacy	statistical analysis of dataset	individual-specific info	trusted curator
Local or Federated Differential Privacy	statistical analysis of dataset	individual-specific info	original users (or delegates)
Secure Multiparty Computation	any query desired	everything other than result of query	original users (or delegates)
Fully Homomorphic (or Functional) Encryption	any query desired	everything (except possibly result of query)	untrusted server

Difference of Means

Outcome: $y_i \in [y_{\min}, y_{\max}]$; $\qquad R = y_{\max} - y_{\min}$

Treatment: $t_i \in \{0, 1\}$

Difference of Means

Outcome: $y_i \in [y_{\min}, y_{\max}]$; $\qquad R = y_{\max} - y_{\min}$

Treatment: $t_i \in \{0, 1\}$

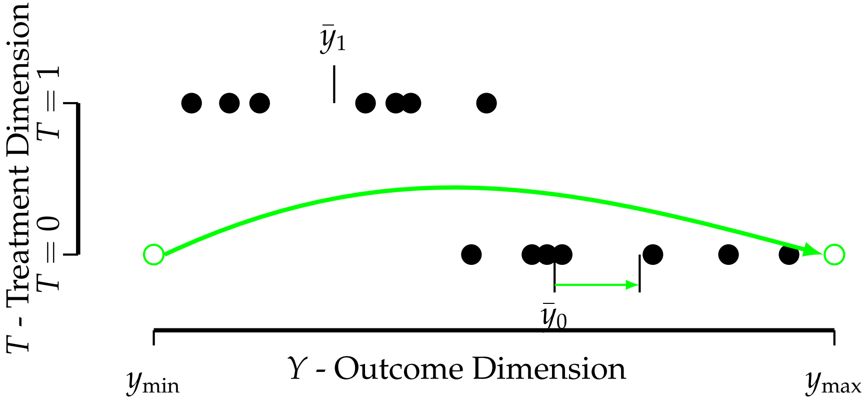
$$n_1 = \sum t_i \qquad n_0 = \sum 1 - t_i$$

$$\bar{y}_1 = \frac{\sum t_i y_i}{n_1} \qquad \bar{y}_0 = \frac{\sum (1 - t_i) y_i}{n_0}$$

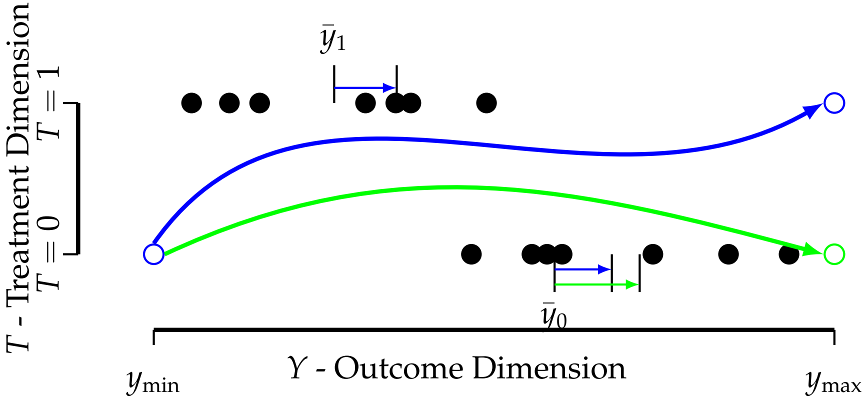
$$sd(y_1) = \sqrt{\frac{\sum t_i (y_i - \bar{y}_1)^2}{n_1}} \qquad sd(y_0) = \sqrt{\frac{\sum (1 - t_i)(y_i - \bar{y}_0)^2}{n_0}}$$

