



Katsushika Hokusai (葛飾 北斎)

Enoshima in Sagami Province (相州江の島)

*from Thirty-six views of Mount Fuji*



Supported by the  
IEEE Information Theory Society  
Distinguished Lecturer Program

# An information theorist's tour of differential privacy

Anand D. Sarwate, Rutgers University

24 October 2025

Centro de Investigación en Matemáticas (CIMAT)  
Guanajuato, Guanajuato, Mexico

# Some thanks and credits

Paper: S., Calmon, Kosut, Sankar, to appear in BITS



Thanks for helpful discussions with  
**Shahab Asoodeh (McMaster)**  
**Flavio Calmon (Harvard)**  
**Oliver Kosut (Arizona State)**  
**Lalitha Sankar (Arizona State)**  
**Mario Diaz (UNAM) - in memoriam**

## Image credits:

- Wikimedia Commons
- ARC Ukiyo-e dataset
- [OpenMoji.org](https://openmoji.org)

# Mario Diaz

We are thankful for his insights

- Mario is a “local” to CIMAT:
  - MS (2013)
  - Postdoc (2018-2019)
- Discussions with Mario were a **strong motivation** for writing a tutorial on the connections between information theory and differential privacy.



# Today's talk

I will tell you stuff you know already (possibly? probably?)

Differential privacy is a way of quantifying how different two distributions are.  
Information theorists also think about this kind of thing.

But: the kinds of questions/settings can have a different flavor. Today:

[hypothesis testing](#)

[f-Divergences](#)

[contraction coefficients](#)

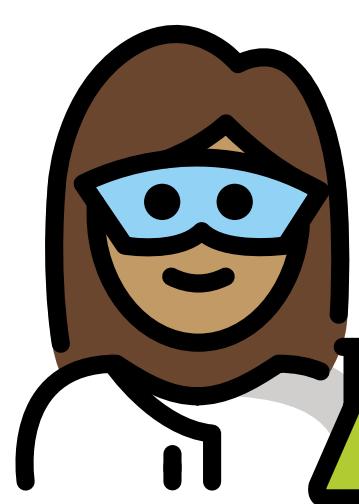
Goals:

- Describe some of these connections for those less familiar
- Open some new avenues for research

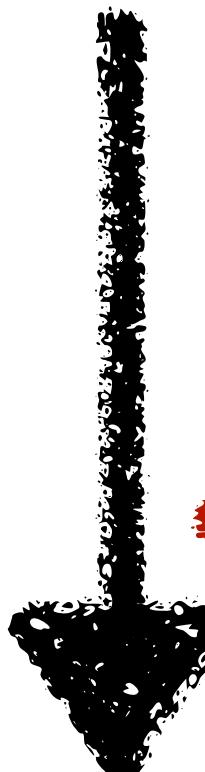
# Privacy at its simplest: protect 1 secret bit

## A binary hypothesis test

Sara



$$s \in \{0,1\}$$



$$Y \sim P_{Y|S=s}$$



$$\hat{s} \in \{0,1\}$$



Benito

Sara wants to share information to a collaborator Benito.

- She shares a random variable  $Y$ .
- The distribution of  $Y$  depends on a **secret**  $s \in \{0,1\}$ .
- How should she design  $Y$  to not reveal  $s$ ?

The **privacy question** is a **hypothesis testing question**:

$$\mathcal{H}_0: Y \sim P_{Y|S=0}$$

$$\mathcal{H}_1: Y \sim P_{Y|S=1}$$



The Lake of Hakone in  
Sagami Province

相州箱根湖水

Sōshū Hakone Kosui

Vista 1

hypothesis testing

# Neyman-Pearson tells us the optimal test

Adversarial inference is a generalized LRT

The optimal test for the adversary is a  
**likelihood ratio test**:

$$\hat{s}(y) = \begin{cases} 1 & \log \frac{P_{Y|S=1}(y)}{P_{Y|S=0}(y)} \geq \tau \\ 0 & \log \frac{P_{Y|S=1}(y)}{P_{Y|S=0}(y)} < \tau \end{cases}$$

## Example

$$\mathcal{H}_0: Y = 0 + Z \sim \mathcal{N}(0, \sigma^2)$$

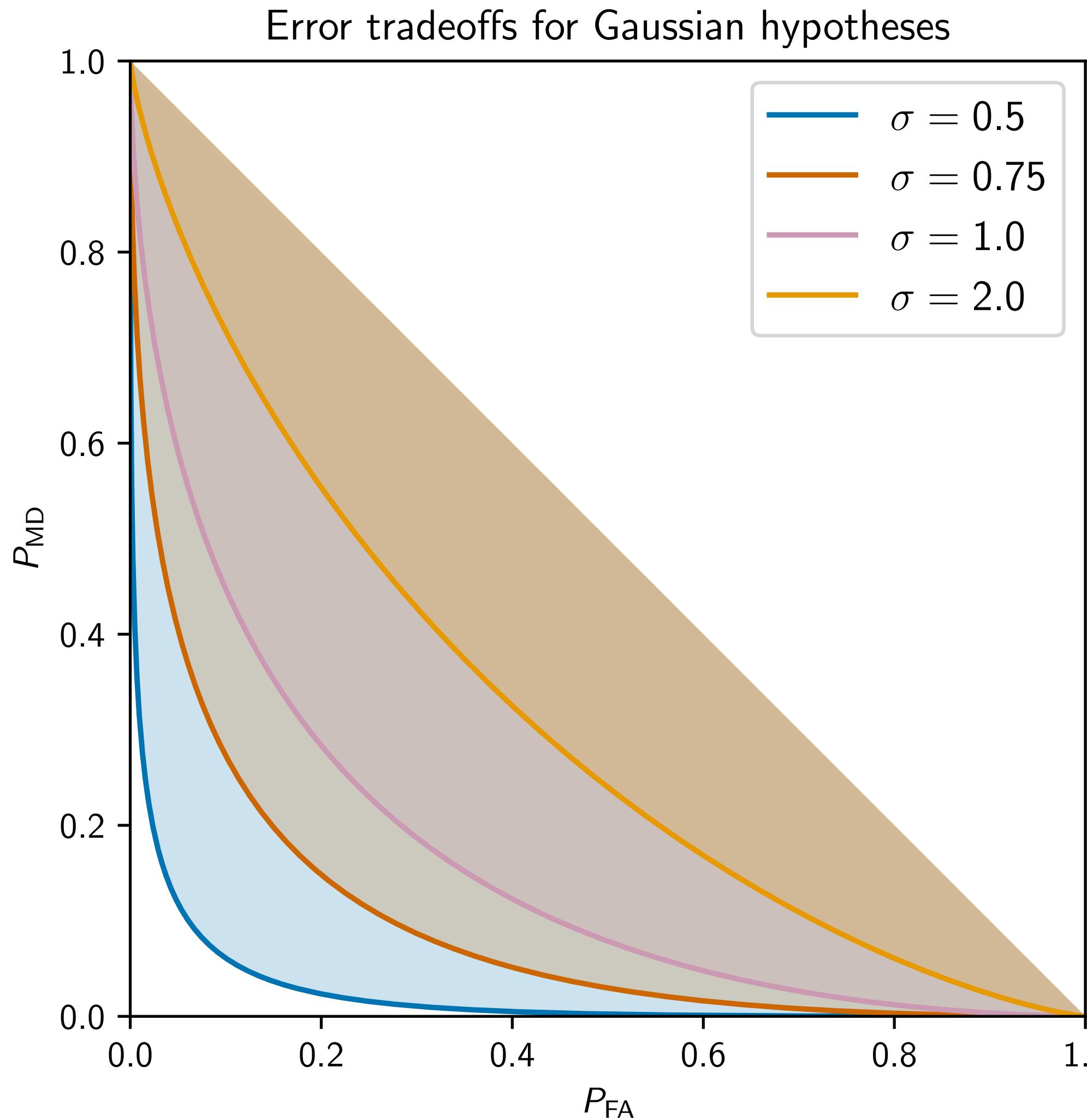
$$\mathcal{H}_1: Y = 1 + Z \sim \mathcal{N}(1, \sigma^2)$$



I ❤️ Neyman-Pearson!

# Example: additive Gaussian noise

Everyone's favorite example: Gaussians!



If the revealed information  $Z$  is **Gaussian**:

$$\mathcal{H}_0: Z \sim \mathcal{N}(0, \sigma^2)$$

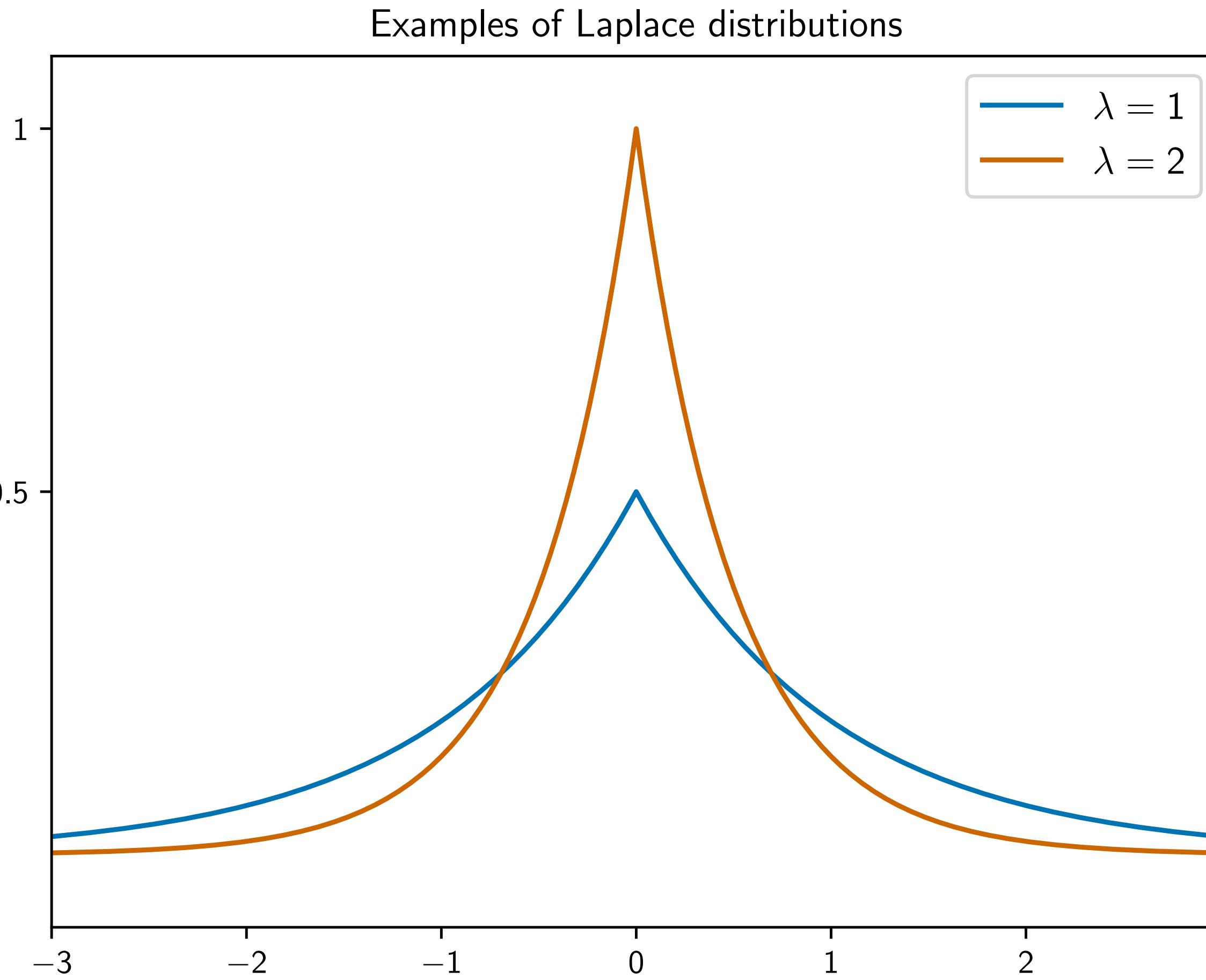
$$\mathcal{H}_1: Z \sim \mathcal{N}(1, \sigma^2)$$

We can write the error probabilities in terms of Q functions:

$$P_{\text{FA}} = Q\left(\frac{t}{\sigma}\right), P_{\text{MD}} = Q\left(\frac{1-t}{\sigma}\right).$$

# Example: additive Laplace noise

We can do more than just Gaussians!



