2014 IEEE 27th Computer Security Foundations Symposium

Differential Privacy: An Economic Method for Choosing Epsilon

Justin Hsu* Marco Gaboardi† Andreas Haeberlen* Sanjeev Khanna*
Arjun Narayan* Benjamin C. Pierce* Aaron Roth*

*University of Pennsylvania, USA †University of Dundee, Scotland

***Abstract*—*Differential privacy* is becoming a gold standard notion of privacy; it offers a guaranteed bound on loss of privacy due to release of query results, even under worst-case assumptions. The theory of differential privacy is an active research area, and there are now differentially private algorithms for a wide range of problems.**

However, the question of when differential privacy works *in practice* has received relatively little attention. In particular, there is still no rigorous method for choosing the key parameter ε, which controls the crucial tradeoff between the strength of the privacy guarantee and the accuracy of the published results.

In this paper, we examine the role of these parameters in concrete applications, identifying the key considerations that must be addressed when choosing specific values. This choice requires balancing the interests of two parties with conflicting objectives: the data *analyst*, who wishes to learn something abou the data, and the prospective *participant*, who must decide whether to allow their data to be included in the analysis. We propose a simple model that expresses this balance as formulas over a handful of parameters, and we use our model to choose ε on a series of simple statistical studies. We also explore a surprising insight: in some circumstances, a differentially private study can be *more accurate* than a non-private study for the same cost, under our model. Finally, we discuss the simplifying assumptions in our model and outline a research agenda for possible refinements.

***Index Terms*—Privacy**

I. Introduction

Protecting privacy is hard: experience has repeatedly shown that when owners of sensitive datasets release derived data, they often reveal more information than intended. Even careful efforts to protect privacy often prove inadequate—a notable example is the Netflix prize competition, which released movie ratings from subscribers. Although the data was carefully anonymized, Narayanan and Shmatikov were later able to "de-anonymize" some of the private records [21].

Privacy breaches often occur when the owner of the dataset uses an incorrect threat model—e.g., they make wrong assumptions about the knowledge available to attackers. In the case of Netflix, Narayanan and Shmatikov had access to auxiliary data in the form of a public, unanonymized data set (from IMDB) that contained similar ratings. Such errors are difficult to prevent without reasoning about arbitrary information that could be (or become) available to an attacker.

One way through this dilemma is to make sure that every computation on sensitive data satisfies *differential privacy* [8], a very strong guarantee: if an individual's data is used in a differentially private computation, the probability of any given result changes by at most a factor of e^{ε} compared to if the individual's data is not used in the computation, where ε is a parameter controlling the tradeoff between privacy and accuracy. Differential privacy impresses by the long list of assumptions it does *not* require: it is not necessary to know what information an attacker has, whether attackers are colluding, or what the attackers are looking for in particular.

But there is one question that users of differential privacy cannot avoid: how to choose the privacy parameter ε. It is the central parameter controlling strength of the privacy guarantee, and hence the number of queries that can be answered privately as well as the achievable accuracy. But ε is also a rather abstract quantity, and it is not clear how to choose an appropriate value in a given situation. This is evident in the literature, where algorithms have been evaluated with ε ranging from as little as 0.01 to as much as 7, often with no explanation or justification. A similar concern applies to the parameter δ in (ε, δ)-differential privacy, a standard generalization of differential privacy [7].

In this paper, we take a step towards a more principled approach by examining the impact of ε and δ on the different actors in a differentially private study: the data analyst, and the prospective participants who contribute private data. We propose a simple model that can be used to calculate a range of acceptable values of ε and δ, based on a few parameters of the study. Our model assumes that the participants are rational and will choose to contribute their data if their expected benefits (i.e., monetary compensation) from the study outweigh the risks (i.e., the bad events that may befall them as a result of their private data being exposed).

To demonstrate our model, we use it to choose ε in a series of case studies. We start by presenting the different parameters of our model, in the simplest situation where the analyst is interested in the result of only one query. Then, we consider a more realistic setting where the analyst wants to answer thousands of queries. Next, we show how our model can incorporate constraints specific to a particular study. Finally, we apply our model to a more accurate study under (ε, δ)-differential privacy. Throughout these examples, we vary the input parameters to our model through four scenarios—a clinical study of smokers, a study of educational data, a study of movie ratings, and a social network study—and show how the conclusions of our model change.

We also find that—perhaps counterintuitively—a study with strong differential privacy guarantees can sometimes be *cheaper* or (given a fixed budget) *more accurate* than an equivalent study without any privacy protections: while a differentially private study requires considerably more participants to account for the additional noise, it substantially

DOI 10.1109/CSF.2014.35

reduces the risks of each participant and thus lowers the compensation that rational participants should demand.

Our model provides a principled way to choose reasonable values for ε and δ based on parameters with more immediate connections to the real world. For many applications of differential privacy, this level of guidance may already prove useful. However, as is often the case, our model relies on some simplifying assumptions; for instance, we assume that participants fear some specific bad events when participating in the study, and that they can estimate their expected cost from these events even when they do not participate in the study. Some applications may require a more detailed model, and we consider possible refinements.

Our main contributions are: (1) a principled approach to choosing the privacy parameter ε for differentially private data analysis (Section IV); and (2) three case studies: a simple one-query study, a more sophisticated study answering many queries (Section V), and a study with external constraints (Section VII-C); and (3) an extension of our model to (ε,δ)-differential privacy (Section VIII). As an application of our model, we consider when a differentially private study can be cheaper than a non-private study (Section VI). We discuss possible extensions of our model in Section IX, and review related work in Section X.

II. Background: Differential privacy

Before describing our model, let us briefly review the core definitions of ε-differential privacy. (We defer the generalization of (ε,δ)-differential privacy to Section VIII.)

Differential privacy [8] is a quantitative notion of privacy that bounds how much a single individual's private data can contribute to a public output. The standard setting involves a *database* of private information and a *mechanism* that calculates an output given the database. More formally, a database D is a multiset of records belonging to some *data universe* $\mathcal{X}$, where a record corresponds to one individual's private data. We say that two databases are *neighbors* if they are the same size and identical except for a single record.[1] A mechanism $\mathcal{M}$ is a randomized function that takes the database as input and outputs an element of the range $\mathcal{R}$.

Definition 1 ([8]). *Given $\varepsilon \geq 0$, a mechanism $\mathcal{M}$ is ε-*differentially private *if, for any two neighboring databases D and D' and for any subset $S \subseteq \mathcal{R}$ of outputs,*

$$\Pr[\mathcal{M}(D) \in S] \;\leq\; e^{\varepsilon} \cdot \Pr[\mathcal{M}(D') \in S]. \qquad (1)$$

Note that S in this definition is any subset of the mechanism's range. In particular, when S is a singleton set $\{s\}$, the definition states that the probability of outputting s on a database D is at most e^{ε} times the probability of outputting s on any neighboring database D'.

For an intuitive reading of Definition 1, let x be an individual in database D, and let D' contain the same data as D except with x's data replaced by default data. Then, the differential privacy guarantee states that the probability of any output of mechanism $\mathcal{M}$ is within an e^{ε} multiplicative factor whether or not x's private data is included in the input. Hence, the parameter ε controls how much the distribution of outputs can depend on data from the individual x.

[1] The original definition of differential privacy [8] is slightly different: it says that neighboring databases are identical, except one has an *additional* record. We use our modified definition since we will assume the database size is public, in which case neighboring databases will have the *same* size.

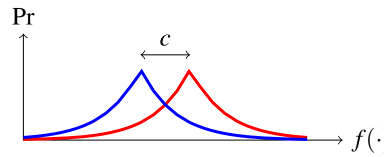

Fig. 1: Probability distributions of the Laplace mechanism for a c-sensitive function on two neighboring databases.

The definition also implies a lower bound: swapping D and D' yields $e^{\varepsilon} \cdot \Pr[\mathcal{M}(D) \in S] \geq \Pr[\mathcal{M}(D') \in S]$, or

$$\Pr[\mathcal{M}(D) \in S] \;\geq\; e^{-\varepsilon} \cdot \Pr[\mathcal{M}(D') \in S]. \qquad (2)$$

That is, the probability of an output in S on a database D is *at least* $e^{-\varepsilon}$ times the probability of an output in S on a neighboring database D'.[2]

A. The Laplace mechanism

The canonical example of a differentially private mechanism is the *Laplace mechanism*.

Theorem 2 ([8]). *Suppose $\varepsilon, c > 0$. A function g that maps databases to real numbers is* c-sensitive *if $|g(D) - g(D')| \leq c$ for all neighboring D, D'. For such a function, the* Laplace mechanism *is defined by $L_{c,\varepsilon}(D) = g(D) + \nu$, where ν is drawn from the Laplace distribution* $\mathrm{Lap}(c/\varepsilon)$*, that is, with probability density function*

$$F(\nu) = \frac{\varepsilon}{2c} \exp\left(\frac{-\varepsilon|\nu|}{c}\right).$$

This mechanism is ε-differentially private.

