


The background of the entire image is a dark grey brick wall. Mounted on this wall is a grid of 20 surveillance cameras, arranged in 5 rows and 4 columns. Most of the cameras are black, but the cameras in the second, third, and fourth columns of the first three rows are silver. A semi-transparent white rectangular box is centered on the wall, containing the title and introductory text. The title 'Understanding differential privacy' is in a large, black, serif font. Below it, a paragraph of text in a smaller, black, sans-serif font explains the concept of differential privacy. At the end of the paragraph, the name 'Christoph F. Kurz' is written in a bold, black, sans-serif font, followed by the text 'works through a simple hypothetical example' in a regular, black, sans-serif font.

# Understanding differential privacy

Differential privacy promises to strike a balance between the protection of privacy for individuals and the generation of insights from data. But how exactly does it work?  
**Christoph F. Kurz** works through a simple hypothetical example





**D**ata protection is a valuable asset, but so is data analysis. Without data analysis there would be no scientific knowledge and no medical, technical or economic advances. Although data can be used for surveillance and manipulation, it also creates transparency and strengthens democracy. In order to unite data protection and data evaluation, personal information should be removed or made completely unrecognisable at the earliest possible stage of the data processing chain.

Differential privacy (DP) is an approach that promises to combine both the protection of privacy and the desire for the greatest possible gain in insight when making use of data. In this way, DP is not a single tool, but rather a requirement that several methods for analysing sensitive personal information have been developed to fulfil. Today, many technology companies such as Facebook and Google use DP for privacy protection of their user data, and it is also being used for the 2020 US Census.

The DP concept was developed at Microsoft by Dwork *et al.*<sup>1</sup> and leverages random noise – random variation around the true value – to falsify data. This conserves statistical properties but makes individual information hard to identify. While adding random noise to data is not a new and exciting innovation, a unique feature of DP is that it clearly quantifies privacy loss and offers privacy guarantees that can be mathematically proven. A privacy parameter regulates exactly how much information can be leaked and how much random noise is introduced during differentially private computing. However, this requires that analysts using DP have to follow certain rules when reporting statistics; it requires them to calculate “differentially private” sums, means, standard deviations and so on.

Why is reporting simple sums and means of large data sets not enough to conserve privacy? Consider the following example. Assume there is a company whose stakeholders require a monthly report on the average salaries of their employees. Let this average salary be €50,000. Now assume that a new colleague, Alice, joins the company and the following month the average salary is reported to be €51,000 – even though no original worker leaves or changes salary. This

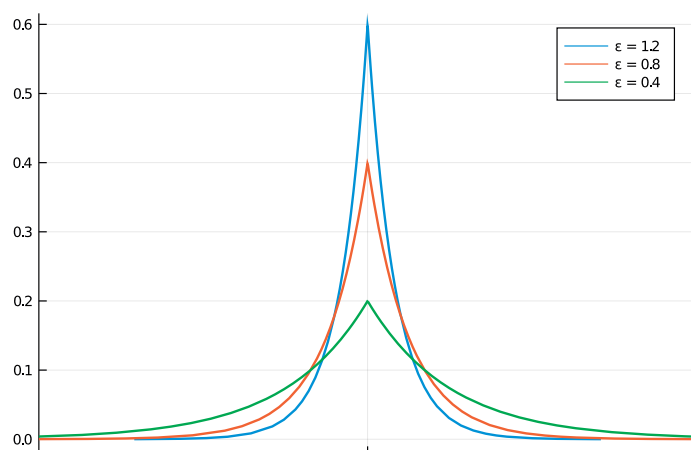
immediately reveals two things: first, Alice’s salary is more than €50,000; second, if it’s a small company with, say, 20 employees, Alice’s salary must be much greater than €50,000 to cause this large shift of the mean. This example illustrates that even aggregate information may allow us to draw conclusions about individuals in the data.

Now, suppose instead that the company runs a differentially private computation on the employee records; in other words, they introduce a carefully tuned amount of random noise into the statistics that are produced. This results in an approximate answer to the question of average salaries. It may now be that the company reports an average salary below €50,000 even after Alice, the new high-earning employee, joins the company.