If the revealed information  $Z$  is **Lapace**:

$$\mathcal{H}_0: X \sim \text{Laplace}(\lambda)$$

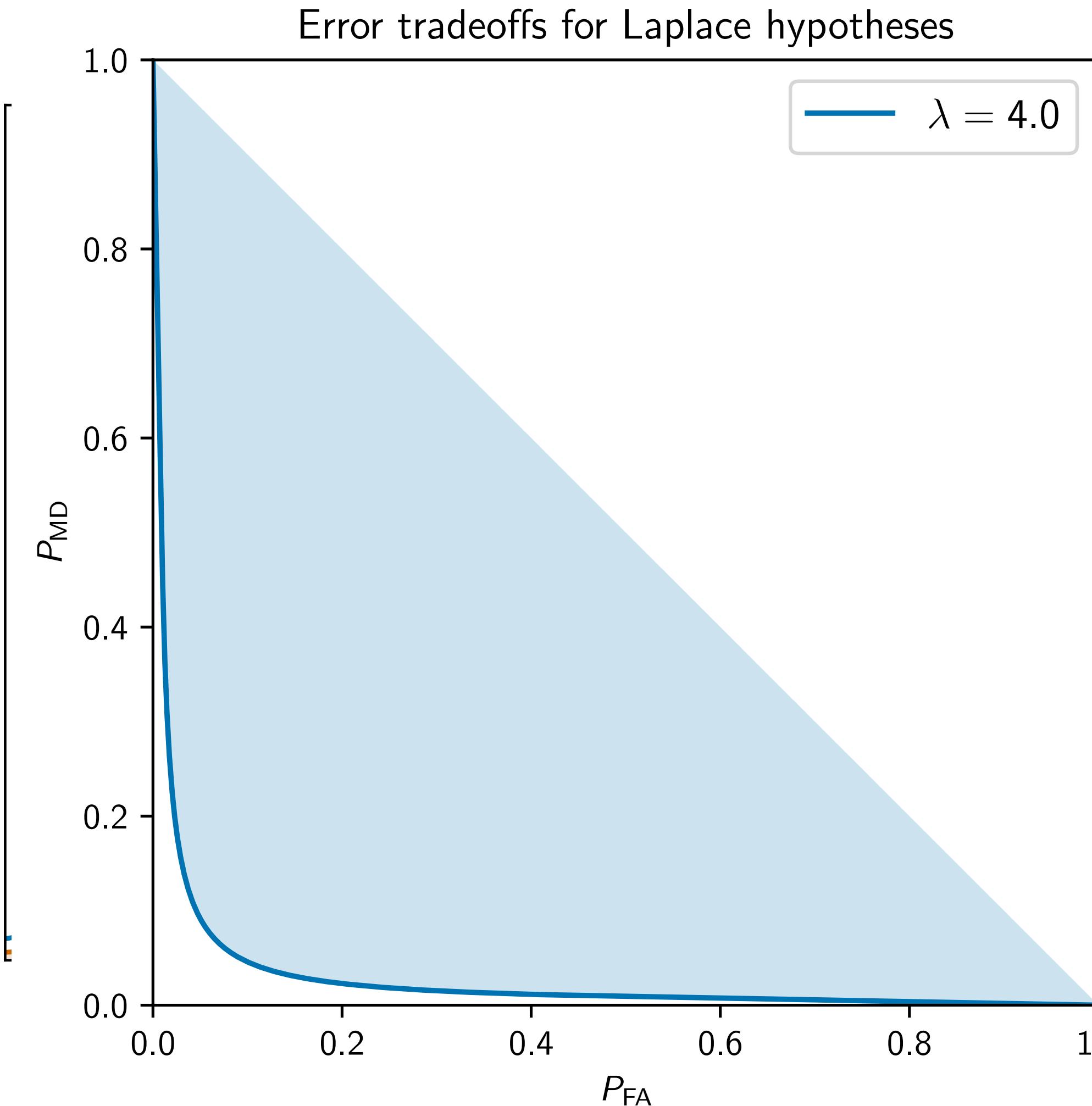
$$\mathcal{H}_1: X \sim 1 + \text{Laplace}(\lambda)$$

Where  $\text{Laplace}(\lambda)$  has density

$$p(z) = \frac{\lambda}{2} \exp(-\lambda |z|).$$

# Error tradeoffs for Laplace noise

Lighter tails give a different shape



The error probabilities for the test are:

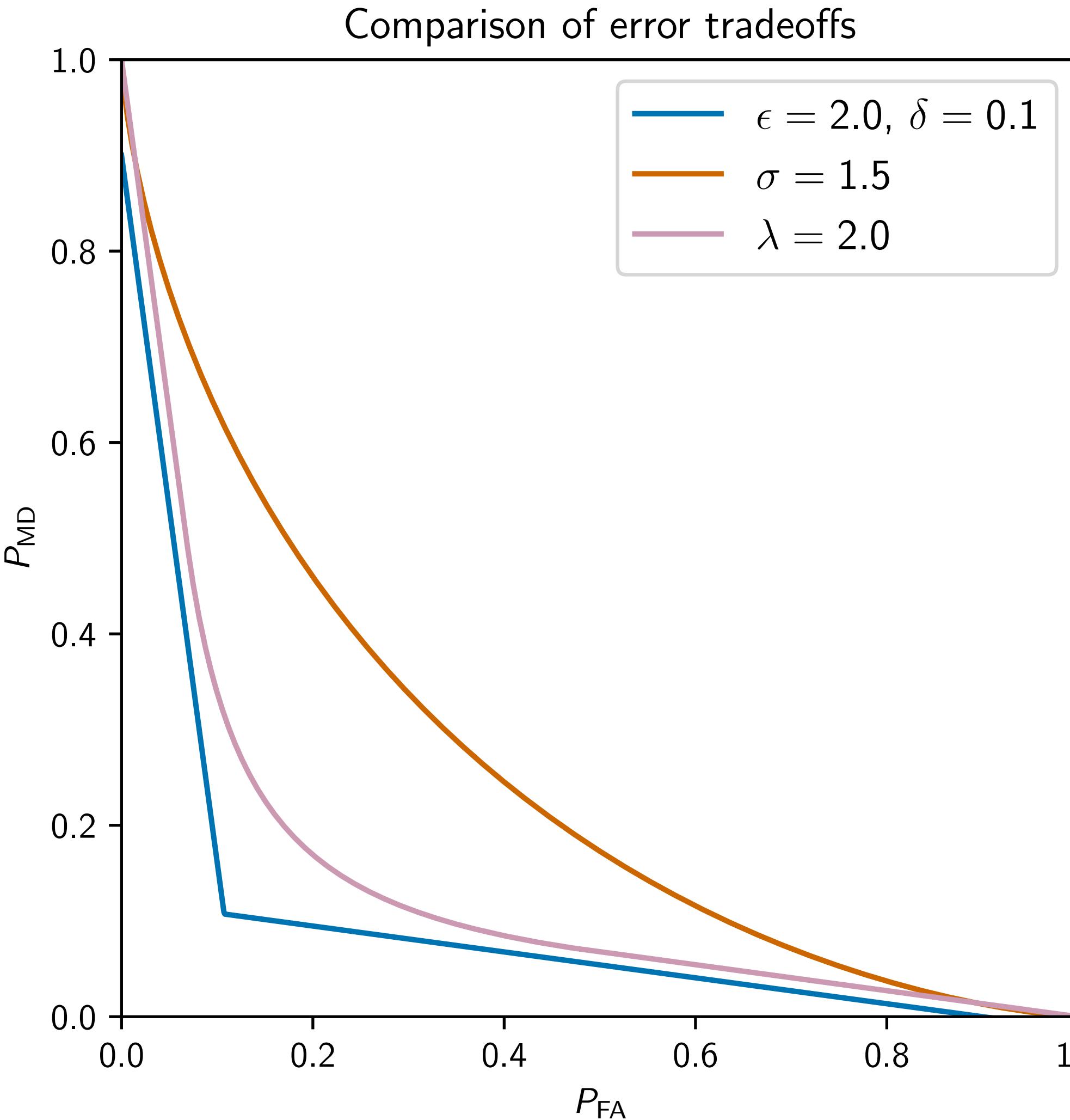
$$P_{\text{FA}} = \int_t^{\infty} \frac{\lambda}{2} \exp(-|t|\lambda) dt$$

$$P_{\text{MD}} = \int_{-\infty}^t \frac{\lambda}{2} \exp(-|A - t|\lambda) dt$$

The tradeoff is similar to the Gaussian but the slope at the corners is different.

# Tradeoffs between $P_{\text{FA}}$ and $P_{\text{MD}}$

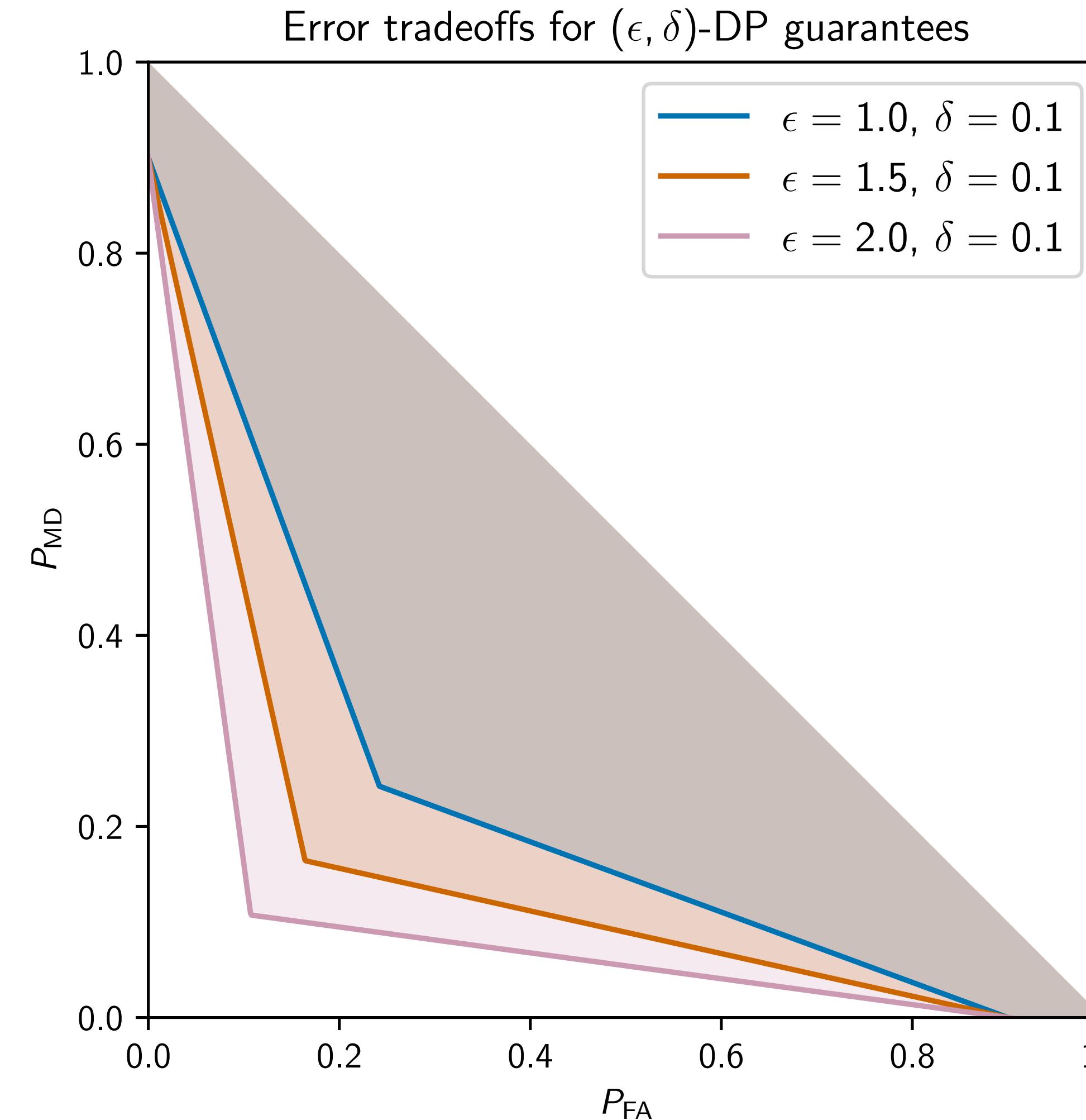
We get more privacy when the hypothesis test is “hard”