The scale c/ε of the Laplace distribution controls its *spread*: the distribution is wider for more sensitive functions (larger c) or stronger privacy guarantees (smaller ε), giving a higher probability of adding more noise.

For example, suppose that we have a database D of medical information and we wish to compute the proportion of smokers in a differentially private way. If the database has N records, define $g(D) = \#(\text{smokers in } D)/N$. Notice that, on any two neighboring databases, this proportion changes by at most $1/N$, since the numerator changes by at most 1 if a single record is altered. Thus, $L(D) = g(D) + \nu$, where $\nu \sim \mathrm{Lap}(1/N\varepsilon)$ is an ε-differentially private mechanism.

Differential privacy has many key benefits, like worst-case adversary assumptions and closure under composition and post-processing. We will take these features for granted;

[2] For example, if $\Pr[\mathcal{M}(D) \in S] = 0$ for some D and S, then $\Pr[\mathcal{M}(D') \in S] = 0$ for *all* databases D'—if some outputs are impossible on one input database, they must be impossible on all inputs.

they have been well-explored in the literature. (We refer the interested reader to the survey by Dwork [6].) Instead, let us take a closer look at the central parameter in the definition: ε.

III. Interpreting ε

A natural interpretation of differential privacy is in terms of *bad events*. For concreteness, let the mechanism be a scientific study, and suppose the individual has a choice to contribute data. Let $\mathscr{E}$ be a set of (real-world) events such that, if the output of the mechanism is fixed, an individual's participation has no effect on the probabilities of events in $\mathscr{E}$ (we will make this more precise below). Note that probabilities of events in $\mathscr{E}$ may still depend on the output of the mechanism. Roughly, $\mathscr{E}$ can be thought of as the set of privacy violation events.

To connect the outputs of the mechanism to the real-world events in $\mathscr{E}$, we imagine two runs of the mechanism: one with x's real data and one with dummy data, holding the other records fixed in both runs. Let x_p be the event "x participates," x_{np} be the event "x does not participate," and R be the output of the mechanism (a random variable). For any event $e \in \mathscr{E}$, the probability of event e if x participates in the study is

$$\Pr[e \mid x_p] = \sum_{r\in\mathscr{R}} \Pr[e \mid x_p, R = r] \cdot \Pr[R = r \mid x_p].$$

We say that events cannot *observe* x's participation if all the differences between the two trials are due to differences in the output: if the output is the same in both trials, the probability of events in $\mathscr{E}$ should be the same. That is, the first probability under the summation is the same assuming event x_p or x_{np}.

By differential privacy (Equation (1)), the second probability is bounded by e^ε times the probability of output r if x does not participate:

$$\begin{aligned} & \textstyle\sum_{r\in\mathscr{R}} \Pr[e \mid x_p, R = r] \cdot \Pr[R = r \mid x_p] \\ = \; & \textstyle\sum_{r\in\mathscr{R}} \Pr[e \mid x_{np}, R = r] \cdot \Pr[R = r \mid x_p] \\ \leq \; & e^\varepsilon \cdot \textstyle\sum_{r\in\mathscr{R}} \Pr[e \mid x_{np}, R = r] \cdot \Pr[R = r \mid x_{np}] = e^\varepsilon \cdot \Pr[e \mid x_{np}], \end{aligned}$$

where the first equality holds since events in $\mathscr{E}$ cannot observe x's participation. In particular, if e is a bad event in $\mathscr{E}$, the probability of e increases by at most a factor e^ε when the individual participates compared to when the individual does not participate. Hence, the interpretation of privacy in terms of bad events: under differential privacy, events in $\mathscr{E}$ will not be much more likely if an individual participates.

Since differential privacy bounds the *multiplicative* change in probabilities, the probability of a likely event may change significantly in *absolute* terms. Thus, the differential privacy guarantee is stronger for events that are very unlikely to happen if the individual does not participate. This is arguably true of most unpleasant events concerning private data: for instance, there is low probability that an individual's entire genome is released if they do not participate in a genetic study.

A. Introducing cost

Of course, not all bad events are equally harmful. To model this fact, we can assign a cost to each event. Specifically, suppose the potential participant has a non-negative *event cost function* $f_\mathscr{E}$ on the space of events $\mathscr{E}$. Let R again be the output of mechanism M, and define the associated *output cost function* f on the space of outputs $\mathscr{R}$ by

$$f(r) = \mathbb{E}_{e\in\mathscr{E}}[f_\mathscr{E}(e) \mid R = r].$$

Note that

$$\mathbb{E}_{e\in\mathscr{E}}[f_\mathscr{E}(e) \mid x_p] = \mathbb{E}_{r\in\mathscr{R}}[f(r) \mid x_p],$$

and similarly with x_{np} instead of x_p, so bounds on the expected value of f carry over to bounds on the expected value of $f_\mathscr{E}$. Thus, the individual need not reason about the set of outputs $\mathscr{R}$ and the output cost function f directly; they can reason just about costs of real-world events, represented by $f_\mathscr{E}$.

Using the differential privacy guarantee, we can bound the expected cost of participating in the study:[3]

$$e^{-\varepsilon} \mathop{\mathbb{E}}_{r\in\mathscr{R}}[f(r) \mid x_{np}] \;\leq\; \mathop{\mathbb{E}}_{r\in\mathscr{R}}[f(r) \mid x_p] \;\leq\; e^\varepsilon \mathop{\mathbb{E}}_{r\in\mathscr{R}}[f(r) \mid x_{np}] \quad (3)$$

In other words, the expected cost of x participating in a study is within an e^ε factor of the expected cost of declining.

Note that $\mathscr{E}$ and the cost function f have a large impact on the expected cost: for instance, if $\mathscr{E}$ contains bad events that will not actually be affected by the output of the mechanism, such as the event that an asteroid impact destroys civilization, the participant's perceived increase in expected cost may be prohibitively (and unjustifiably) large.

In general, the question of what events a differentially private study may be responsible for (i.e., what events should be in $\mathscr{E}$) is not a purely technical question, and could conceivably be handled by the legal system—just as laws describe who is liable for bad events, perhaps laws should also describe which events a private mechanism is liable for. Accordingly, our model does not specify precisely which events to put in $\mathscr{E}$, as long as they do not depend directly on the individual's participation. For our examples, we will consider events that can directly result from running a private mechanism.

B. The challenge of setting ε

So far, we have considered what ε means for the participant: higher values of ε lead to increases in expected cost. As we will soon see, there is another important consideration: ε controls how much noise is needed to protect privacy, so it has a direct impact on accuracy.

This is the central tension—abstractly, ε is a knob that trades off between privacy and utility. However, most prior work (we discuss some exceptions in Section X) focuses on how the knob works rather than how it should be set. High-level discussions about setting ε tend to offer fairly generic guidance, for example reasoning that a $e^{0.1} \sim 1.1$ factor increase in the probability of a bad event that is already very

[3] For one direction,

$$\begin{aligned} \mathop{\mathbb{E}}_{r\in\mathscr{R}}[f(r) \mid x_p] &= \sum_{r\in\mathscr{R}} \Pr[R = r \mid x_p] \cdot f(r) \\ &\leq \sum_{r\in\mathscr{R}} e^\varepsilon \Pr[R = r \mid x_{np}] \cdot f(r) = e^\varepsilon \cdot \mathop{\mathbb{E}}_{r\in\mathscr{R}}[f(r) \mid x_{np}]. \end{aligned}$$

Note that the inequality requires $f(r) \geq 0$. The other direction is similar, appealing to Equation (2).

improbable is a minor concern, so 0.1 is a sensible value for ε. On the other hand, experimental evaluations of differential privacy, where a concrete choice of ε is required, often just pick a value (ranging from 0.01 [25] to 7 [19]) with little justification.

In a sense, the difficulty of choosing ε is a hidden consequence of a key strength of differential privacy: its extreme simplicity. That is, ε is difficult to think about precisely because it rolls up into a single parameter a fairly complex scenario involving at least two parties with opposing interests (the analyst and the participants), as well as considerations like compensating individuals for their risk.

Our goal in this paper is to unpack this complexity and offer a more ramified model with more intuitive parameters.

IV. A TWO-PARTY MODEL

We propose a simple model for choosing ε, involving two rational parties: a data analyst and an individual considering whether to participate in the analyst's study.

A. The analyst's view

The analyst's goal is to conduct a study by running a private mechanism, in order to learn (and publish) some useful facts. The analyst's main concern is the *accuracy* $A_{\mathscr{M}}$ of the mechanism's result, with respect to some benchmark.