Reporting meaningful statistics without revealing some information on individuals is impossible. In the example above, it is reasonable to assume that each worker is at least above a certain salary – such as a minimum wage – and may not exceed some upper limit. Given more information on the company structure and number of employees, it may be possible to define a range of likely salaries for a certain individual. In consequence, every report of summary statistics provides at least some information on individuals and violates personal privacy. Even DP cannot fix that. Instead, DP provides a measure for *how much* individual privacy is violated and makes it possible to set an upper bound for this violation. In other words, DP allows us to set a threshold for the amount of privacy loss.

### The privacy loss parameter

An essential component of a differentially private computation is the privacy loss parameter, usually denoted by  $\epsilon$ . This parameter determines the amount of noise added to the computation, and therefore defines what can be learned about an individual as a result of their private information being included in a differentially private analysis. The parameter  $\epsilon$  can be freely chosen: the bigger  $\epsilon$  is, the more accurate the statistical results can be, but the worse the privacy protection gets. From a privacy perspective, the ideal  $\epsilon$  is zero, but that would yield useless results because they would not be allowed to contain any information. Although guidelines for



**Figure 1:** Different Laplace distributions with scale parameter  $1/\epsilon$  centred around the true value. The bigger the acceptable privacy loss (i.e., the higher the value of  $\epsilon$ ), the sharper the peak becomes, while the tails approach zero more quickly.

► choosing  $\epsilon$  have not yet been developed, a rule of thumb is to set it to a small number, between approximately 0.001 and 1.<sup>2</sup> Choosing  $\epsilon$  can be thought of as using a tuning knob to balance privacy and accuracy, determining the amount of noise to be added to the computation.

A common way to add noise in a differentially private computation is to use the Laplace distribution. Because it has the form of two exponential distributions joined together, the Laplace distribution is also known as the double exponential distribution, or the two-sided exponential distribution.<sup>3</sup> A DP algorithm generates a Laplace distribution with the location parameter at the true value and the scale parameter as  $1/\epsilon$ . It then outputs a randomly drawn value from this distribution as the new value. This means that  $\epsilon$  directly controls the variance of the distribution, and therefore its randomness. A lower  $\epsilon$  guarantees more variation and therefore higher privacy because the random value is more likely to be further from the true value (see Figure 1). Conversely, increasing  $\epsilon$  concentrates the results more strongly around the true value, resulting in higher analytical accuracy.

One of the most powerful features of DP is its robustness under composition – that is, the accumulation of risk across multiple analyses. In our earlier example, it might be possible that the company releases average salaries stratified by age groups and sex. One individual's data would then be included in several publications, allowing an attacker to combine this information, assuming the attacker knows the age and sex

of an individual. In contrast to many other anonymisation methods, DP still makes it possible to reason about and bound the privacy risk that accumulates when multiple differentially private computations are performed on the same individual's data. Executing an  $\epsilon$ -DP algorithm twice results in a cumulative privacy risk no greater than  $2\epsilon$ . Triple execution results in  $3\epsilon$ , and so on. So, if we anticipate that, say, five different DP calculations will be made on the same data, we need only set each DP algorithm to  $\epsilon/5$  to limit the privacy risk to  $\epsilon$ .

## Reporting meaningful statistics without revealing some information on individuals is impossible

### The privacy bounding risk

The parameter  $\epsilon$  not only configures the amount of noise but also sets an upper limit for the privacy loss. To understand this, we now hypothetically assume our example company asks each employee if they want their salary information included in the monthly reports. The company provides a strong incentive for employees to include their data, and they also promise a very strong  $\epsilon = 0.01$  to preserve privacy. We now take the perspective of the new employee, Alice: should she have her salary included in the reports? This problem basically results in two databases,  $D_1$  and  $D_2$ . Let  $D_2$  contain the information from Alice, in addition to

all the other employees, while  $D_1$  does not. This means  $D_2$  differs from  $D_1$  in only one additional entry. Using these two databases, we can define how  $\epsilon$  bounds the privacy loss:

$$\Pr[A(D_1) = R] \leq \exp(\epsilon) \cdot \Pr[A(D_2) = R]$$

where  $A$  is an algorithm or a process, and  $R$  is the result of this query. This equation is the formal definition of DP.