- A **privacy guarantee** is made by the tradeoff between probabilities of
  - **false alarm** (Type I error) and
  - **missed detection** (Type 2 error)
- If the likelihood ratio is small, the test will have a higher error.
- We can use a version of the ROC curve to visualize the kinds of guarantees.

# Tradeoffs between $P_{\text{FA}}$ and $P_{\text{MD}}$

Commonly used noise distributions do not meet the lower bound



A **privacy guarantee** is made by the tradeoff between probabilities of **false alarm** (Type I error/false positive) and **missed detection** (Type 2 error/false negative).

The DP definition prescribes a **piecewise linear lower bound** on the error tradeoff:

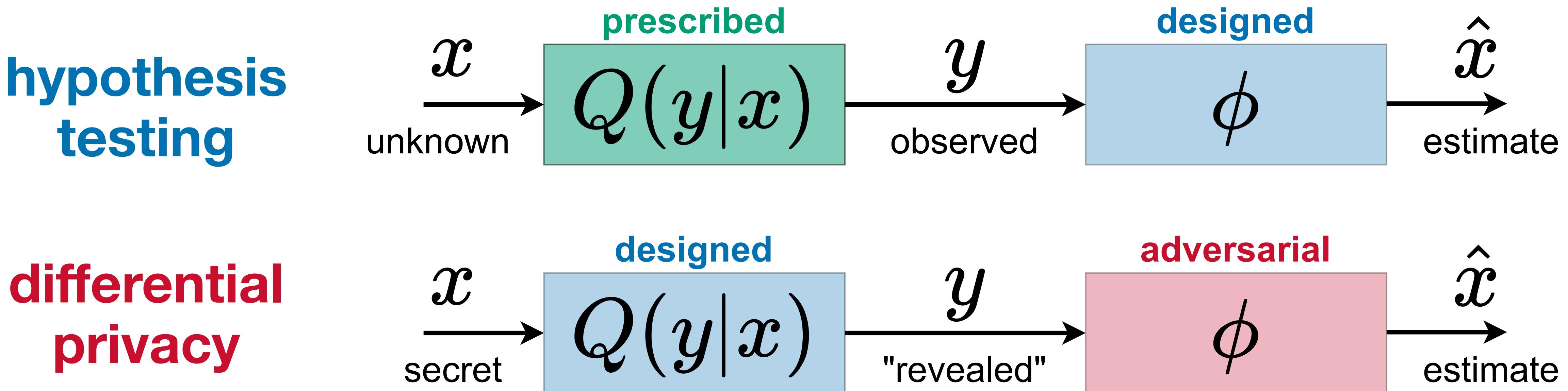
$$P_{\text{FA}} + e^\epsilon P_{\text{FA}} \geq 1 - \delta$$

$$e^\epsilon P_{\text{FA}} + P_{\text{MD}} \geq 1 - \delta$$

# Privacy versus testing

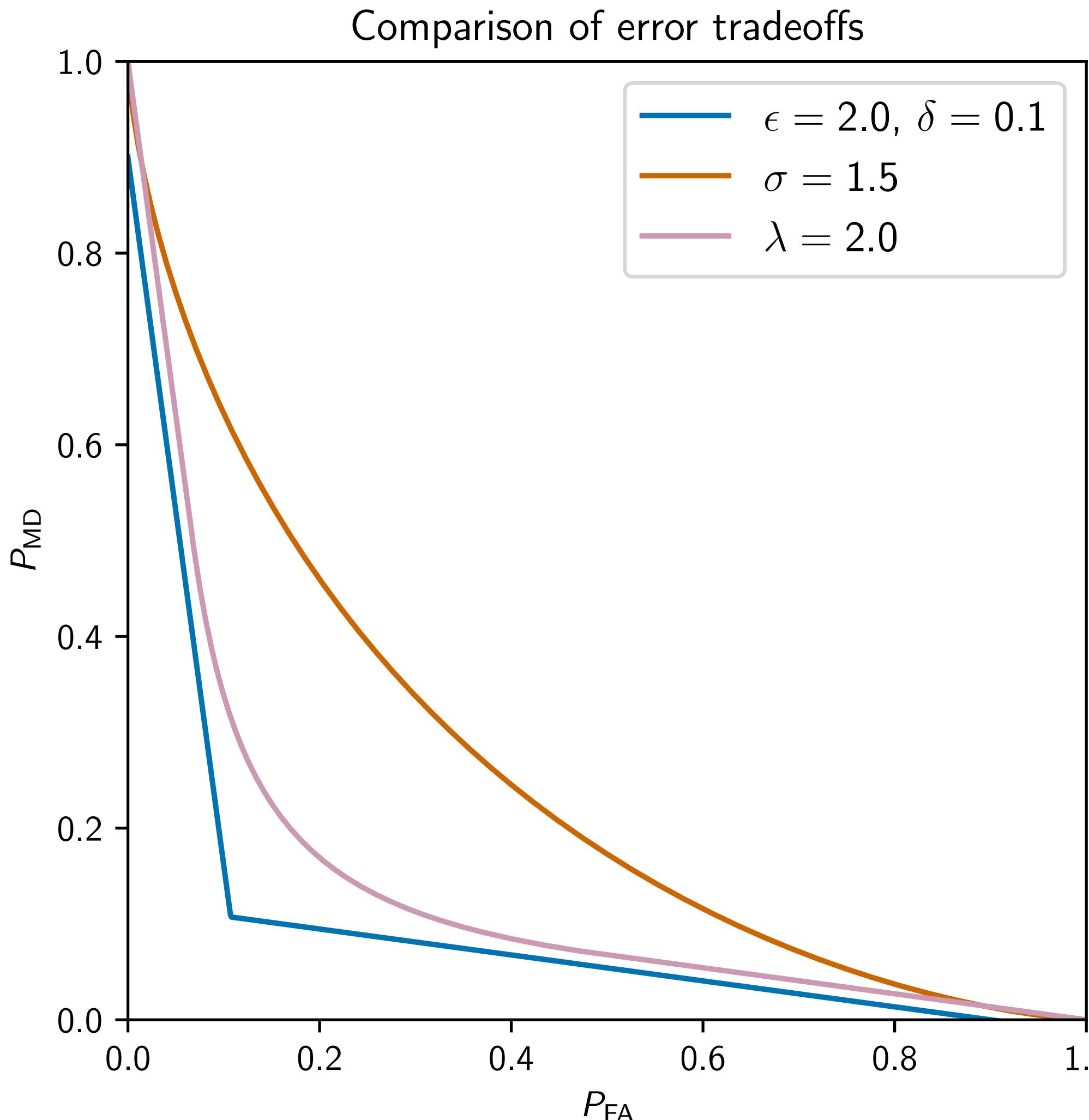
We get to design the test!

The key difference between **hypothesis testing** (as we usually encounter it) and **(differential) privacy** is that we design the **likelihoods** but not the **test**!



# Comparing hypothesis tests

Asking if one tradeoff curve is above another



We can sometimes compare two hypothesis testing problems by **comparing the optimal tradeoff curves**.

$$\mathcal{H}_0: Y \sim \mu \quad \mathcal{H}_1: Y \sim \nu$$

$$\mathcal{H}'_0: Y \sim \mu' \quad \mathcal{H}'_1: Y \sim \nu'$$

If the curve for the test  $(\mu', \nu')$  is above the curve for  $(\mu, \nu)$  then the test  $(\mu', \nu')$  is “harder” = more private:

$$T(\mu, \nu) \leq T(\mu', \nu')$$

Sunset Across Ryōgoku  
Bridge from  
Ommayagashi

御厩川岸より両国橋夕陽  
見

Ommayagashi yori  
Ryōgoku-bashi yūhi-mi



## Vista 2

differential privacy the normal way

# The “standard” approach to explaining DP

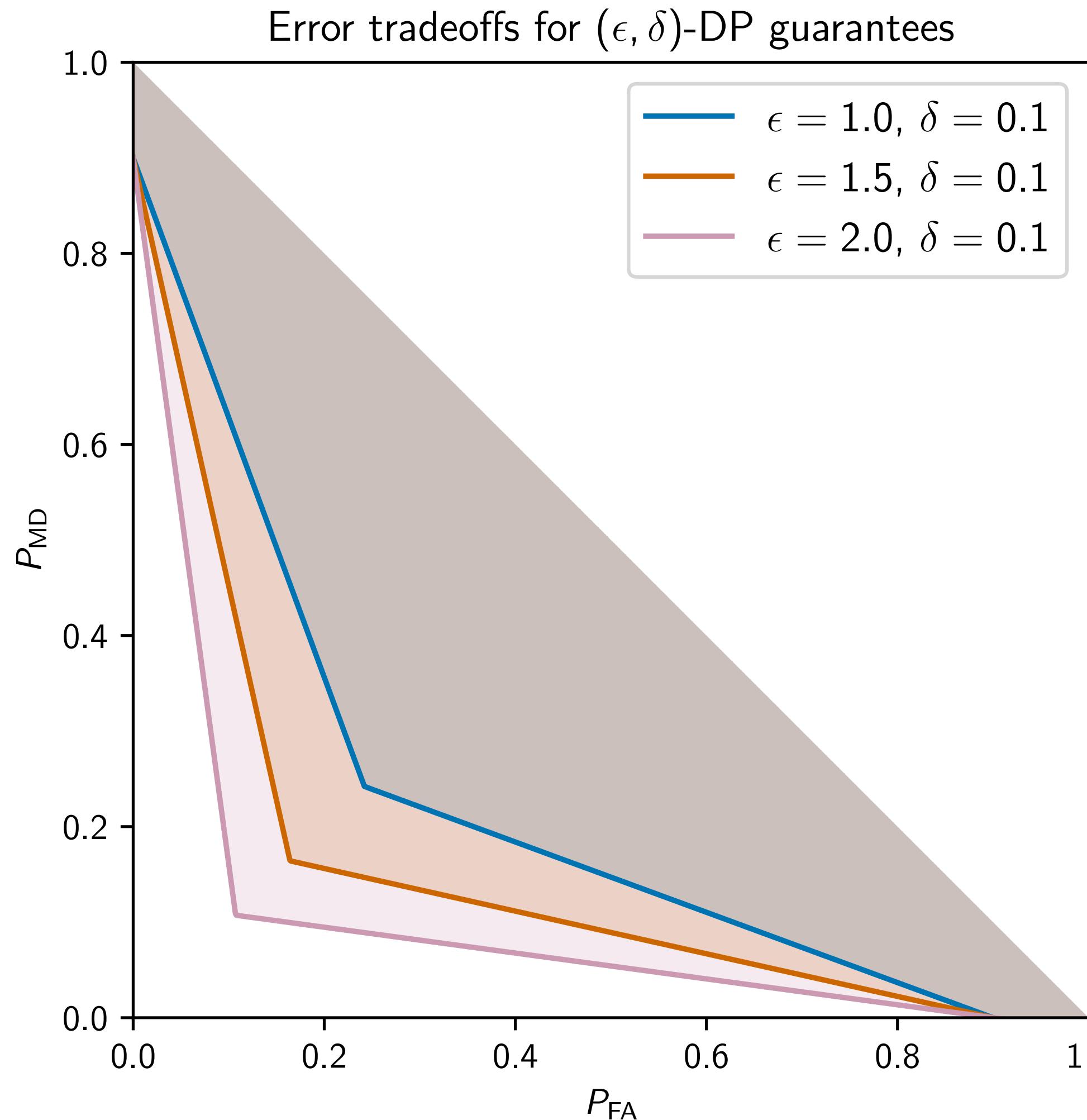
## Neighboring databases of individual records