One natural benchmark is the "true answer" for the non-differentially-private version of the study, which we call the *sample statistic*. Compared to this standard, the error in a private study is due entirely to noise added to preserve privacy. This error is determined partly by ε, but also can depend on N, the number of records in the analyst's database: if a larger number of records leads to less privacy loss to any individual, less noise is needed to protect privacy.[4]

Another possible benchmark is the true answer on the entire population, which we call the *population statistic*. This is the natural benchmark when we want to infer properties of the population, given only a random sample of individual data (here, the database). For this benchmark, an additional source of error is *sampling error*: the degree to which the sample is not perfectly representative of the population. This error tends to decrease as N increases: larger samples (databases) are more representative. This error is not due to differential privacy and so is independent of ε.

Since these errors typically decrease as N increases, the analyst would conduct huge studies were it not for a second constraint: *budget*. Each individual in the study needs to be compensated for their participation, so the analyst can't afford studies of infinite size. The analyst model describes these two conflicting constraints.

Definition 3. *The* analyst *runs a private mechanism M parameterized by ε and N. The mechanism comes with a real-valued* accuracy function *$A_{\mathscr{M}}(\varepsilon,N)$, where smaller values of $A_{\mathscr{M}}(\varepsilon,N)$ correspond to more accurate results. (We will omit the subscript when the mechanism is clear.) The analyst wants a target accuracy α, and so requires that $A_{\mathscr{M}}(\varepsilon,N) \leq \alpha$. Finally, the analyst has a* budget *B to spend on the study.*

Depending on what the analyst is trying to learn, he may be able to tolerate a lower or higher total error. In general, the analyst may have a utility function that quantifies how bad a specific amount of error is. Though our model can be extended to handle this situation, for simplicity we assume that the analyst cannot tolerate inaccuracy beyond the target level and is otherwise equally happy.

B. The individual's view

We next consider the individuals who might want to contribute their information to a database in exchange for payment. Study participants may want compensation for various reasons; for example, they may want a flat compensation just for their time. Even though our model can be easily extended to handle per-participant costs, for simplicity we do not consider this cost. Instead, we focus on the compensation most relevant to privacy: participants may face personal harm if their data is revealed, so they are willing to join the study only if they are adequately compensated for the risk they take. A simple way to model the individual's risk is via a cost function f, as described in Section II.

We suppose the individual is offered a choice between participating in a study and declining, but the study will always take place. Our model does not say whether to run the study or not—are the study's potential discoveries worth the potential (non-privacy-related) harm to individuals?[5] Instead, we assume that some authority has decided that the study will take place, and the individual only gets to decide whether to participate or not. Thus, the individual participates only if they are compensated for their *marginal* increase in expected cost.

From the interpretation of differential privacy in terms of bad events (Section III), an individual's expected cost should increase by at most an e^{ε} factor if she decides to participate in a study. There is one detail we need to attend to: our previous calculation of the marginal increase in cost depends on the probability of each possible output of the mechanism. This probability should be interpreted as taken over not only the randomness in the mechanism, but also over the uncertainty of an individual about the rest of the database.

To make this point clearer, we separate these two sources of randomness in the calculation of the marginal increase in cost for a specific individual x. Let $\mathscr{D}$ be the set of all possible databases of size N, and let E be x's expected cost if she decides *not* to participate. Unpacking,

$$\begin{aligned} E = \mathbb{E}[f(\mathscr{M}(D))] &= \sum_{s\in\mathscr{R},D^*\in\mathscr{D}} \Pr[D=D^*, s=\mathscr{M}(D)]\cdot f(s) \\ &= \sum_{D^*\in\mathscr{D}} \Pr[D=D^*]\cdot \sum_{s\in\mathscr{R}} \Pr[s=\mathscr{M}(D) \mid D=D^*]\cdot f(s), \end{aligned}$$

[4]For example, if the Laplace mechanism is used to release an average value, the sensitivity of the underlying function depends on N: as N increases, the sensitivity decreases, so less noise is required to achieve a given level of privacy.

[5]Indeed, the difference in harm between running a study and not running the study may be very large: for instance, running a study may discover that smoking causes lung cancer, increasing insurance costs for all smokers.

where $\Pr[D = D^*]$ encodes an individual's belief about the contents of the entire database, and by extension an individual's belief about the output of the mechanism run on the entire database. E represents an upper bound on the individuals' beliefs about how much the study will cost them if they do *not* participate in the study. For example, a study might discover that people in a certain town are likely to have cancer—this knowledge could harm all the residents of the town, not just the participants. Similarly, if C is the individual's expected cost if they *do* participate and y is any record in D (representing a default or dummy record),

$$\begin{aligned} C &= \mathbb{E}[f(\mathcal{M}(D \cup x \setminus y))] \\ &= \sum_{s \in \mathcal{R}, D^* \in \mathcal{D}} \Pr[D = D^*, s = \mathcal{M}(D \cup x \setminus y)] \cdot f(s) \\ &= \sum_{D^* \in \mathcal{D}} \Pr[D = D^*] \cdot \sum_{s \in \mathcal{R}} \Pr[s = \mathcal{M}(D \cup x \setminus y) \mid D = D^*] \cdot f(s). \end{aligned}$$

But the inner summation is the individual's expected cost when the rest of the database is known to be D^*. By Equation (3), we bound the increase of cost C if x participates (for any y):

$$\begin{aligned} & \textstyle\sum_{s \in \mathcal{R}} \Pr[s = \mathcal{M}(D \cup x \setminus y) \mid D = D^*] \cdot f(s) \\ \leq \quad & e^{\varepsilon} \textstyle\sum_{s \in \mathcal{R}} \Pr[s = \mathcal{M}(D) \mid D = D^*] \cdot f(s). \end{aligned}$$

Repacking the expressions for E and C, we get $C \leq e^{\varepsilon} E$. Hence the individual's *marginal* cost of participation $C - E$ satisfies $C - E \leq e^{\varepsilon} E - E = (e^{\varepsilon} - 1)E$.

Now, we are ready to define a model for the individual.

Definition 4. *The* individuals *are offered a chance to participate in a study with a set level of ε for some payment. Each individual considers a space of real-world events that, conditioned on the output of the study and the database size, are independent of their participation.*

Each individual also has a non-negative cost function on this space, which gives rise to a non-negative cost function f on the space of outputs of the mechanism, and base cost $\mathbb{E}[f(R)]$*, where R is the random output of the mechanism without the individual's data. Let E be an upper bound on the individual's base costs. The individual participates only if they are compensated for the worst-case increase in their expected cost by participating: $(e^{\varepsilon} - 1)E$.*

Note the requirement on the space of bad events: we condition on the output of the mechanism, as well as the size of the database. Intuitively, this is because the size of the database is usually published. While such information may sometimes be private, it is hard to imagine conducting a study without anyone knowing how many people are in it—for one thing, the size controls the budget for a study. By this conditioning, we require that an adversary cannot infer an individual's participation even if he knows both the database size and the output of the mechanism.

C. Combining the two views

To integrate the two views, we assume that the analyst directly compensates the participants. Supposing that the analyst has total budget B, since N individuals need to be paid $(e^{\varepsilon} - 1)E$ each, we have the following budget constraint:[6]

$$(e^{\varepsilon} - 1)EN \leq B \tag{4}$$

This constraint, combined with the analyst's accuracy constraint $A_M(\varepsilon, N) \leq \alpha$, determines the feasible values of N and ε. In general, there may be no feasible values: in this case, the mechanism cannot meet the requirements of the study. On the other hand, there may be multiple feasible values. These trade off between the analyst's priorities and the individual's priorities: larger values of ε and smaller values of N make the study smaller and more accurate, while smaller values of ε and larger values of N give a stronger guarantee to the individuals. In any case, feasible values of N and ε will give a study that is under budget, achieves the target accuracy, and compensates each individual adequately for their risk.

Note that the payments depend on the particular study only through the E parameter—different studies require different data, which may lead to different base costs—and the ε parameter, which controls the privacy guarantee; other internal details about the study do not play a role in this model. By using differential privacy as an abstraction, the model automatically covers differentially private mechanisms in many settings: offline, interactive, distributed, centralized, and more. Further, the model can be applied whether the analyst has benevolent intentions (like conducting a study) or malicious ones (like violating someone's privacy). Since differential privacy does not make this distinction, neither does our model.

D. Deriving the cost E

While the expected cost of not participating in a study may seem like a simple idea, there is more to it than meets the eye. For instance, the cost may depend on what the individuals believe about the outcome of the study, as well as what bad events individuals are worried about. The cost could even depend on prior private studies an individual has participated in—the more studies, the higher the base cost.

Since individuals have potentially different beliefs about this cost, the analyst must be sure to offer enough payment to cover each individual's expected cost. Otherwise, there could be *sampling bias*: individuals with high cost could decline to participate in the study. While the analyst would like to offer each individual just enough compensation to incentivize them to participate, this amount may depend on private data. Thus, we model the analyst as paying each individual the same amount, based on some maximum expected cost E.