Differential privacy ensures that an individual will be subject to roughly the same privacy risk, whether or not their data are included in a differentially private analysis. We consider the privacy risk associated with the release of data to be the possible harm that an individual may encounter as a result of the assumption that an observer forms based on the release of the data.<sup>2</sup>

Using the formula above, Alice can now calculate the privacy risk of having her salary data included in the monthly reports. If  $\epsilon = 0.01$ , then Alice's privacy risk resulting from having her salary data included in a differentially private computation grows, at most, by a multiplicative factor of  $\exp(0.01) = 1.01$ . So how does this help Alice make an informed decision on whether to include her information in the salary reports?

Assume Alice does not want to disclose her salary to her colleagues. She predicts the probability that her colleagues discover her salary at 5%. However, the incentive given to every employee to include their salary information in the monthly reports is a €100 voucher for the staff canteen. Using the DP formula, Alice can calculate her privacy risk as  $5\% \cdot \exp(0.01) = 5.05\%$ . Thus, DP guarantees that, given the company uses a value of  $\epsilon = 0.01$ , the estimated probability that Alice's salary information becomes public can increase from 5% to 5.05% at most. Note that this can drastically increase with higher  $\epsilon$  values. For example, setting  $\epsilon = 0.8$  would more than double the risk from 5% to  $5\% \cdot \exp(0.8) = 11.13\%$ .

In this way, DP allows Alice to decide between the trade-off of getting the €100 voucher and the increased risk of having her salary identified. Be aware that this is a very hypothetical and simplified example and may not accurately reflect Alice's privacy risk, as we see in the next section.

### The attacker's perspective

A better way to understand the DP formula



**Christoph Kurz** is a postdoctoral researcher in health economics at the Ludwig-Maximilians-Universität Munich, Germany.

## Key messages

- **Differential privacy (DP) combines the requirement to protect the privacy of individuals with the desire to gain insight by analysing data.**
- **Privacy loss in a DP analysis can be quantified, proven and regulated.**
- **Still, there are many pitfalls of DP that make it difficult to use for the average researcher.**

is to take the position of an attacker. For example, an attacker gets to see result  $R$  of algorithm  $A$  and wants to draw conclusions on individual properties of some included person. In the simplest case, the attacker wants to know if Alice's salary is included in the monthly reports, that is, whether  $R$  is based on  $D_1$  (Alice not included) or  $D_2$  (Alice included). The DP formula states that  $A$  is an  $\epsilon$ -DP algorithm if the probability of  $A$  producing  $R$  when executed on  $D_1$  is at most  $\exp(\epsilon)$  more likely than the probability of  $A$  producing  $R$  when executed on  $D_2$ . Note that the DP definition is symmetric: it makes no difference whether  $D_2$  differs from  $D_1$  by one additional entry, or  $D_1$  differs from  $D_2$  by a missing one. This interchangeability allows a lower bound to be set for the difference of probabilities, by  $\exp(-\epsilon)$ .

However, an algorithm is not differentially private if it only protects Alice's entry. It must protect every single entry in a database. Any

$\epsilon$ -DP algorithm must return equally likely results, no matter if it is based on  $D_1$  or  $D_2$ . How likely depends on  $\epsilon$ . Recall that  $\epsilon$  defines the amount of added noise: the smaller it is, the more similar the results are.

Let us revisit our example of Alice and the average salaries. In the introduction, we learned that outliers present extra challenges when reporting averages. Remember that Alice's salary is much higher than the average, and therefore causes a large shift of €1,000 to the average when included. So, what happens when we add noise picked from the Laplace distribution with scale parameter  $1/\epsilon$  to our computation? This would result in the two databases shown in Figure 2(a).

It is obvious that these distributions make it extremely easy for an attacker to distinguish between a result from  $D_1$  or  $D_2$ . Again,  $D_2$  contains Alice's information and  $D_1$  does not. The difference between the curves is very large because the results differ in more than one unit (here, €), even though the databases differ in only one entry. High values have a disproportionately large effect on averages. The ratio of the two Laplace distributions in this example is at most  $1,000\epsilon$ , so using a parameter of  $1/\epsilon$  only gives  $1,000\epsilon$ -DP.