In the textbook approach to describing DP we have several ingredients:

1. **Data space:**  $\mathcal{X}$ , often modeled as records from  $n$  individuals.
2. **Neighborhood relationship  $\sim$ :** for  $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$  we write  $\mathbf{x} \sim \mathbf{x}'$  if they are “neighbors”.
  - Example: each person has 1 bit so  $\mathcal{X} = \{0,1\}^n$  and  $\mathbf{x} \sim \mathbf{x}'$  if they differ in one position.
3. **Output space:**  $\mathcal{Y}$ , depends on the functionality/what we want to release.
  - Example: If we want the average of data  $\mathcal{X} = [0,1]^n$ , we have  $\mathcal{Y} = [0,1]$ .
  - Example: If we want to train a classifier using data  $\mathcal{X} = (\mathbb{R}^d \times \{0,1\})^n$ ,  $\mathcal{Y} = \mathbb{R}^d$ .
4. **Algorithm:** a randomized map/conditional distribution/channel  $Q: \mathcal{X} \rightarrow \mathcal{Y}$ .

# DP makes many hypothesis tests hard

Protecting many single bits simultaneously



Compared to our single private bit  $s$ , in DP we want **many hypothesis tests to hard for the adversary**. For every  $\mathbf{x} \sim \mathbf{x}'$  the test

$$\mathcal{H}_0: \mathbf{y} \sim Q(\cdot | \mathbf{x})$$

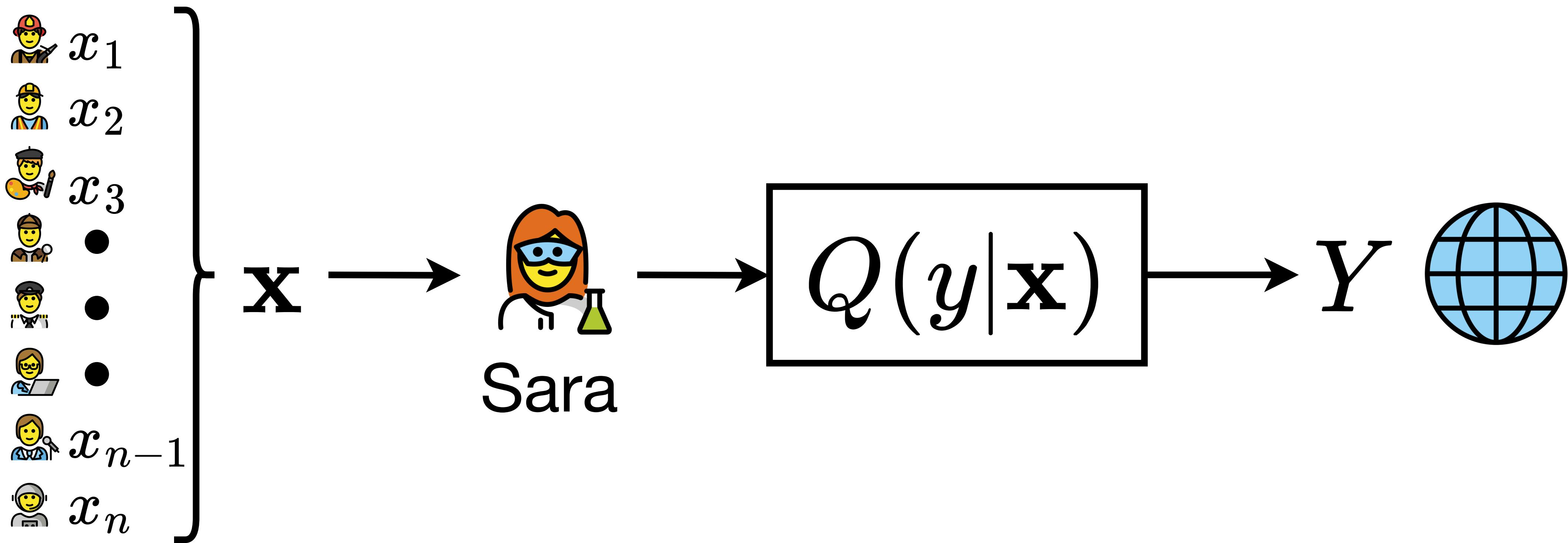
$$\mathcal{H}_1: \mathbf{y} \sim Q(\cdot | \mathbf{x}')$$

should have a large probability of error.

When can we do this? When **neighboring data sets make similar output distributions**.

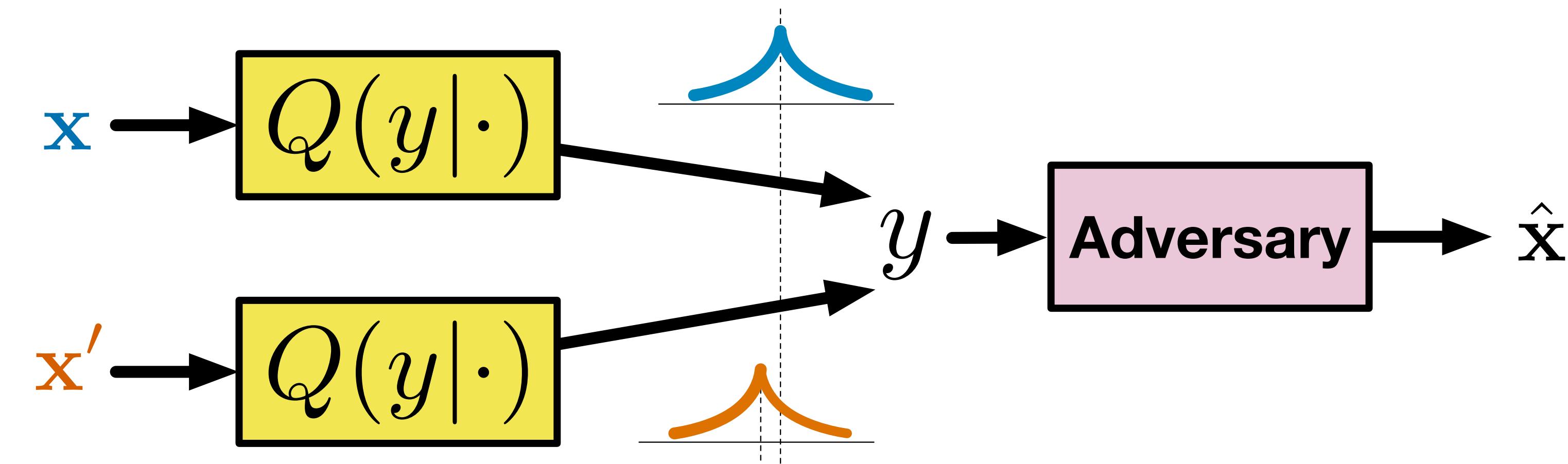
# In a snapshot

Replacing a single bit with a database



# The hypothesis testing in DP

DP is a property of the channel



A channel/“mechanism”/algorithm  $Q$  is  **$(\epsilon, \delta)$ -differentially private** if

$$Q(\mathcal{T} | \mathbf{x}) \leq e^\epsilon Q(\mathcal{T} | \mathbf{x}') + \delta$$

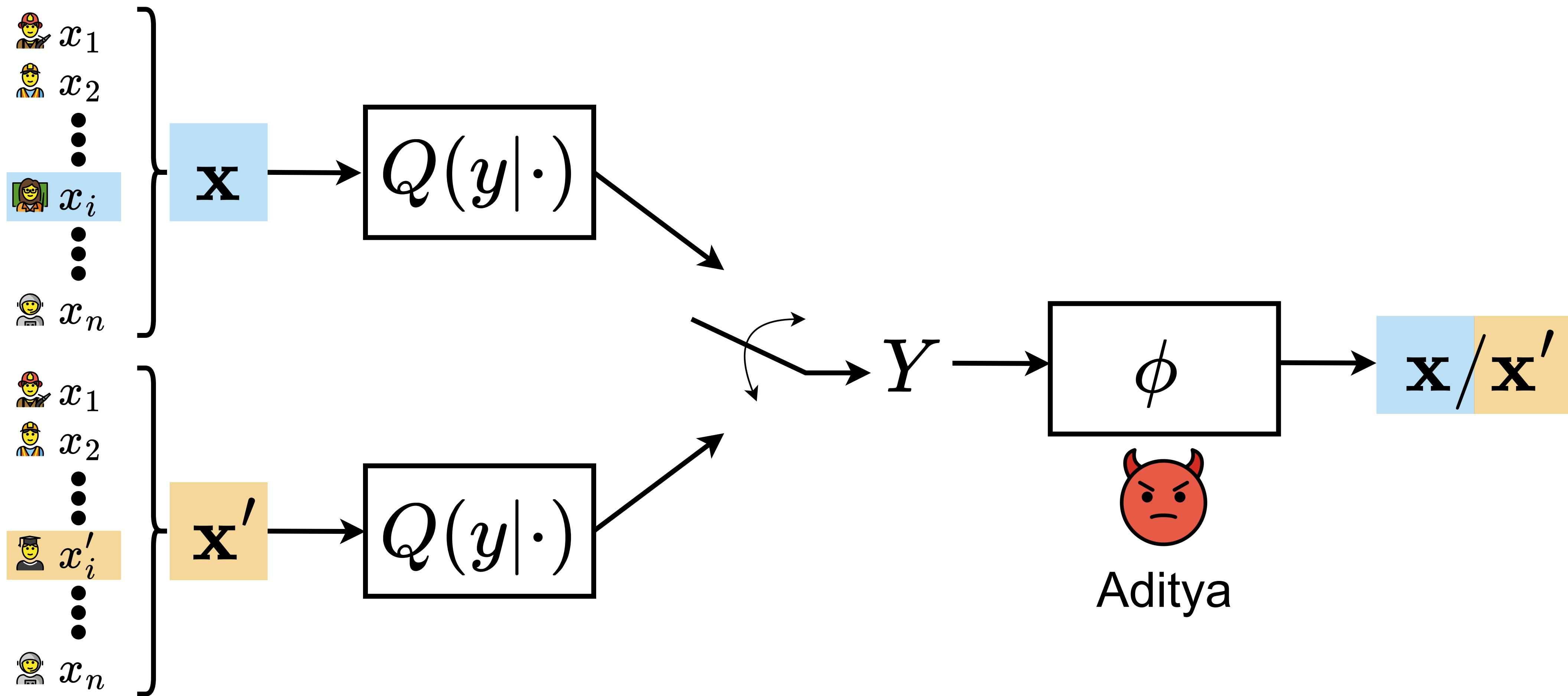
For all measurable subsets  $\mathcal{T} \subseteq \mathcal{Y}$  and all  $\mathbf{x} \sim \mathbf{x}'$ .