Even if this maximum expected cost is difficult to perfectly calculate in practice, it can be estimated in various ways: reasoning about specific bad events and their costs, conducting surveys, etc. While there has been work on using auctions to discover costs related to privacy [11], [18], [2], [5], [24],

[6]Since we do not consider compensating participants for their time (though our model can be extended to cover this case), the "budget" should be thought of as the cost needed to cover privacy-related harm, part of a potentially larger budget needed to conduct the study.

estimating this cost in a principled way is an area of current research. Therefore, we will not pick a single value of E for our examples; rather, we show how different values of E affect our conclusions by considering estimates for a few scenarios.

Remark 5. *Our goal in the following sections is to demonstrate how our model works in a simple setting. Hence, we will consider studies with rather simplistic statistical analyses. As a result, the number of participants (and costs) required to achieve the accuracy may seem unreasonably high. There is a vast literature on sophisticated study design; more advanced methods (such as those underlying real medical studies) can achieve better accuracy at more modest cost.*

V. A SIMPLE STUDY

In this section, we will show how to apply our model to a simple study that answers some queries about the database.

A. A basic example: estimating the mean

Suppose we are the analyst, and we want to run a study estimating the proportion of individuals in the general population with some property P; we say this target proportion μ is the *population mean*. We also have a measure of accuracy $A(\varepsilon,N)$ (which we define below), a target accuracy level α and a fixed budget provided by the funding agency.

First, we specify our study. For any given N and ε, we will recruit N subjects to form a private database D_N. We model the participants as being chosen independently and uniformly at random, and we consider the database D_N as a random variable. (We sometimes call the database the *sample*.) We then calculate the proportion of participants with property P (the *sample mean*)—call it $g(D_N)$. Since g is a $1/N$-sensitive function, we release it using the ε-private Laplace mechanism by adding noise $\nu(\varepsilon,N)$ drawn from $\mathrm{Lap}(1/N\varepsilon)$ to $g(D_N)$.

Now, we need to specify the accuracy function $A(\varepsilon,N)$ of this study. In general, there are several choices of what A can measure. In this example we will fix the desired error T, and we say the mechanism *fails* if it exceeds the error guarantee T. More precisely, the failure probability p is

$$p := \Pr[\,|g(D_N)+\nu(\varepsilon,N)-\mu| \geq T].$$

As analysts, our goal is to ensure that p is at most some given level α (i.e., α measures how confident we are).

While it would be natural to define $A(\varepsilon,N)$ as precisely this failure probability (and requiring $A(\varepsilon,N) \leq \alpha$), the analysis is simpler if we define $A(\varepsilon,N)$ to be an *upper bound* on the failure probability. Then, $A(\varepsilon,N) \leq \alpha$ is a sufficient condition on the true failure probability p being bounded by α.

There are two sources of error for our mechanism: from the sample mean deviating from the population mean ($|g(D_N)-\mu|$), and from the Laplace noise added to protect privacy ($|\nu(\varepsilon,N)|$). Let us bound the first source of error.

Theorem 6 (Chernoff Bound)**.** *Suppose $\{X_i\}$ is a set of N independent, identically distributed $0/1$ random variables with mean μ and sample mean $Y = \frac{1}{N}\sum X_i$. For $T \in [0,1]$,*

$$\Pr[\,|Y-\mu| \geq T] \;\leq\; 2\exp\left(\frac{-NT^2}{3\mu}\right).$$

Assuming that our sample D_N is drawn independently from the population, we can model each individual as a random variable X_i which is 1 with probability μ, and 0 otherwise. Then, the Chernoff bound is a bound on the probability of the sample mean $g(D_N)$ deviating from the true proportion μ that we are interested in. Note that the sample must be free of sampling bias for this to hold—inferring population statistics from a non-representative sample will skew the estimates. This is why we must compensate participants so that they are incentivized to participate, regardless of their private data.

Similarly, we use the following result to bound the second source of error, from adding Laplace noise.

Lemma 7 (Tail bound on Laplace distribution)**.** *Let ν be drawn from* $\mathrm{Lap}(\rho)$*. Then,*

$$\Pr[\,|\nu| \geq T] \;\leq\; \exp\left(-\frac{T}{\rho}\right).$$

Now, since the total error of the mechanism is the difference between the sample mean and the population mean plus the Laplace noise, if the output of the Laplace mechanism deviates from the population mean by at least T, then either the sample mean deviates by at least $T/2$, or the Laplace noise added is of magnitude at least $T/2$. Therefore, we can bound the failure probability p by

$$\Pr[\,|g(D_N)-\mu| \geq T/2] + \Pr[\,|\nu| \geq T/2].$$

Consider the first term. Since $\mu \leq 1$, the Chernoff bound gives

$$\Pr[\,|g(D_N)-\mu| \geq T/2] \;\leq\; 2e^{-NT^2/12\mu} \;\leq\; 2e^{-NT^2/12}.$$

The tail bound on the Laplace distribution gives

$$\Pr[\,|\nu(\varepsilon,N)| \geq T/2] \;\leq\; \exp\left(-\frac{TN\varepsilon}{2}\right),$$

since we added noise with scale $\rho = 1/\varepsilon N$. Therefore, given a target accuracy α, an upper bound on the failure probability p is

$$p \leq A(\varepsilon,N) := 2\exp\left(-\frac{NT^2}{12}\right) + \exp\left(-\frac{TN\varepsilon}{2}\right), \qquad (5)$$

the accuracy constraint $A(\varepsilon,N) \leq \alpha$ suffices to guarantee that $p \leq \alpha$ as desired.

For the budget side of the problem, we need to compensate each individual by $(e^{\varepsilon}-1)E$, according to our individual model. If our budget is B, the budget constraint is Equation (4): $(e^{\varepsilon}-1)EN \;\leq\; B$.

Our goal is to find ε and N that satisfy this budget constraint, as well as the accuracy constraint Equation (5). While it is possible to use a numerical solver to find a solution, here we derive a closed-form solution. Eliminating ε and N from these constraints is difficult, so we find a sufficient condition on feasibility instead. First, for large enough ε, the sampling error dominates the error introduced by the Laplace noise. That is, for $\varepsilon \geq T/6$,

$$\exp\left(-\frac{TN\varepsilon}{2}\right) \leq \exp\left(-\frac{NT^2}{12}\right),$$

so it suffices to satisfy this system instead:

$$3\exp\left(\frac{-NT^2}{12}\right) \leq \alpha \quad \text{and} \quad (e^{\varepsilon}-1)EN \leq B. \tag{6}$$

Figure 2 gives a pictorial representation of the constraints in Equation (6). For a fixed accuracy α, the blue curve (marked α) contains values of ε, N that achieve error α. The blue shaded region (above the α curve) shows points that are feasible for that accuracy—there, ε, N give accuracy better than α. The red curve (marked B) and red shaded region (below the B curve) show the same thing for a fixed budget B. The intersection of the two regions (the purple area) contains values of ε, N that satisfy both the accuracy constraint, and the budget constraint. Figure 3 shows the equality curves for Equation (6) at different fixed values of α and B.

Solving the constraints for N, we need

$$N \geq \frac{12}{T^2}\ln\frac{3}{\alpha}. \tag{7}$$

Taking equality gives the loosest condition on ε, when the second constraint becomes

$$\varepsilon \leq \ln\left(1+\frac{BT^2}{12E\ln\frac{3}{\alpha}}\right).$$

Thus, combining with the lower bound on ε, if we have

$$\frac{T}{6} \leq \varepsilon \leq \ln\left(1+\frac{BT^2}{12E\ln\frac{3}{\alpha}}\right), \tag{8}$$

then the study can be done at accuracy α, budget B. Since we have assumed $\varepsilon \geq T/6$, this condition is sufficient but not necessary for feasibility. That is, to deem a study *infeasible*, we need to check that the *original* accuracy constraint (Equation (5)) and budget constraint (Equation (4)) have no solution.

For a concrete instance, suppose we want to estimate the true proportion with ± 0.05 accuracy (5% additive error), so we take $T = 0.05$. We want this accuracy except with at most $\alpha = 0.05$ probability, so that we are 95% confident of our result. If Equation (8) holds, then we can set $\varepsilon = T/6 = 0.0083$ and N at equality in Equation (7), for $N \approx 20000$.

For the budget constraint, let the budget be $B = 3.0 \times 10^4$. If we now solve Equation (8) for the base cost E, we find that the study is feasible for $E \leq E_{feas} \approx 182$. This condition also implies a bound on an individual's cost increase, which we have seen is $(e^{\varepsilon}-1)E$. To decide whether a study is feasible, let us now now estimate the cost E and compare it to E_{feas} in various scenarios.

B. Analyzing the costs scenarios

We consider participant costs for four *cost scenarios*; these will be our running examples for the remainder of the paper.

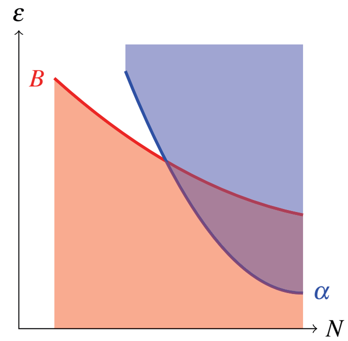

Fig. 2: Feasible ε, N, for accuracy α and budget B.