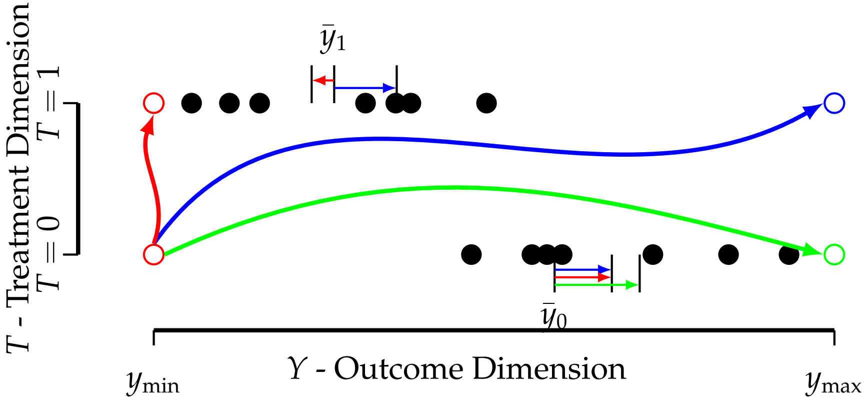
Statistic	Formula	Sensitivity
Difference of Means	$\bar{y}_1 - \bar{y}_0$	$\frac{R}{n_1+1} + \frac{R}{n_0+1}$
Standard Error	$\sqrt{\frac{sd(y_1)^2}{n_1} + \frac{sd(y_0)^2}{n_0}}$	$R\sqrt{\frac{N^*-1}{N^{*3}}}$

where $N^* = \min\,(n_0, n_1)$



$\bar{y}_1$
T - Treatment Dimension
$T = 1$
$T = 0$
$\bar{y}_0$
$y_{\min}$
Y - Outcome Dimension
$y_{\max}$


T - Treatment Dimension
$T = 1$
$T = 0$
$\bar{y}_1$
$\bar{y}_0$
$y_{\min}$
Y - Outcome Dimension
$y_{\max}$



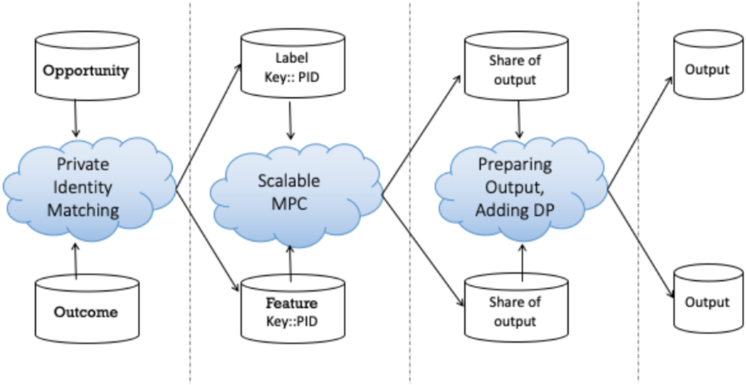
$\bar{y}_1$
$\bar{y}_0$
T - Treatment Dimension
$T = 1$
$T = 0$
Y - Outcome Dimension
$y_{\min}$
$y_{\max}$

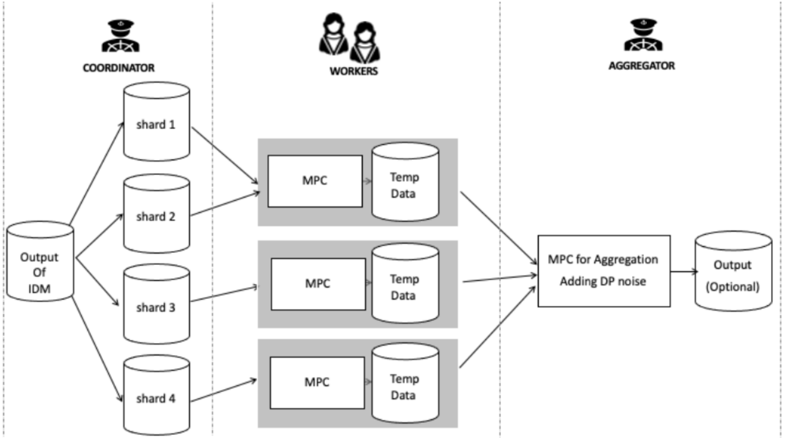
Alg.1 Differentially Private Diff.of Means Estimate