To get  $\epsilon$ -DP we must add more noise. How much noise depends on the maximum contribution of individual entries. In the case of Alice, we have to add 1,000 times more noise by using a Laplace distribution with scale parameter  $1,000/\epsilon$  to make  $D_1$  and  $D_2$  less distinguishable (see Figure 2(b)). However,

**An algorithm is not differentially private if it only protects one entry. It must protect every single entry in a database**

this choice might change if higher salaries than Alice's (or extremely low salaries) are possible for future hires. In practice it is best to estimate the individual contributions very conservatively. For example, if we assume that all salaries are between €10,000 and €100,000 then the greatest possible difference is €90,000. This occurs when the attacker tries to determine whether a new hire earns €10,000 or €100,000. As before, we must add  $\text{Laplace}(90,000/\epsilon)$  noise before dividing by the number of employees to achieve  $\epsilon$ -DP. If later it turns out that there are values outside the assumed extremes, these values must be clamped. This skews the statistics but is required to guarantee DP.

## Summing up

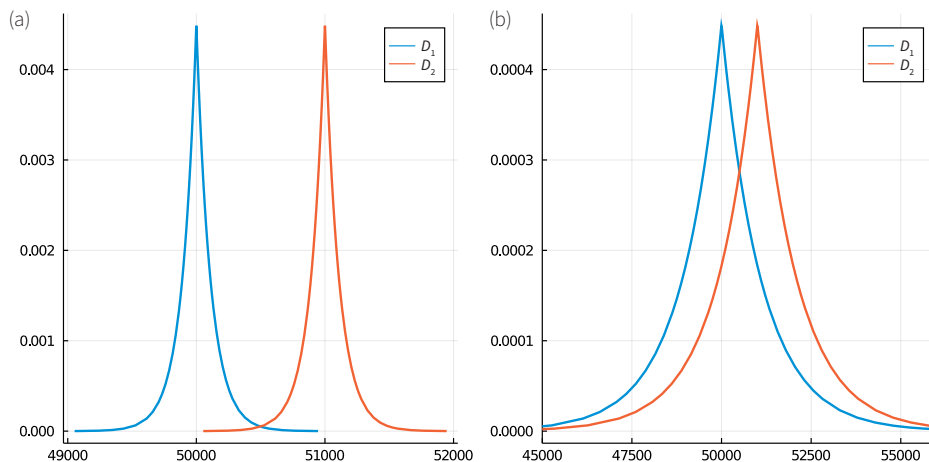
Differential privacy is a powerful concept but difficult to apply in practice. It requires a lot of consideration when reporting averages and it becomes even more complicated for other statistical measures such as variances, maximum values or histograms. Helpfully, there exists open-source software that automates DP, such as OpenDP by Microsoft and Harvard ([bit.ly/3ejoK2h](https://bit.ly/3ejoK2h)), diffprivlib by IBM ([bit.ly/3ekICCT](https://bit.ly/3ekICCT)) and Google's own library ([bit.ly/3x8lxvb](https://bit.ly/3x8lxvb)). ■

## Disclosure statement

The author declares no competing interests.

## References

1. Dwork, C., McSherry, F., Nissim, K. and Smith, A. (2006) Calibrating noise to sensitivity in private data analysis. In S. Halevi and T. Rabin (eds), *Theory of Cryptography* (pp. 265–284). Berlin: Springer.
2. Nissim, K., Steinke, T., Wood, A., Altman, M., Bembeneke, A., Bun, M., Gaboardi, M., O'Brien, D. R. and Vadhan, S. (2017) Differential privacy: A primer for a non-technical audience. Working paper, Privacy Tools for Sharing Research Data project, Harvard University. [bit.ly/2PivfKs](https://bit.ly/2PivfKs)
3. Geraci, M. and Borja, M. C. (2018) Notebook: The Laplace distribution. *Significance*, 15(5), 10–11.



**Figure 2:** (a) High  $\epsilon$  makes  $D_1$  and  $D_2$  easy to distinguish from the attacker perspective. (b) Lower  $\epsilon$  makes  $D_1$  and  $D_2$  less distinguishable.