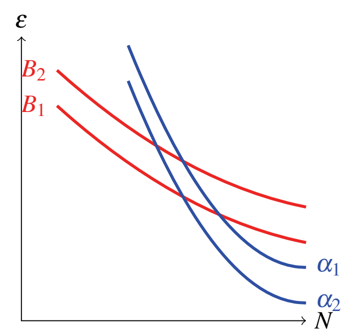

Fig. 3: Constant accuracy curves for $\alpha_1 < \alpha_2$, constant budget curves for $B_1 < B_2$

Smoking habits. The data comes from smokers who fear their health insurance company will find out that they smoke and raise their premiums. The average health insurance premium difference per year between smokers and nonsmokers is \$1,274 [20]. Thus, some participants fear a price increase of \$1,274.[7] Since we pay the same amount to everyone, smoker or not, we base our estimate of the base cost on the smokers' concerns. To estimate the base cost E, we first estimate the probability that the insurance company concludes that an individual smokes, even if they do not participate in the study. This is possible: perhaps people see her smoking outside.

So, suppose the participants think there is a moderate, 20% chance that the insurance company concludes that they are smokers, even if they do not participate. Thus, we can estimate the base cost by $E = 0.20 \cdot 1274 = 254.8$—this is the cost the participants expect, even if they do not participate. Since this is far more than $E_{feas} \approx 182$, the study may not be feasible. To check, we plug the exact accuracy constraint (Equation (5)) and budget constraint (Equation (4)) into a numeric solver, which reports that the system has no solution for ε, N. Thus, we conclude that the study not feasible.

[7]Note that this covers the cost for just one year. Throughout our scenarios, we focus on immediate, concrete costs, rather than projecting out costs far into the future.

Perhaps this is not so surprising: since smoking is often done in public, it may not be considered truly private information—indeed, the probability of the bad event if the individual did not participate was already fairly large (we estimated 20%). For more typical privacy scenarios, this may not be the case.

Educational data. The data consists of students' educational records, including courses taken and grades. Suppose individuals fear their record of classes and grades are published, perhaps causing them to be fired and forced to switch to a job with lower pay. The mean starting salary of college graduates in the United States is $45,000 [1]; let us suppose that they face a pay cut of 30% ($12,500) if their records become public.

If the individual does not participate in the study, it is still possible that an employer uncovers this data: perhaps someone steals the records and publishes them, or maybe the outcome of the study (without the individual) can be combined with public information to infer an individual's grades. However, complete disclosure seems like a low probability event; let us suppose the chance is one in a thousand (0.001). Then, we can estimate the base cost by $E = 0.01 \cdot 12500 = 12.5$. Since this is less than E_{feas}, the study is feasible.

Movie ratings data. The data consists of movie ratings; suppose individuals fear their private movie ratings are published, much like the case of the Netflix challenge [21]. Again, while an exact monetary amount of harm is not obvious, we can consider the monetary cost of this disclosure, legally speaking. The Video Privacy Protection Act of 1998 is an American law stating that punitive damages for releasing video rental records should be at least $2,500.

If the individual does not participate in the study, it is possible that their movie ratings are still published—perhaps their internet service provider is monitoring their activities, or someone guesses their movies. Again, disclosure seems like a low probability event; let us suppose the chance is one in ten thousand (0.00001). Then, we can estimate the base cost by $E = 0.0001 \cdot 2500 = 0.25$. Since this is less than E_{feas}, the study is feasible.

Social networks. Many social networks allow anonymous personas. Suppose individuals fear their true identity will be revealed from studies on the social network structure, much like deanonymization attacks against anonymous Twitter accounts [22]. The exact monetary amount harm of disclosure is not clear, but the dangers can be serious: consider dissidents operating anonymous accounts. Suppose we estimate the cost of disclosure to be very high, say $100,000.

If the individual's information is not included in the social network data, it is unlikely that their identity is revealed—perhaps there is a physical privacy breach. If the individual takes strong precautions to guard their privacy, then disclosure seems like a very low probability event if they do not participate in this study. Let us suppose the chance is one in a hundred thousand (0.00001). Then, we can estimate the base cost by $E = 0.00001 \cdot 100000 = 1$. This is less than E_{feas}, so the study is feasible.

C. *A more realistic example: answering many queries*

When working out the simple study above, we needed to derive the accuracy totally from scratch. Moreover, the mechanism is not very powerful—it can answer only a single query! In this section, we address these issues by considering a more sophisticated algorithm from the privacy literature: the multiplicative weights exponential mechanism (MWEM) [12], [13]. As part of our analysis, we will show how to directly plug established accuracy results for MWEM into our model.

MWEM is a mechanism that can answer a large number of *counting queries*: queries of the form "What fraction of the records in the database satisfy property P?" For example, suppose that the space of records is bit strings of length d, i.e. $\mathscr{X} = \{0,1\}^d$. Each individual's bit string can be thought of as a list of attributes: the first bit might encode the gender, the second bit might encode the smoking status, the third bit might encode whether the age is above 50 or not, etc. Then, queries like "What fraction of subjects are male, smokers and above 50?", or "What proportion of subjects are female nonsmokers?" are counting queries.

To use our model, we will need an accuracy bound for MWEM [13]: For a data universe $\mathscr{X}$, set of queries $\mathscr{C}$ and N records, the ε-private MWEM answers all queries in $\mathscr{C}$ within additive error T with probability at least $1-\beta$, where

$$T = \left(\frac{128 \ln|\mathscr{X}| \ln\left(\frac{32|\mathscr{C}|\ln|\mathscr{X}|}{\beta T^2}\right)}{\varepsilon N} \right)^{1/3}.$$

We again define the accuracy function $A(\varepsilon, N)$ to be the probability β of exceeding error T on any query. Solving,

$$A(\varepsilon, N) := \beta = \frac{32|\mathscr{C}|\ln|\mathscr{X}|}{T^2} \exp\left(-\frac{\varepsilon N T^3}{128 \ln|\mathscr{X}|}\right),$$

Like the previous example, we want to satisfy the constraints $A(\varepsilon, N) \leq \alpha$ and $(e^{\varepsilon} - 1)EN \leq B$.

For a concrete example, suppose we want $\mathscr{X} = \{0,1\}^8$ so $|\mathscr{X}| = 2^8$ and accuracy $T = 0.2$ for 20% error, with bad accuracy at most 5% of the time, so $\alpha = 0.05$. Further, we want to answer 10000 queries, so $|\mathscr{C}| = 10000$.

We can now carry out cost estimates, supposing that the budget is $B = 2.0 \times 10^6$. Looking at the movie ratings scenario when $E = 0.25$, the constraints are satisfiable: take $N = 8.7 \times 10^5$, $\varepsilon = 2.3$, compensating each individual $(e^{\varepsilon} - 1)E = 2.2$.

For the social network scenario where $E = 1$, the constraints are also satisfiable: take $N = 1.3 \times 10^6$, $\varepsilon = 1.5$, compensating each individual $(e^{\varepsilon} - 1)E = 3.5$. The calculations are similar for the other cost scenarios; since those examples have higher base cost E, the required budget will be higher.

VI. The True Cost of Privacy

Now that we have a method for estimating the cost of a private study, we can compare this cost to that of an equivalent non-private study. Again, we consider only costs arising from compensation for harm to privacy.

Differential privacy requires additional noise to protect privacy, and requires a larger sample to achieve the same accuracy. Hence, one would expect private studies to be more expensive than equivalent non-private studies. While this is true if individuals are paid the same in both cases, differential privacy also has an advantage: it can bound the harm to individuals, whereas—as deanonymization attacks have shown—it is difficult for non-private studies to make any guarantees about privacy.

When an individual participates in a non-private study, it is very possible that their information can be completely recovered from published results. Thus, to calculate the cost for the non-private study, we consider a hypothetical world where non-private study participants are compensated in proportion to their worst-case cost W, i.e., their cost for having their data published in the clear.

It would be unreasonable for the non-private study to pay each individual their full worst-case cost: even in the most poorly designed non-private study, it is hard to imagine every individual having their data published in the clear. More likely, attackers may be able to recover some fraction of the data; for instance, if attackers are analyzing correlations with public datasets, the public datasets may contain information for only a portion of all the study participants. Thus, we suppose a non-private study might expose up to a ϕ fraction of the participants, and the non-private study compensates each participant with a ϕ fraction of their worst-case cost, i.e., ϕW.

Consider the mean-estimation study from Section V. For the non-private study with N' individuals, we directly release the sample mean $g(D) = \frac{1}{N'}\sum X_i$.[8] Thus, we do not have to bound the error from Laplace noise—all the error is due to the sample mean deviating from the population mean.

We can then give conditions under which the private study is cheaper than the public study. (The proof can be found in the extended version of this paper [14].)