[Dwork-Kenthapadi-McSherry-Mironov-Naor 2006]

[Wasserman-Zhou 2010]

# Neighboring datasets in a picture

## The adversary's hypothesis test



# Some notes on the definition

DP's underlying assumptions are slightly different

- **Differential privacy is a pretty stringent requirement:** The probability of any event is similar under  $\mathbf{x}$  and any other neighboring  $\mathbf{x}'$ .
- **Guarantee is on conditional probabilities given the data:** same risk holds regardless of side information (e.g. linkage attacks).
- **There is no statistical assumption on the data:**  $\mathbf{x}$  is not drawn from some (prior) distribution.
- **The data itself is considered identifying:** no notion of some parts being personally identifiable information (PII) and others not.

# Sensitivity of scalar functions

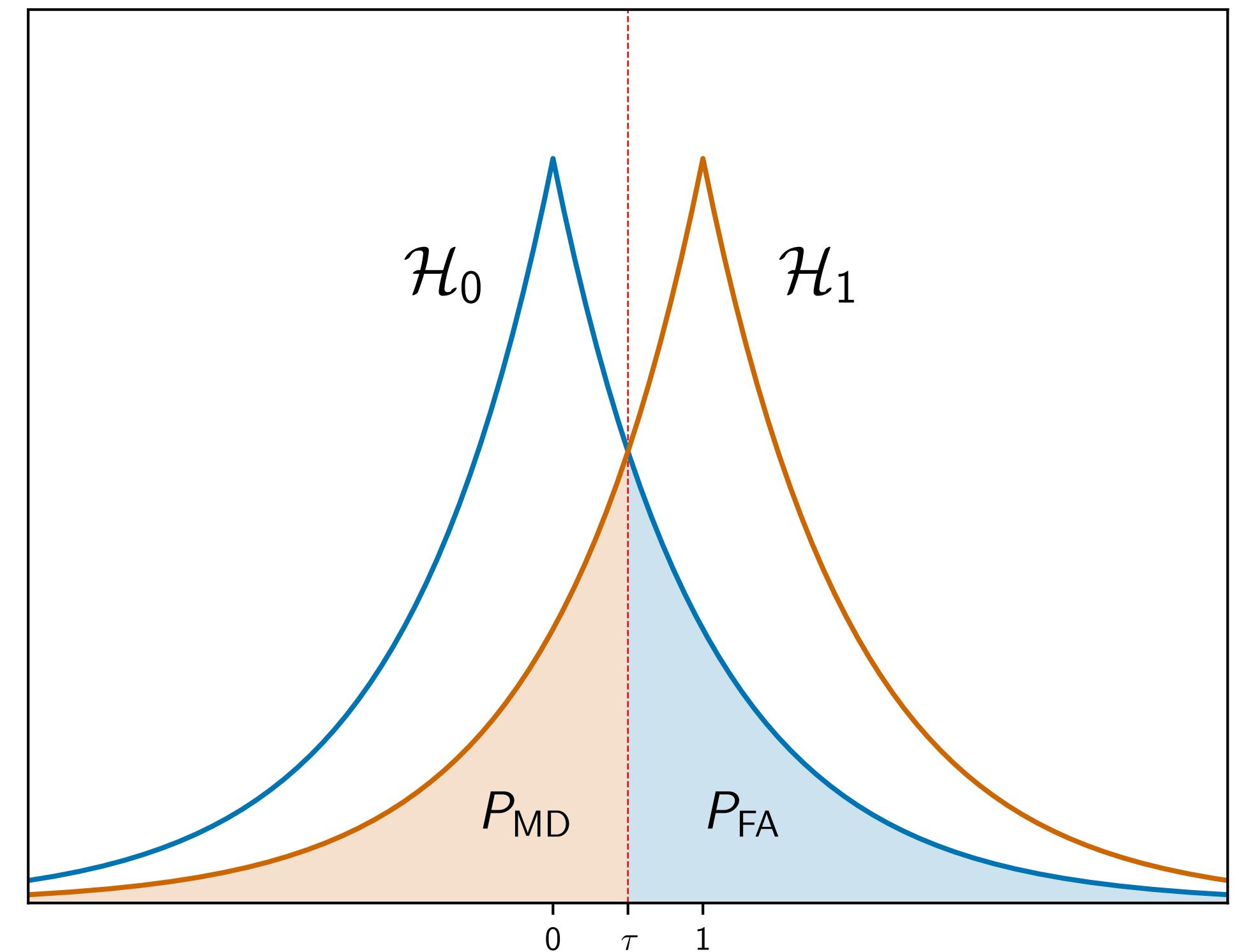
## Understanding the distance between hypotheses

In DP, we usually want to approximate some function of the data.

Suppose we want  $f: \mathcal{X} \rightarrow \mathbb{R}$ . We want the test to be hard for any pair  $(\mathbf{x}, \mathbf{x}')$  which are “neighbors”  $(\mathbf{x} \sim \mathbf{x}')$ .

If  $f(\cdot)$  is small for all neighbors, this should be easier.

Example:  $f(x) = \frac{1}{n} \sum_{i=1}^n x_i$  can change by at most  $\frac{1}{n}$  for  $x_i \in [0,1]$ .



# Sensitivity of scalar functions

Understanding the distance between hypotheses

The **global sensitivity** of a scalar function  $f(\cdot)$  is

$$\Delta(f) = \max_{\mathbf{x} \sim \mathbf{x}'} |f(\mathbf{x}) - f(\mathbf{x}')|.$$

If we use additive noise (like in the Laplace and Gaussian case) we have

$$\mathcal{H}_0: Z \sim p(z - f(\mathbf{x})) \quad \text{vs.} \quad \mathcal{H}_1: Z \sim p(z - f(\mathbf{x}'))$$

We can make a guarantee for all “neighbors” if following test is hard:

$$\mathcal{H}_0: Z \sim p(z) \quad \text{vs.} \quad \mathcal{H}_1: Z \sim p(z - \Delta(f)).$$

# Example: the sample mean

## Computing the MSE as a function of privacy risk

Suppose we have data in  $\mathcal{X} = [A, B]^n$  and want to estimate the mean:

$$\hat{\mu}(\mathbf{x}^n) = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j + Z$$

- Sensitivity of  $\hat{\mu}(\mathbf{x}^n)$  is  $(B - A)/n$ .
- $Z \sim \text{Laplace}(n\epsilon/(B - A))$  will guarantee  $(\epsilon, 0)$ -DP.
- MSE of  $\hat{\mu}(\mathbf{x}^n)$  is  $2/\lambda^2 = 2\frac{(B - A)^2}{n^2\epsilon^2}$ .

# The privacy-utility tradeoff

How much do we lose when we guarantee privacy?

Adding  $\text{Laplace}(\lambda)$  noise guarantees privacy, but at what cost? The MSE is:

$$2/\lambda^2 = 2 \frac{(B - A)^2}{n^2 \epsilon^2}$$

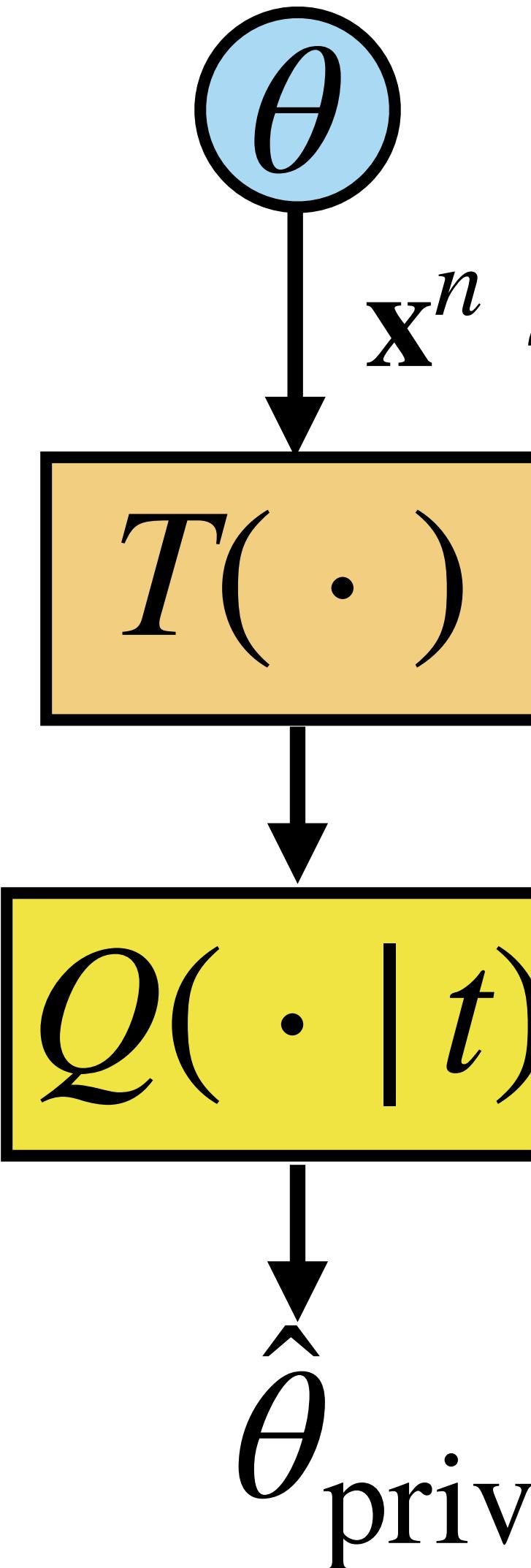
So we can see that **less privacy risk (smaller  $\epsilon$ ) induces more MSE.**

We can try to optimize the privacy mechanism if we know the utility function (like squared error).

This is what people call the **privacy-utility tradeoff**.

# Point estimation with differential privacy

Adding noise to sufficient statistics



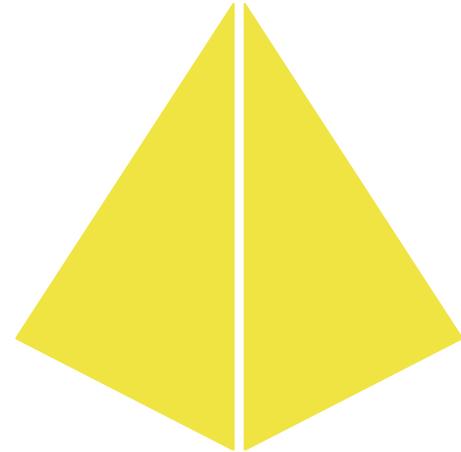
A typical DP approach to statistical estimation  
(Smith 2009):

- Model data as drawn i.i.d.  $\sim p(\mathbf{x} | \theta)$ .
- Compute a sufficient statistic  $T(\mathbf{x}^n)$  for  $\theta$ .
- Add noise to  $T(\mathbf{x}^n)$  to guarantee DP.
- Compute a “plug-in” estimate from noisy  $T(\mathbf{x}^n)$ .