1. `Calculate` $\bar{y}_1 - \bar{y}_0$
2. `Calculate` $\Delta f = \frac{x_{\max} - x_{\min}}{N_1 + 1} + \frac{x_{\max} - x_{\min}}{N_0 + 1}$
3. `Draw` $Z \sim f_{Laplace}(\mu = 0, b = \Delta f / \epsilon)$
4. `Release` $M(X) = \bar{y}_1 - \bar{y}_0 + Z$

Privacy-Preserving Randomized Controlled Trials: A Protocol for Industry Scale Deployment

Mahnush Movahedi*
Benjamin M. Case
James Honaker
Andrew Knox
Li Li
Yiming Paul Li
Sanjay Saravanan
Shubho Sengupta
Erik Taubeneck
Facebook Inc
Menlo Park, CA


Opportunity
Outcome
Private Identity Matching
Label Key:: PID
Feature Key::PID
Scalable MPC
Share of output
Share of output
Preparing Output, Adding DP
Output
Output


COORDINATOR
WORKERS
AGGREGATOR
Output
Of
IDM
shard 1
shard 2
shard 3
shard 4
MPC
Temp
Data
MPC
Temp
Data
MPC
Temp
Data
MPC for Aggregation
Adding DP noise
Output
(Optional)

Algorithm 1 Differentially Private RCT

Input:

- x_T: user-level outcomes for the test group
- x_C: user-level outcomes for the control group
- R: upper bound of user-level outcomes (lower bound = 0)
- ρ_1: zCDP privacy budget for point estimate
- ρ_2: zCDP privacy budget for standard error
- α: significance level of confidence interval (e.g., 10%)

Output: [DP lift $- w$, DP lift $+ w$] confidence interval

1: Clamp/Winsorize:

$$Y_i = \begin{cases} X_i & \text{if } X_i \leq R \\ R & \text{if } X_i > R. \end{cases}$$

2: Calculate sample means, variances, and counts: $\bar{y}_T$, $\bar{y}_C$, s_T^2, s_C^2, n_T, n_C.

3: lift $\leftarrow \bar{y}_T - \bar{y}_C$.

4: Standard error of lift: $se_{\text{lift}} \leftarrow \sqrt{s_T^2/n_T + s_C^2/n_C}$.

5: Sensitivity of lift: $\Delta_{\text{lift}} \leftarrow \frac{R}{n_T} + \frac{R}{n_C}$.

6: Sensitivity of the standard error of lift: $\Delta_{se_{\text{lift}}} \leftarrow \sqrt{\frac{N^*-1}{N^{*3}}}R$, where $N^* = \min(n_T, n_C)$.

7: Draw scalar random noise $Z_1 \sim \text{Normal}\left(0, \frac{\Delta_{\text{lift}}^2}{2\rho_1}\right)$, $Z_2 \sim \text{Normal}\left(0, \frac{\Delta_{se_{\text{lift}}}^2}{2\rho_2}\right)$.

8: DP lift $\leftarrow$ lift $+ Z_1$, where $Z_1 \sim \text{Normal}\left(0, \frac{\Delta_{\text{lift}}^2}{2\rho_1}\right)$.

9: DP $se_{\text{lift}} \leftarrow se_{\text{lift}} + Z_2$, where $Z_2 \sim \text{Normal}\left(0, \frac{\Delta_{se_{\text{lift}}}^2}{2\rho_2}\right)$.

10: $w = \sqrt{(se_{\text{lift}} + Z_2)^2 + \frac{\Delta_{\text{lift}}^2}{2\rho_1}} \cdot z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the critical value of standard normal at $1 - \alpha/2$.