Theorem 8. *Given a target error $T \geq 0$ and target accuracy $\alpha > 0$, the private mean estimation study will be cheaper than the non-private mean estimation study exposing a fraction ϕ of participants if the following (sufficient, but not necessary) condition holds:*

$$\frac{T}{6} \leq \ln\left(1 + \frac{\phi W \ln \frac{1}{2\alpha}}{96 E \ln \frac{3}{\alpha}}\right) \tag{9}$$

The non-private study needs at least N' individuals, where

$$N' \geq \frac{1}{8T^2} \ln \frac{1}{2\alpha}. \tag{10}$$

[8]Remark 5 holds here as well: more sophisticated statistical experiments can achieve accuracy for less resources, while we use a simple analysis. However, we do so for *both* the private and non-private studies, so comparing the relative costs is fair.

For our example calculations below, we take $\phi = 1/500 = 0.002$.[9] Now that we have a sufficient condition for when the private study is cheaper than the public study, let us turn back to our four cost scenarios. From Section V, recall we wanted to estimate the mean of a population to accuracy $T = 0.05$, with failure probability at most $\alpha = 0.05$.

Smoking habits. Recall that we estimated the base cost E to be 254.8. The worst case cost is at least the rise in health insurance premium, so we let $W = 1274$. Plugging in these numbers, Equation (9) does not hold. So, the private study is not necessarily cheaper.

Educational data. Recall that we estimated the base cost E to be 12.5. The worst case cost is at least the loss in salary: \$12,500; we take this to be W. Plugging these numbers into Equation (9), we find that the private study is cheaper.

Movie ratings. Recall that we estimated the base cost E to be 0.25, and we estimated the worst case disclosure cost to be at least the damages awarded under the Video Privacy Protection Act. So we let $W = 2500$. Plugging these numbers into Equation (9), we find that the private study is cheaper.

Social networks. Recall that we estimated the base cost E to be 1. The worst case cost is at least the cost of discovery: \$100,000; we let this be W. Plugging these numbers into Equation (9), we find that the private study is cheaper.

Let us compare the size and costs of the two studies, just for the movie ratings scenario. By Equation (10), the non-private study needs $N' \geq 115$. As expected, the non-private study needs fewer people to achieve the same accuracy compared to the private study ($N = 20000$), since no noise is added. However, the total cost for non-private study would be $B' = \phi W N = 0.002 \cdot 2500 \cdot 115 \approx 575$. The equivalent private study, with $E = 0.25, \varepsilon = 0.0083, N = 20000$ costs $(e^{\varepsilon} - 1) \cdot EN \approx 40$.

If both private and non-private studies have the same budget, the private study can buy more participants to further improve its accuracy. Thus, this private study is *more accurate* and *cheaper* (and more private!) than the non-private version.

VII. Extending the Model

So far, we have considered just two constraints on ε and N: expected cost to the individuals (expressed as a budget constraint), and accuracy for the analyst. Other constraints may be needed to model finer details—we will refer to these additional constraints as *side conditions*. In this section, we first consider generic upper and lower bounds on ε—these follow from the definition of differential privacy. Then, we present a case study incorporating side conditions.

A. Upper bounds on ε

While the definition of differential privacy is formally valid for any value of ε [9], values that are too large or too small give

[9]Precise measurements of the success rate of real deanonymization attacks are hard to come by, for at least two reasons: first, published deanonymization attacks aim to prove a concept, rather than violate as many people's privacy as possible. Second, published deanonymization attacks generally do not have the luxury of knowing the original data, so they are necessarily conservative in reporting success rates. Adversarial attacks on privacy need not satisfy these constraints.

weak guarantees. For large values of ε, the upper bound on the probability $\Pr[\mathscr{M}(D) \in S]$ can rise above one and thus become meaningless: for instance, if $\varepsilon = 20$, Equation (1) imposes no constraint on the mechanism's output distribution unless $\Pr[\mathscr{M}(D') \in S] \le e^{-20}$.

To demonstrate this problem, we describe an ε-private mechanism for large ε which is not intuitively private. Consider a mechanism M with range $\mathscr{R}$ equal to data universe $\mathscr{X}$, and consider a targeted individual J. When J is in the database, the mechanism publishes their private record with probability $p^* > 1/|\mathscr{X}|$, otherwise it releases a record at random.

We first show that this mechanism is ε-differentially private, for a very large ε. Let j be J's record, and let

$$p = \frac{1 - p^*}{|\mathscr{X}| - 1} < \frac{1}{|\mathscr{X}|}$$

be the probability of releasing a record $s \neq j$ when J is in the database. Consider two databases $D \cup i$ and $D \cup j$, where i is any record. For M to be ε-differentially private, it suffices that

$$\begin{aligned} & e^{-\varepsilon} \Pr[\mathscr{M}(D \cup i) = j] \\ & \quad \le \Pr[\mathscr{M}(D \cup j) = j] \le e^{\varepsilon} \Pr[\mathscr{M}(D \cup i) = j] \\ & e^{-\varepsilon} \Pr[\mathscr{M}(D \cup i) = s] \\ & \quad \le \Pr[\mathscr{M}(D \cup j) = s] \le e^{\varepsilon} \Pr[\mathscr{M}(D \cup i) = s], \end{aligned}$$

for all $s \neq j$. Rewriting, this means

$$e^{-\varepsilon} \frac{1}{|\mathscr{X}|} \le p^* \le e^{\varepsilon} \frac{1}{|\mathscr{X}|} \quad \text{and} \quad e^{-\varepsilon} \frac{1}{|\mathscr{X}|} \le p \le e^{\varepsilon} \frac{1}{|\mathscr{X}|}.$$

By assumption, the left inequality in the first constraint and the right inequality in the second constraint hold. Thus, if

$$\varepsilon \ge \ln(p^* |\mathscr{X}|), \tag{11}$$

the first constraint is satisfied. Since the probabilities over all outputs sums to one, we also know $p^* + (|\mathscr{X}| - 1)p = 1$. So,

$$\varepsilon \ge \ln\left(\frac{1}{p|\mathscr{X}|}\right) \ge \ln\left(\frac{|\mathscr{X}| - 1}{|\mathscr{X}|(1 - p^*)}\right) \tag{12}$$

suffices to satisfy the second constraint.

Therefore, M is ε-differentially private if ε satisfies these equations. For instance, suppose $|\mathscr{X}| = 10^6$, and $p^* = 0.99$. M almost always publishes J's record (probability 0.99) if J is in the database, but it is still ε-differentially private if $\varepsilon \ge 14$.

Clearly, a process that always publishes a targeted individual's data if they are in the database and never publishes their data if they are not in the database is blatantly non-private. This ε-private mechanism does nearly the same thing: with probability $p^* = 0.99$, it publishes J's record with probability at least $p^* = 0.99$ if J is in the database, and with probability $1/|\mathscr{X}| = 10^{-6}$ if J is not. Evidently, values of ε large enough to satisfy both Equations (11) and (12) do not give a very useful privacy guarantee.

B. Lower bounds on ε

While choosing ε too large will compromise the privacy guarantee, choosing ε too small will ruin *accuracy*—the mechanism must behave too similarly for databases that are very different. For example, let D, D' be arbitrary databases of size N, and let $0 < \varepsilon \le \frac{1}{N}$. Since the two databases have the same size, we can change D to D' by changing at most N rows. Call the sequence of intermediate neighboring databases $D_1, \cdots, D_{N-1}$. By differential privacy,

$$\begin{aligned} \Pr[\mathscr{M}(D) \in S] &\le e^{\varepsilon} \Pr[\mathscr{M}(D_1) \in S] \\ \Pr[\mathscr{M}(D_1) \in S] &\le e^{\varepsilon} \Pr[\mathscr{M}(D_2) \in S] \\ \cdots \quad \Pr[\mathscr{M}(D_{N-1}) \in S] &\le e^{\varepsilon} \Pr[\mathscr{M}(D') \in S]. \end{aligned}$$

Combining, $\Pr[\mathscr{M}(D) \in S] \le e^{N\varepsilon} \Pr[\mathscr{M}(D') \in S]$. Similarly, we can use Equation (2) to show $\Pr[\mathscr{M}(D) \in S] \ge e^{-N\varepsilon} \Pr[\mathscr{M}(D') \in S]$. But we have taken $\varepsilon \le 1/N$, so the exponents are at most 1 and at least -1. So, the probability of every event is fixed up to a multiplicative factor of at most e, whether the input is D or D'. (Differential privacy with $\varepsilon = 1$ guarantees this for neighboring databases, but here D and D' may differ in many—or all!—rows.) Such an algorithm is probably useless: its output distribution depends only weakly on its input.

C. Case Study: Educational statistics

Putting everything together, we now work through an example with these added constraints on ε, together with a limit on the study size. We consider the same mean estimation study from Section V, except now with side constraints.

Concretely, suppose that we are in the educational data scenario, where each student's record contains class year (4 possible values), grade point average (rounded to the nearest letter grade, so 5 possible values), years declared in the major (4 possible values), and order of courses in the major (100 possible combinations). The total space of possible values is $|\mathscr{X}| = 4 \cdot 5 \cdot 4 \cdot 100 = 8000$.