All we need to know is the sensitivity of  $T(\cdot)$ .

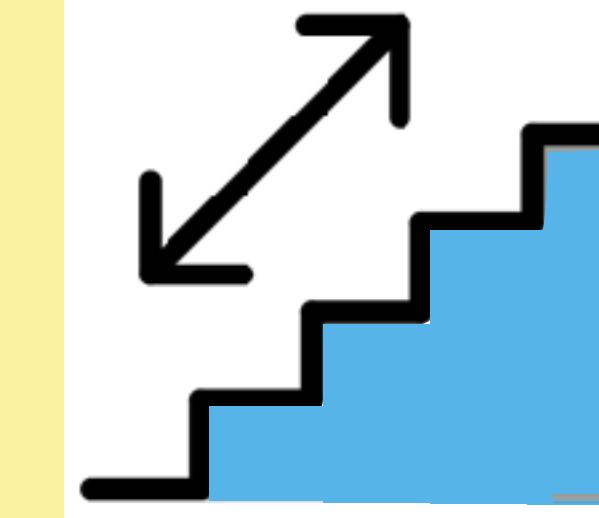
# What kind of noise should we use?

So many different choices: a non-comprehensive list



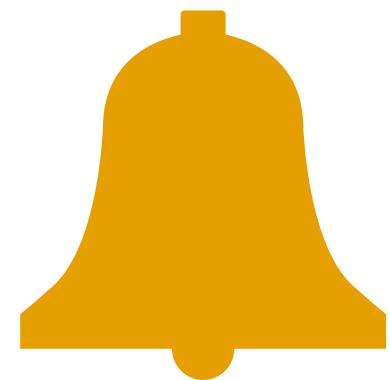
## Variations on geometric noise

Ghosh, Roughgarden,  
Sundarajan (2009/2012)  
Gupte, Sundararajan (2010)  
Balcer, Vadhan (2017)



## Variations on staircases

Geng, Viswanath (2015, 2016)  
Kairouz, Oh, Viswanath (2014)  
Geng, Kairouz, Oh, Viswanath (2015)  
Kalantari, Sankar, Sarwate (2018)



## Variations on Gaussians

Cannone, Kamath, Steinke (2020)  
Sadeghi, Korki (2022)  
Muthukrishnan, Kalyani (2023)  
Rinberg, Shumailov, Cummings,  
Papernot (2024)

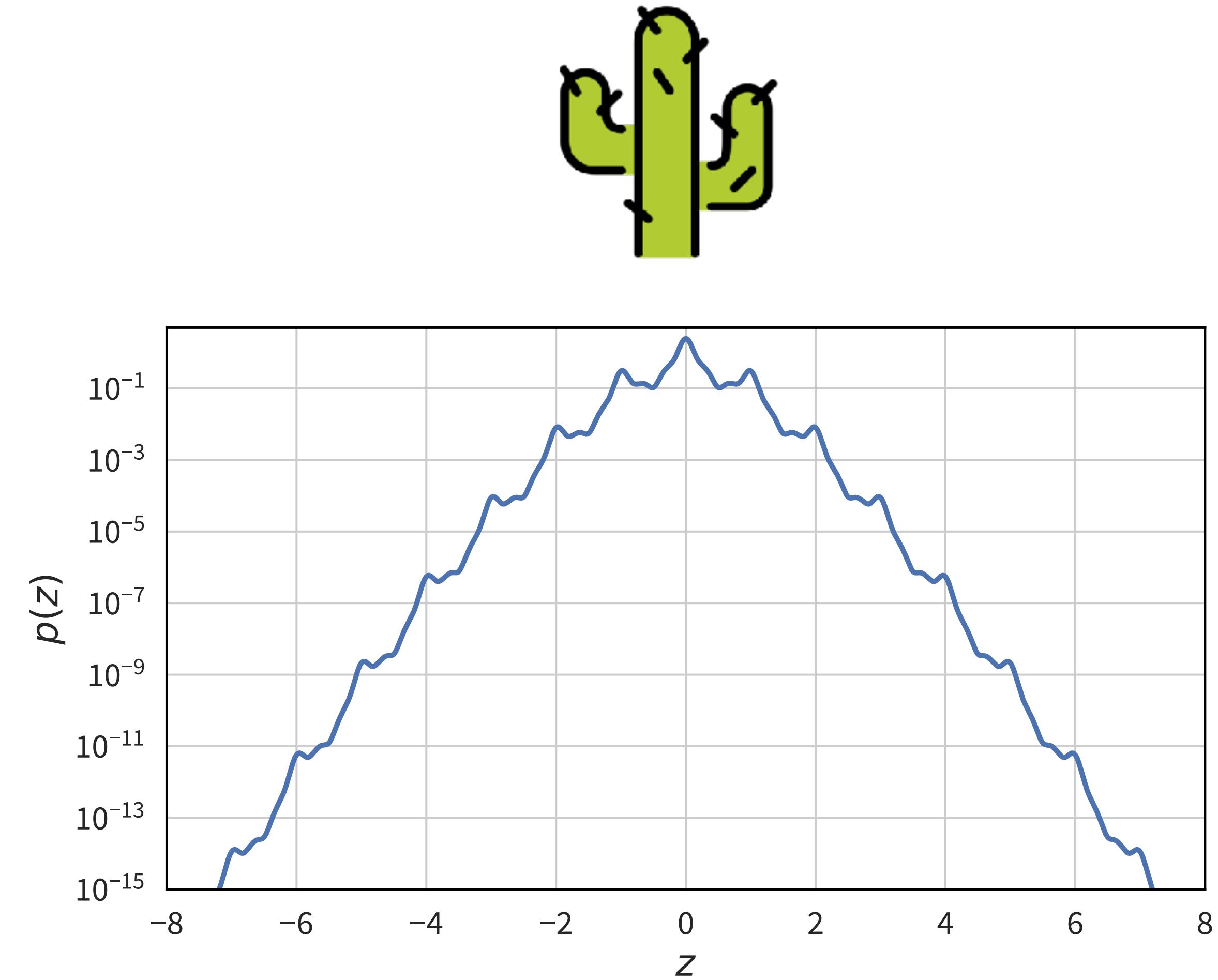
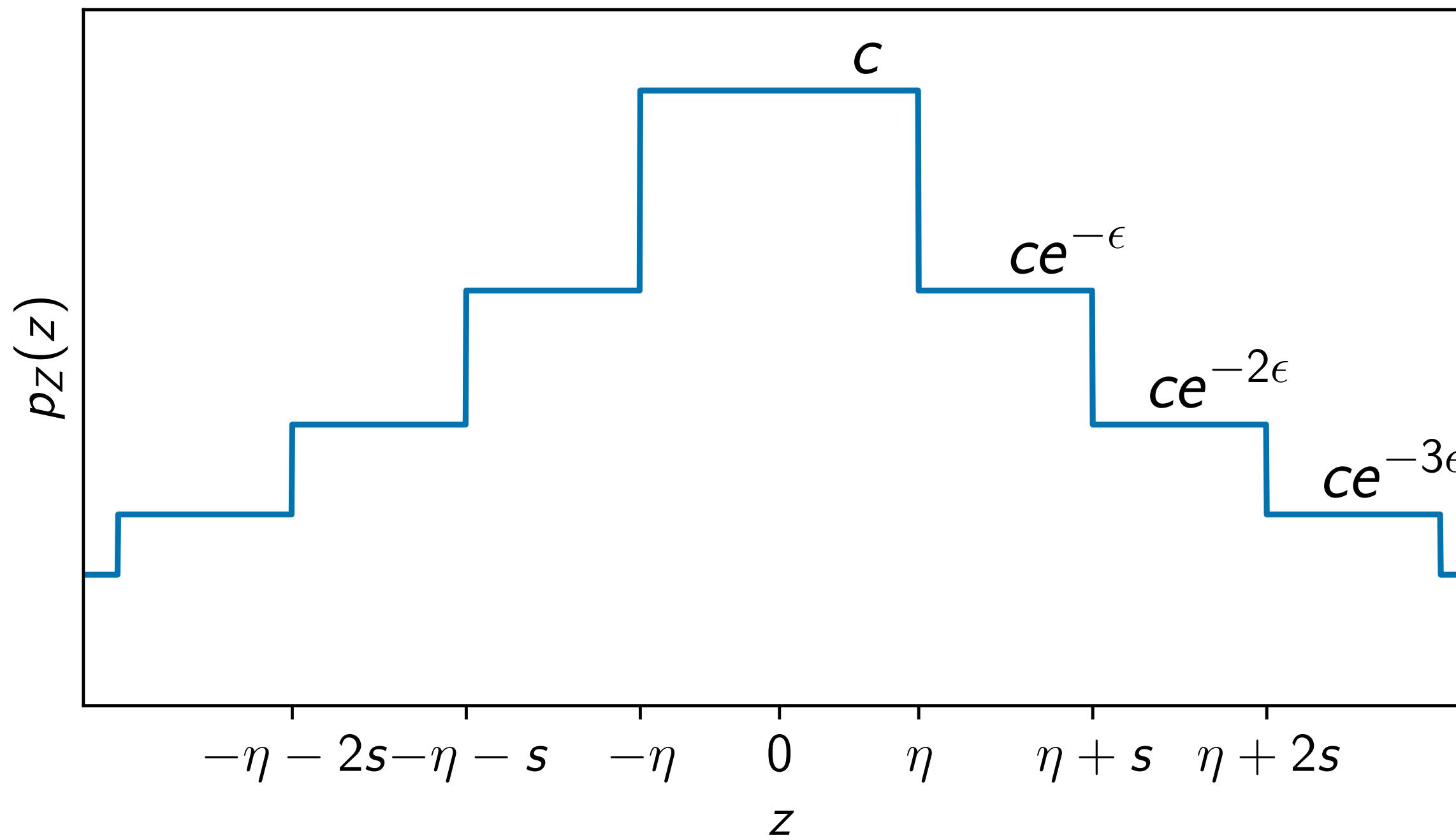
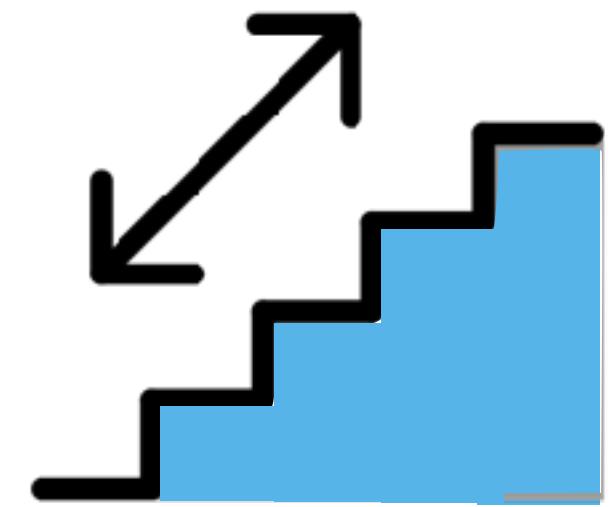


## “Other”

Geng, Ding, Guo, Kumar (2019/2020)  
Dong, Su, Zhang (2021)  
Alghamdi, Asoodeh, Calmon, Kosut,  
Sankar, Wei (2022)  
Alghamdi, Asoodeh, Calmon, Gomez,  
Kosut, Sankar (2023)