We now add in our side conditions. First, suppose that we have data for $N = 1000$ students, spanning several years. It may not be simple to expand the study size—perhaps this data for all the students, or perhaps we only have access to recent data. the only way to collect more data is to graduate more students. We also want the upper and lower bounds on ε discussed above to hold.

For the accuracy, recall from Section V that if T is the desired additive error and α is the probability we do not achieve this accuracy, the accuracy constraint is

$$2 \exp\left(-\frac{NT^2}{12}\right) + \exp\left(-\frac{TN\varepsilon}{2}\right) \le \alpha.$$

In this example it is not very natural to think of a total budget for compensation, since all the data is assumed to have been already collected. Instead, we know the privacy harm for any individual is at most $(e^{\varepsilon} - 1) \cdot E$, and we will bound the maximum allowed privacy harm per student. Suppose it is $B_0 = 10$, giving the constraint $(e^{\varepsilon} - 1) \cdot E \le B_0$.

To capture the side conditions, we add a few more constraints. For the population, we require $N \le 1000$. For the upper bound on ε, we do not want Equations (11) and (12) to both hold, so we add a constraint

$$\varepsilon \leq \max \left[\ln(0.1 \cdot |\mathscr{X}|), \ln\left(\frac{|\mathscr{X}|-1}{|\mathscr{X}|(1-0.1)}\right)\right].$$

For the lower bound on ε, we add the constraint $\varepsilon \geq 1/N$.

Putting it all together, with base cost $E = 12.5$, record space size $|\mathscr{X}| = 8000$ and allowed harm per student $B_0 = 10$, and target error $T = \alpha = 0.05$, the feasibility of this study is equivalent to the following system of constraints.

$$2\exp(-0.0002 \cdot N) + \exp(-0.025 N\varepsilon) \leq 0.05,$$

$$(e^{\varepsilon} - 1) \cdot 12.5 \leq 10, \qquad N \leq 1000,$$

$$1/N \leq \varepsilon \leq \max(\ln(800), \ln(1.11))$$

Note that we are requiring the same accuracy as in our original study in Section V, and in fact the original study without the side constraints is feasible. However, a numeric solver shows that these constraints are not feasible, so this study is not feasible in this setting.[10]

VIII. What about δ?

In this section, we extend our model to an important generalization of ε-differential privacy, known as (ε, δ)-differential privacy.

Definition 9 ([7])**.** *Given $\varepsilon, \delta \geq 0$, a mechanism M is* (ε, δ)-differentially private *if for any two neighboring database D, D', and for any subset $S \subseteq \mathscr{R}$ of outputs,*

$$\Pr[\mathscr{M}(D) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathscr{M}(D') \in S] + \delta.$$

Intuitively, this definition allows a δ probability of failure where the mechanism may violate privacy. For instance, suppose that s is an output that reveals user x's data. For a database D' that does not contain user x's information, suppose $\Pr[\mathscr{M}(D') \in S] = 0$. Under ε-differential privacy, M can never output s on any database. However, under (ε, δ)-differential privacy, M may output s with probability up to δ, when fed any neighboring database D. In particular if $D = D' \cup x \setminus y$, $\Pr[\mathscr{M}(D) \in S]$ may be up to δ: even though M never outputs x's records on databases without x, M can output x's record when she is in the database with probability δ.

A. Modeling δ

By considering "blatantly non-private" mechanisms that nevertheless satisfy (ε, δ)-privacy, we can upper bound δ. For example, for a database with N records and for $\delta = 1/N$, the mechanism that randomly outputs a record from the database is $(0, \delta)$-private. This mechanism is intuitively non-private, so we require $\delta \ll 1/N$ for a more reasonable guarantee.

For a more principled method of picking this parameter, we can model the costs associated with different levels of δ. The first step is to bound the increase in expected cost for participating in an (ε, δ)-private mechanism. We assume a bound W on an individual's cost if their data is publicly revealed, since with probability δ the mechanism may do just that. Then, we can bound an individual's increase in expected cost when participating in an (ε, δ)-private study. (The proof can be found in the extended version of this paper [14].)

Proposition 10. *Let $\mathscr{M}$ be an (ε, δ)-private mechanism with range $\mathscr{R}$, and let f be a non-negative cost function over $\mathscr{R}$. Let $W = \max_{s \in \mathscr{R}} f(s)$. Then, for neighboring databases D, D',*

$$\mathbb{E}[f(\mathscr{M}(D))] \leq e^{\varepsilon}\mathbb{E}[f(\mathscr{M}(D'))] + \delta W.$$

We can now incorporate the δ parameter into our model.

Definition 11 ((ε, δ)-private analyst model)**.** *An* (ε, δ)-private analyst *is an analyst with accuracy $A_{\mathscr{M}}$ a function of ε, N, δ.*

Definition 12 ((ε, δ)-private individual model)**.** *An* (ε, δ)-private individual *is an individual with a* worst-case cost W, *which measures the cost of publicly revealing the individual's private information. The individual wants to be compensated for her worst-case marginal cost of participating under these assumptions: $e^{\varepsilon}E + \delta W - E = (e^{\varepsilon} - 1)E + \delta W$.*

Since (ε, δ)-privacy is weaker than pure ε-privacy, why is it a useful notion of privacy? It turns out that in many cases, (ε, δ)-private algorithms are more accurate than their pure privacy counterparts; let us consider such an example.

B. Revisiting MWEM

In Section V, we analyzed the cost of MWEM. We will now revisit that example with an (ε, δ)-private version of MWEM. The setting remains the same: we wish to answer a large number of counting queries with good accuracy, while preserving privacy.

The main difference is the accuracy guarantee, due to Hardt and Rothblum [13]. Suppose the space of records is $\mathscr{X}$ and we want to answer queries $\mathscr{C}$ to accuracy T with probability at least $1 - \beta$. The (ε, δ)-private MWEM has accuracy

$$T = \frac{8(\ln|\mathscr{X}|\ln(1/\delta))^{1/4} \ln^{1/2}\left(\frac{32|\mathscr{C}|\ln|\mathscr{X}|}{\beta T^2}\right)}{N^{1/2}\varepsilon^{1/2}}.$$

We define our accuracy measure $A(\varepsilon, N)$ to be the failure probability β. Solving, this means

$$A(\varepsilon, N) := \beta = \frac{32|\mathscr{C}|\ln|\mathscr{X}|}{T^2} \exp\left(-\frac{\varepsilon N T^2}{8(\ln|\mathscr{X}|\ln(1/\delta))^{1/2}}\right).$$

If α is the target accuracy, we need $A(\varepsilon, N) \leq \alpha$.

For the budget constraint, we need $(e^{\varepsilon} - 1)EN + \delta W N \leq B$. Suppose we are in the social network scenario we described in Section V, with the same budget $B = 2.0 \times 10^6$ we used for the $(\varepsilon, 0)$-private MWEM algorithm. We use our running estimate of the base cost for this scenario, $E = 1$, and the worst-case cost, $W = 10^6$. For the other parameters, suppose the records are bit strings with 15 attributes (versus 8 before): $\mathscr{X} = \{0,1\}^{15}$ and $|\mathscr{X}| = 2^{15}$. We want to answer

[10] To be precise, we have shown that this *particular* mechanism (i.e., the Laplace mechanism) is not feasible—there may be other, more clever mechanisms that feasibly compute what we want. We are not aware of any such mechanisms, but we also cannot rule it out.

$|\mathscr{C}| = 200000$ queries (versus 10000 before), to 5% error (versus 20% before), so $T = 0.05$, with probability at least 95% ($\alpha = 0.05$, same as before).

Plugging in the numbers, we find that the accuracy and budget constraints can both be satisfied, for $\varepsilon = 0.9$, $\delta = 10^{-8}$, and $N = 9.1 \times 10^5$. Each individual is compensated $(e^{\varepsilon} - 1)E + \delta W = 1.46$, for a total cost of $1.9 \times 10^6 \leq B$. Thus, the (ε, δ)-private version of MWEM answers more queries, over a larger space of records, to better accuracy, than the $(\varepsilon, 0)$-version we previously considered.

IX. Discussion

A. Is all this complexity necessary?

Compared to earlier threat models from the differential privacy literature, our model may seem overly complex: the original definition from Dwork et al. [8] had only one parameter (ε), while our model involves a number of different parameters (α, $A_{\mathscr{M}}(\varepsilon, N)$, B, and E). So, at first glance, the original model seems preferable. However, we argue that this complexity is present in the real world: the individuals really *do* have to consider the possible consequences of participating in the study, the researchers really *do* require a certain accuracy, etc. The original definition blends these considerations into a single, abstract number ε. Our model is more detailed, makes the choices explicit, and forces the user to think quantitatively about how a private study could affect real events.

B. Possible refinements

The key challenge in designing any model is to balance complexity and accuracy. Our model is intended to produce reasonable suggestions for ε in most situations while keeping only the essential parameters. Below, we review some areas where our model could be refined or generalized.

Estimating the base cost. Our model does not describe how to estimate the base cost for individuals. There is no totally rigorous way to derive the base cost: this quantity depends on how individuals perceive their privacy loss, and how individuals think about uncertain events. These are both active areas of research—for instance, research in psychology has identified a number of cognitive biases when people reason about uncertain events [15]. Thus, if the consequences of participation are uncertain, the individuals might under- or over-estimate their expected cost.[11] More research is needed to incorporate these (and other) aspects of human behavior into our model.

Another refinement would be to model individuals heterogeneously, with different base costs and desired compensations. For instance, an individual who has participated in many studies may be at greater risk than an individual who has never participated in any studies. However, care must be taken to avoid sampling bias when varying the level of compensation.

Empirical attacks on privacy. Our model assumes that all ε-private studies could potentially increase the probability of bad events by a factor of e^{ε}. It is not clear whether (a) this is true for private algorithms considered in the literature, and (b) whether this can be effectively and practically exploited. The field of differential privacy (and our model) could benefit from empirical attacks on private algorithms, to shed light on how harm actually depends on ε, much like parameters in cryptography are chosen to defend against known attacks.

Collusion. Our model assumes the study will happen regardless of a single individual's choice. However, this may not be realistic if individuals collude: in an extreme case, all individuals could collectively opt out, perhaps making a study impossible to run. While widespread collusion could be problematic, assumptions about the size of limited coalitions could be incorporated into our model.

Large ε. As ε increases, our model predicts that the individual's marginal expected harm increases endlessly. This is unreasonable—there should be a maximum expected cost for participating in a study, perhaps the worst case cost W. The cost curve could be refined for very small and very large values of ε.

Modeling the cost of non-private studies. Our comparison of the cost of private and non-private studies uses a very crude (and not very realistic) model of the cost of non-private studies. More research into how much individuals want to be compensated for their private data would give a better estimate of the true tradeoff between private and non-private studies.

X. Related work

There is by now a vast literature on differential privacy, which we do not attempt to survey. We direct the interested reader to an excellent survey by Dwork [6].

The question of how to set ε has been present since the introduction of differential privacy. Indeed, in early work on differential privacy, Dwork [6] indicates that the value of ε, in economic terms or otherwise, is a "social question." Since then, few works have taken an in-depth look at this question. Works applying differential privacy have used a variety of choices for ε, mostly ranging from 0.01–10, with little or no convincing justification.

The most detailed discussion of ε we are aware of is due to Lee and Clifton [16]. They consider what ε means for a hypothetical adversary, who is trying to discover whether an individual has participated in a database or not. The core idea is to model the adversary as a Bayesian agent, maintaining a belief about whether the individual is in the database or not. After observing the output of a private mechanism, he updates his belief depending on whether the outcome was more or less likely if the individual had participated.

As Lee and Clifton show, ε controls how much an adversary's belief can change, so it is possible to derive a bound on ε in order for the adversary's belief to remain below a given threshold. We share the goal of Lee and Clifton of deriving a bound for ε, and we improve on their work. First, the "bad event" they consider is the adversary discovering an individual's participation in a study. However, by itself, this

[11] In some experiments, people give up their private data for as little as a dollar [3].

knowledge might be relatively harmless—indeed, a goal of differential privacy is to consider harm beyond reidentification.

Second, and more seriously, the adversary's Bayesian updates (as functions of the private output of the mechanism) are themselves differentially private: the distribution over his posterior beliefs conditioned on the output of the mechanism is nearly unchanged regardless of a particular agent's participation. In other words, an individual's participation (or not) will usually lead to the same update by the adversary. Therefore, it is not clear why an individual should be concerned about the adversary's potential belief updates when thinking about participating in the study.

Related to our paper, there are several papers investigating (and each proposing different models for) how rational agents should evaluate their costs for differential privacy [26], [11], [23], [4], [18]. We adopt the simplest and most conservative of these approaches, advocated by Nissim, et al. [23], and assume that agents costs are upper bounded by a linear function of ε.

Alternatively, privacy (quantified by ε) can be thought of as a fungible commodity, with the price discovered by a market. Li, et al. [17] consider how to set arbitrage-free prices for queries. Another line of papers [11], [10], [18], [2], [5], [24] consider how to discover the value of ε via an auction, when ε is set to be the largest value that the data analyst can afford.

XI. Conclusion

We have proposed a simple economic model that enables users of differential privacy to choose the key parameters ε and δ in a principled way, based on quantities that can be estimated in practice. To the best of our knowledge, this is the first comprehensive model of its kind. We have applied our model in two case studies, and we have used it to explore the surprising observation that a private study can be cheaper than a non-private study with the same accuracy. We have discussed ways in which our model could be refined, but even in its current form the model provides useful guidance for practical applications of differential privacy.

Acknowledgments

We thank Emilio Jesús Gallego Arias for fruitful discussions that improved this work; we also thank the anonymous reviewers for their close reading and suggestions. This work was supported by NSF grants CCF-1101389, CCF-1116961, CNS-1065060 and CNS-1054229, and a Google Focused Research Award. Marco Gaboardi has been supported by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement No. 272487.

References

[1] NACE Salary survey, Jan. 2013.
[2] C. Aperjis and B. Huberman. A market for unbiased private data: Paying individuals according to their privacy attitudes. 2012.
[3] A. R. Beresford, D. Kübler, and S. Preibusch. Unwillingness to pay for privacy: A field experiment. *Economics Letters*, 2012.
[4] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan. Truthful mechanisms for agents that value privacy. In *Proc. EC*, pages 215–232. ACM, 2013.
[5] P. Dandekar, N. Fawaz, and S. Ioannidis. Privacy auctions for recommender systems. In *Proc. WINE*, pages 309–322, 2012.
[6] C. Dwork. Differential privacy: A survey of results. In *Proc. TAMC*, Apr. 2008.
[7] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. *Proc. EUROCRYPT*, 2006.
[8] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proc. TCC*, Mar. 2006.
[9] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Differential privacy – a primer for the perplexed. In *Conf. of European Statisticians, Joint UNECE/Eurostat work session on statistical data confidentiality*, 2011.
[10] L. K. Fleischer and Y.-H. Lyu. Approximately optimal auctions for selling privacy when costs are correlated with data. In *Proc. EC*, New York, NY, USA, 2012. ACM.
[11] A. Ghosh and A. Roth. Selling privacy at auction. In *Proc. EC*, pages 199–208. ACM, 2011.
[12] M. Hardt, K. Ligett, and F. McSherry. A simple and practical algorithm for differentially private data release. In *Proc. NIPS*, pages 2348–2356, 2012.
[13] M. Hardt and G. N. Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *Proc. FOCS*, pages 61–70. IEEE, 2010.
[14] J. Hsu, M. Gaboardi, A. Haeberlen, S. Khanna, B. C. Pierce, and A. Roth. Differential privacy: An economic method for choosing epsilon. http://privacy.cis.upenn.edu/papers/epsilon.pdf.
[15] D. Kahneman, P. Slovic, and A. Tversky. *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press, 1982.
[16] J. Lee and C. Clifton. How much is enough? choosing ε for differential privacy. In *Proc. ISC*, ISC'11, 2011.
[17] C. Li, D. Y. Li, G. Miklau, and D. Suciu. A theory of pricing private data. In *Proc. ICDT*, pages 33–44, 2013.
[18] K. Ligett and A. Roth. Take it or leave it: Running a survey when privacy comes at a cost. In *Proc. WINE*, pages 378–391, 2012.
[19] A. Machanavajjhala, A. Korolova, and A. D. Sarma. Personalized social recommendations - accurate or private? *PVLDB*, 4(7):440–450, 2011.
[20] J. Moriarty, M. Branda, K. Olsen, N. Shah, B. Borah, A. Wagie, J. Egginton, and J. Naessens. The effects of incremental costs of smoking and obesity on health care costs among adults: A 7-year longitudinal study. *Journal of Occupational and Environmental Medicine*, 54(3):286, 2012.
[21] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *Proc. S&P*. IEEE, May 2008.
[22] A. Narayanan and V. Shmatikov. De-anonymizing social networks. In *IEEE Symposium on Security and Privacy (S&P), Oakland, California*, pages 173–187, 2009.
[23] K. Nissim, C. Orlandi, and R. Smorodinsky. Privacy-aware mechanism design. In *Proc. EC*. ACM, 2012.
[24] A. Roth and G. Schoenebeck. Conducting truthful surveys, cheaply. In *Proc. EC*. ACM, 2012.
[25] A. D. Sarwate, K. Chaudhuri, and C. Monteleoni. Differentially private support vector machines. *CoRR*, abs/0912.0071, 2009.
[26] D. Xiao. Is privacy compatible with truthfulness? In *Proc. ITCS*, pages 67–86, 2013.