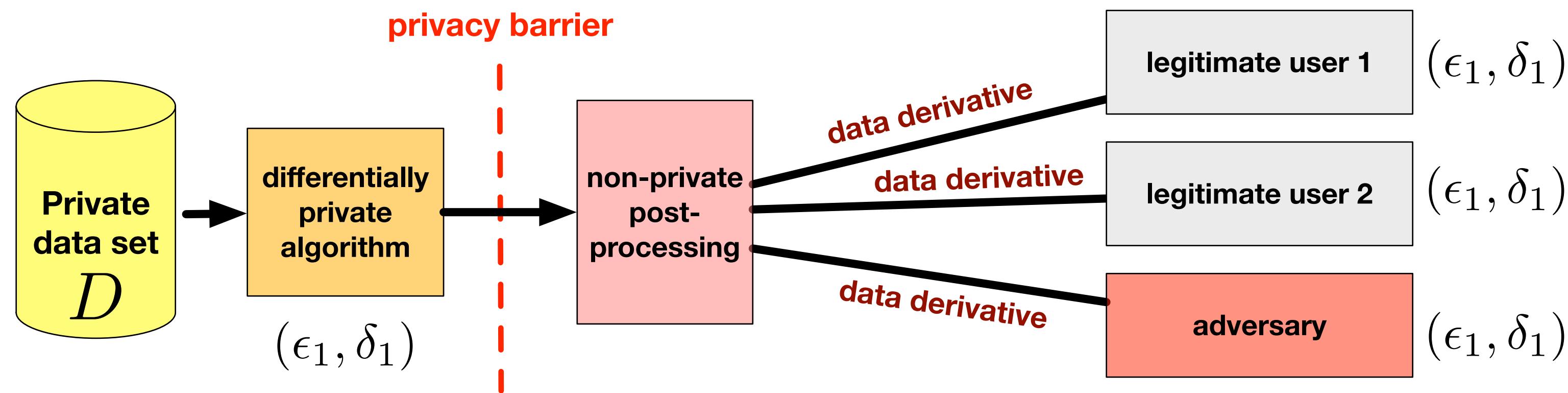
# “Optimal” noise distributions

## Beyond Gaussian and Laplace



# Post-processing invariance and composition

## Nice properties of differential privacy



- **Side-information resilience:** measures the additional risk regardless of what is known already.
- **Post-processing invariance:** once we publish something the risk cannot increase from additional computations.
- **Composition:** quantifies how privacy loss “adds up” over multiple releases.



Umezawa in Sagami  
Province

相州梅沢庄

Soshū Umezawanoshō

## Vista 3

$f$ -divergences/composition

# Privacy loss random variables

## Characterizing the distribution of the LLR

The fundamental quantity of interest is called the **privacy loss random variable** (PLRV). Switching notation a little bit, we can define the variable

$$L_{x,x'} = \log \frac{dP_{Y|X=x}}{dP_{Y|X=x'}}(Y)$$

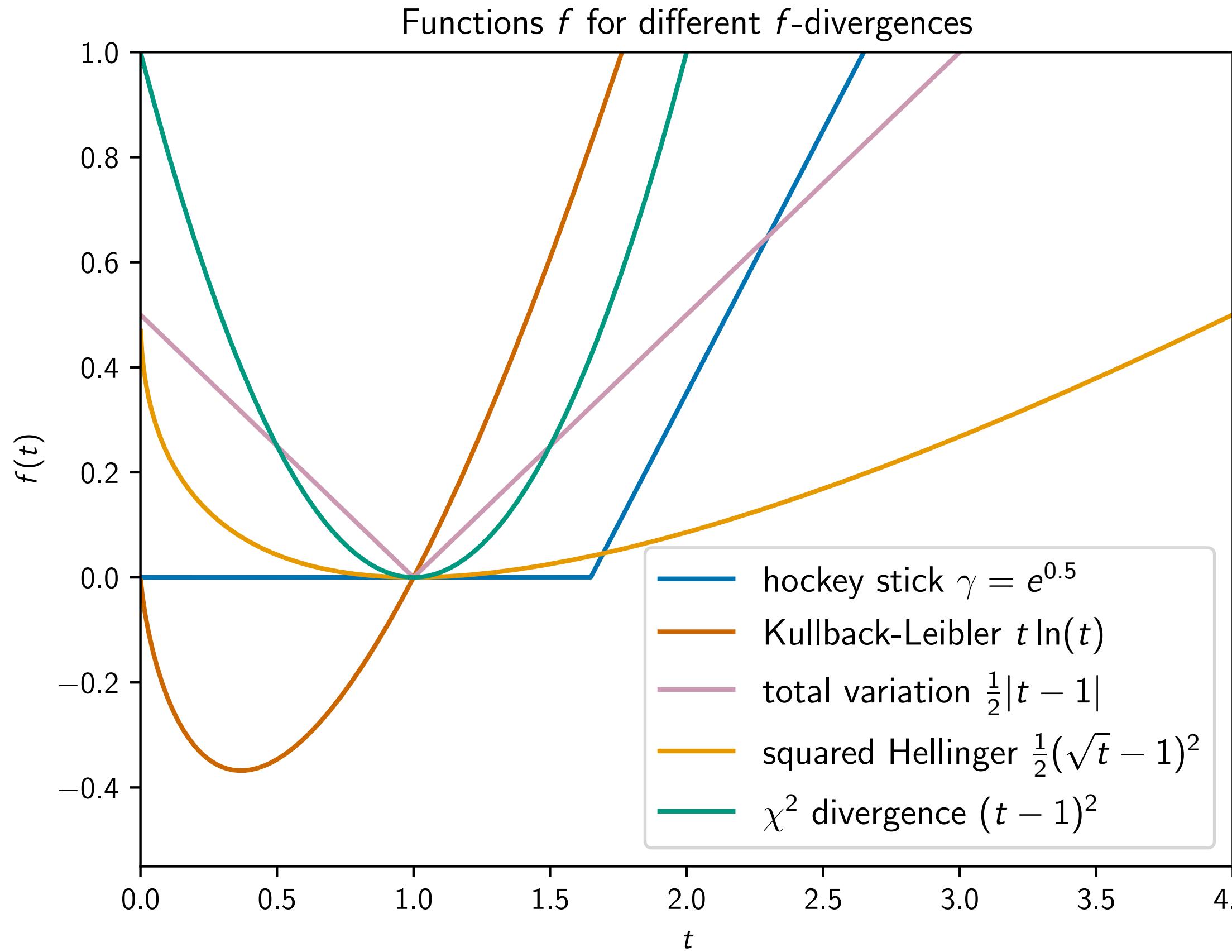
$$\mathbb{E}[L] = D_{\text{KL}}(P_{Y|x} \| P_{Y|x'})$$

The distribution of the PLRV is sometimes called the **privacy loss distribution** (PLD).

**A challenge:** this is defined for a single pair of inputs  $(x, x')$ . We would like to only deal with the “worst case” pair of inputs.

# Generalized divergences and the divergence

## How different are these two distributions?



More generally, look at  $f$ -divergences: for any convex  $f(\cdot)$ ,

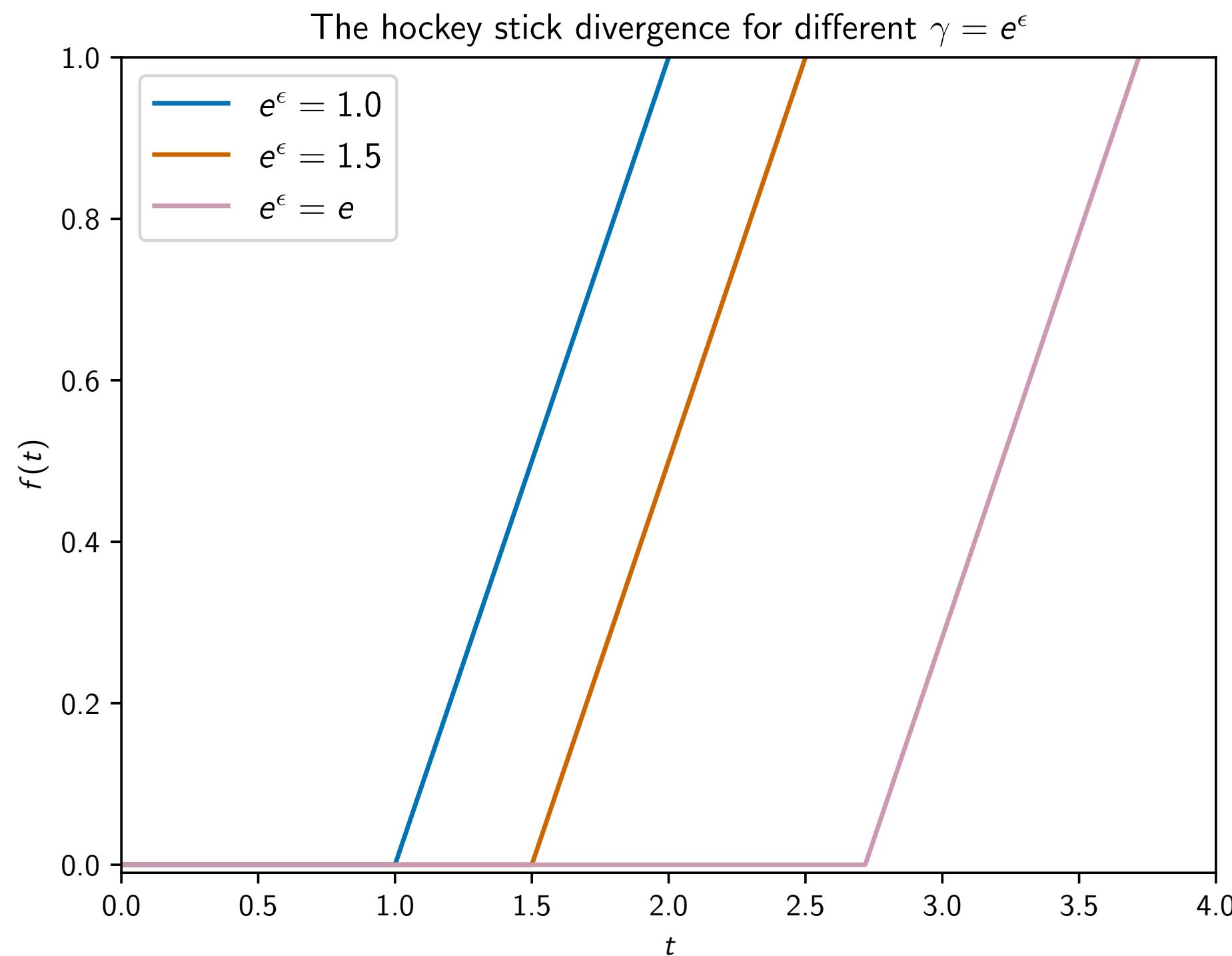
$$D_f(\mu\|\nu) = \mathbb{E}_\nu \left[ f\left(\frac{d\mu}{d\nu}\right)\right].$$

If we choose  $f(t) = (t - \gamma)^+$  we get the “hockey stick”, or  $E_\gamma$  divergence:

$$E_\gamma(\mu\|\nu) = \int_{\Omega} \left( \frac{d\mu}{d\nu} - \gamma \right)^+ d\nu = \sup_A [\mu(A) - \gamma\nu(A)].$$

# The PLRV and divergences

## Interpreting DP as an $f$ -divergence



For an  $(\epsilon, \delta)$ -DP mechanism  $P_{Y|X}$  we can take  $\gamma = e^\epsilon$  to get:

$$\delta_{P_{Y|X}}(\epsilon) = \sup_{x \sim x'} \mathbb{E}_{e^\epsilon} \left( P_{Y|X=x} \middle\| P_{Y|X=x'} \right) = \sup_{x \sim x'} \mathbb{E} \left[ (1 - e^{\epsilon - L})^+ \right]$$

Where  $L$  is the PLRV corresponding to  $(\mu, \nu)$ .

# Dominating pairs of distributions

## Interpreting DP as an $f$ -divergence

We would like to find a **tight dominating pair** of distributions  $(\mu, \nu)$  corresponding to a pair of inputs  $x$  and  $x'$ :

$$\sup_{x \sim x'} E_\gamma(P_{Y|X=x} \| P_{Y|X=x'}) = E_\gamma(\mu \| \nu).$$

However, this does not mean  $(\mu, \nu)$  corresponds to a **worst-case pair** of neighboring inputs. In fact, a worst-case pair may not exist!

- It's sufficient to look at  $\mu$  and  $\nu$  to be univariate distributions on  $[0,1]$ .
- We can use these dominating pairs to bound the loss for compositions.

# Composition and divergences

## Adding things up

If we have multiple releases with PLRVs  $L_1, L_2, \dots, L_k$ :

$$L = L_1 + L_2 + \dots + L_k.$$

Building from measure concentration (Dwork, Rothblum, Vadhan (2010)) or exact composition (Kairouz, Oh, Viswanath (2015), Murtagh, Vadhan (2016)) there are (at least) three main branches of work on composition:

- Large deviations
- Central limit theorem
- Numerical approaches

# A taxonomy of composition analyses

## (on non-interactive settings, also non-exhaustive)

### Large deviations

Bun and Steinke (2016)

Abadi, Chu, Goodfellow, McMahan,  
Mironov, Talwar, Zhang (2016)

Mironov (2017)

Wang, Balle, Kasiviswanathan (2018)

### CLT and characteristic functions

Dong, Roth, Su (2019)

Sommer, Meiser, Mohammadi (2020)

Cannone, Kamath, Steinke (2020)

Zhu, Dong, Wang (2022)

### Numerical methods

Koskela, Jälkö, Prediger, Honkela (2021/2021)

Gopi, Lee, Wutschitz (2021)

Ghazi, Kamath, Kumar, Manurangsi (2022)

Doroshenko, Ghazi, Kamath,  
Kumar, Manurangsi (2022)

### Warning about subsampling!

Lebeda, Regehr, Kamath, Steinke (2024)

Chua, Ghazi, Kamath, Kumar, Manurangsi, Sinha,  
Zhang (2024)

# Tilting a distribution

## Maintaining exactness for composition

Look at the cumulant generating function:

$$K_L(t) = \log \mathbb{E} [e^{tL}]$$

and the **“tilted” random variable** (for continuous  $L$ )

$$\tilde{L}_t \sim p_{\tilde{L}}(y) = e^{-K_L(t)+ty} p_L(y)$$

Then under some mild assumptions

$$\delta(\varepsilon) = \frac{1}{2\pi i} \int_{t-\infty i}^{t+\infty i} e^{K_L(z)-\varepsilon z - \log z - \log(1+z)} dz.$$

Can use this to derive a **“saddle-point” accountant** in terms of the exponent.

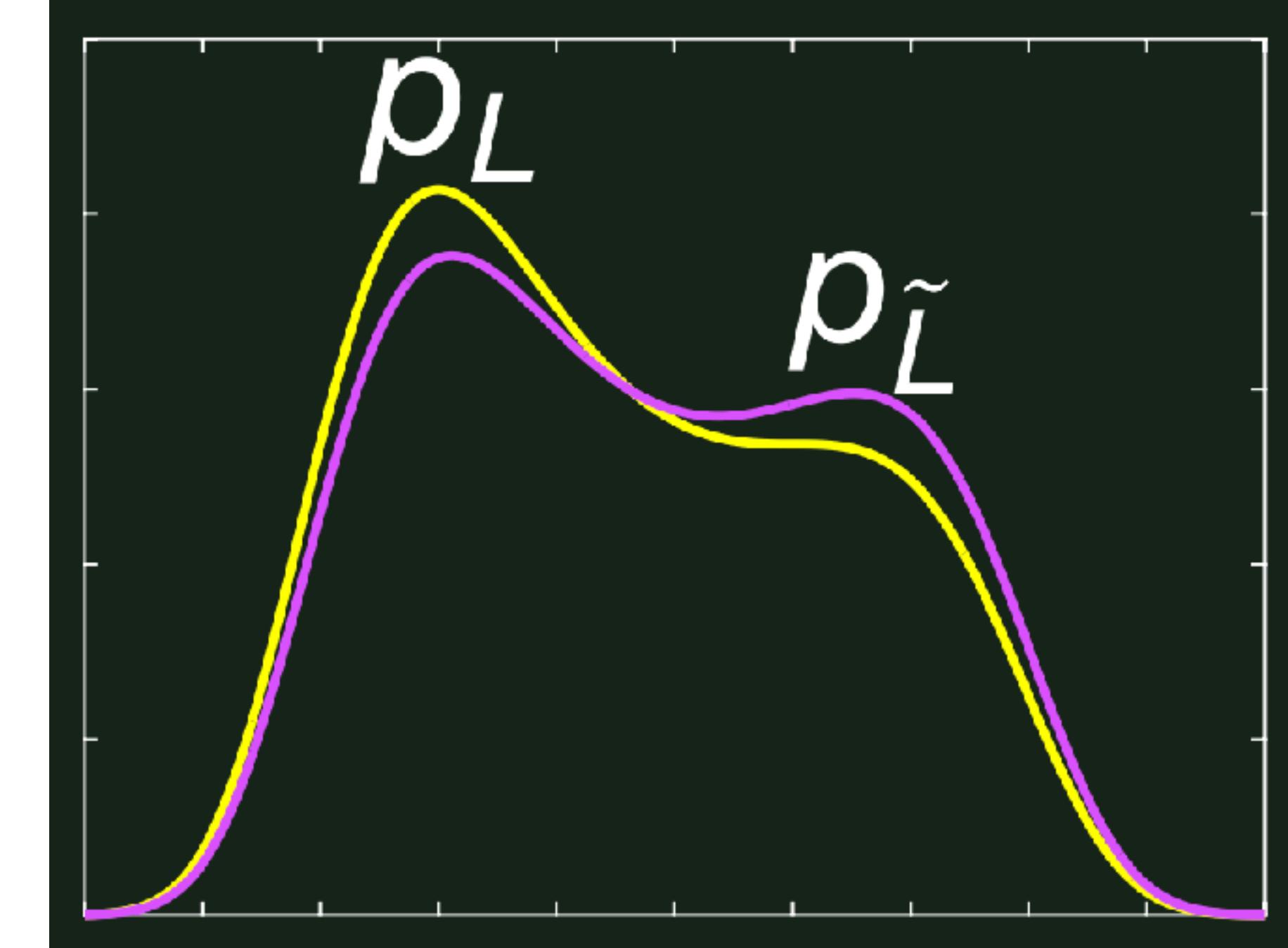
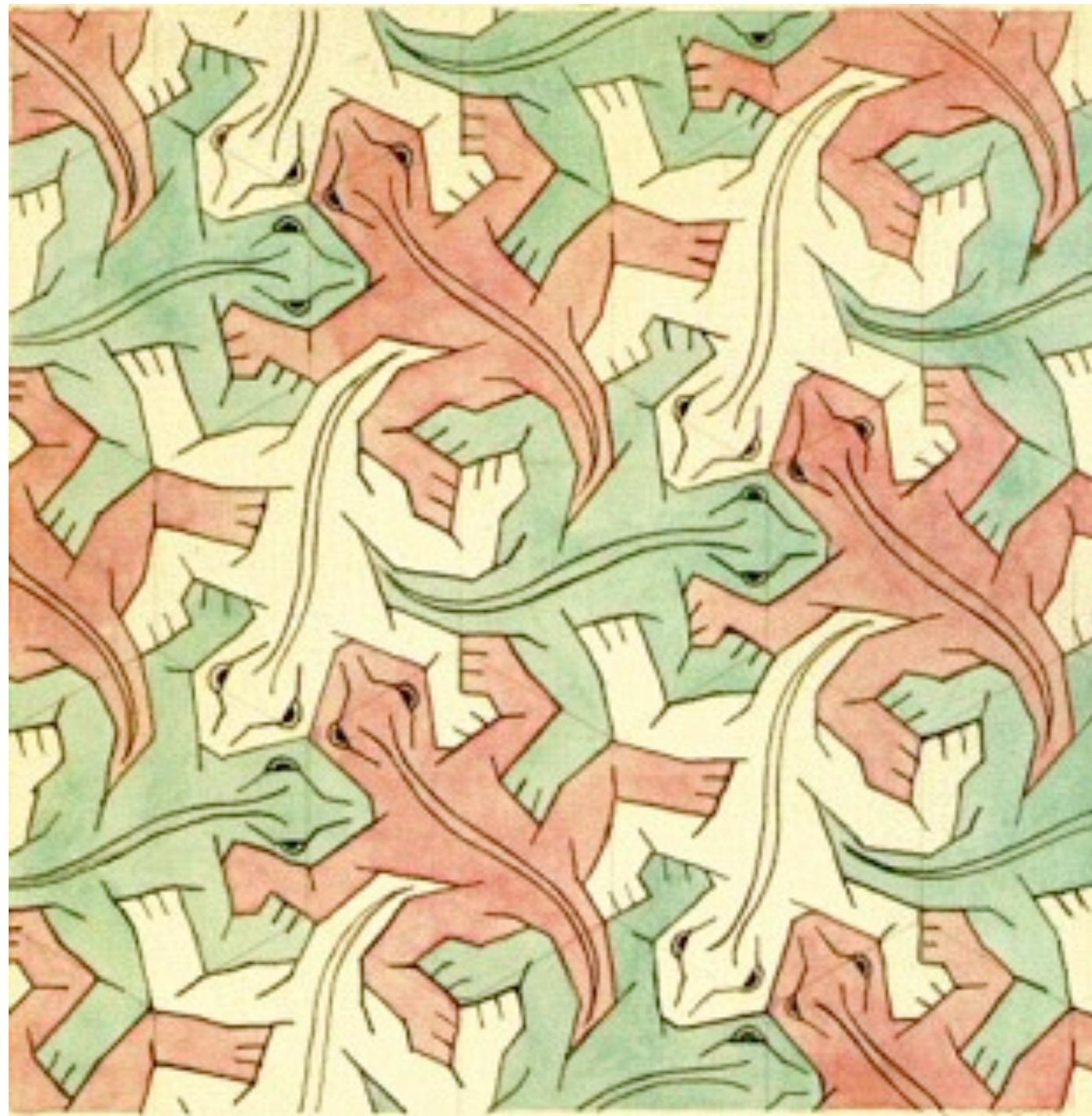


Figure: Oliver Kosut

# Tilting in other contexts

## (Beyond *Don Quixote*)



A tiling (not tilting) by  
M.C. Escher, not F. Esscher

Tilting is an old idea (originally in Esscher (1932)) and used in risk analysis, rejection sampling/importance sampling, and elsewhere. Connections to:

- Exponential families
- Edgeworth expansion
- Temperature in Boltzman/Gibbs distributions

Perhaps of interest to folks here? Botev (2017) uses it to exact iid simulation from the truncated multivariate normal distribution.

## Shichiri Beach in Sagami Province

相州七里浜

Soshū Shichiri-ga-hama



## Vista 4

contraction coefficients/iteration

# Maximum likelihood and ERM

## Optimization and privacy

Most ML/AI training (and classical stats) uses optimization to minimize some loss:

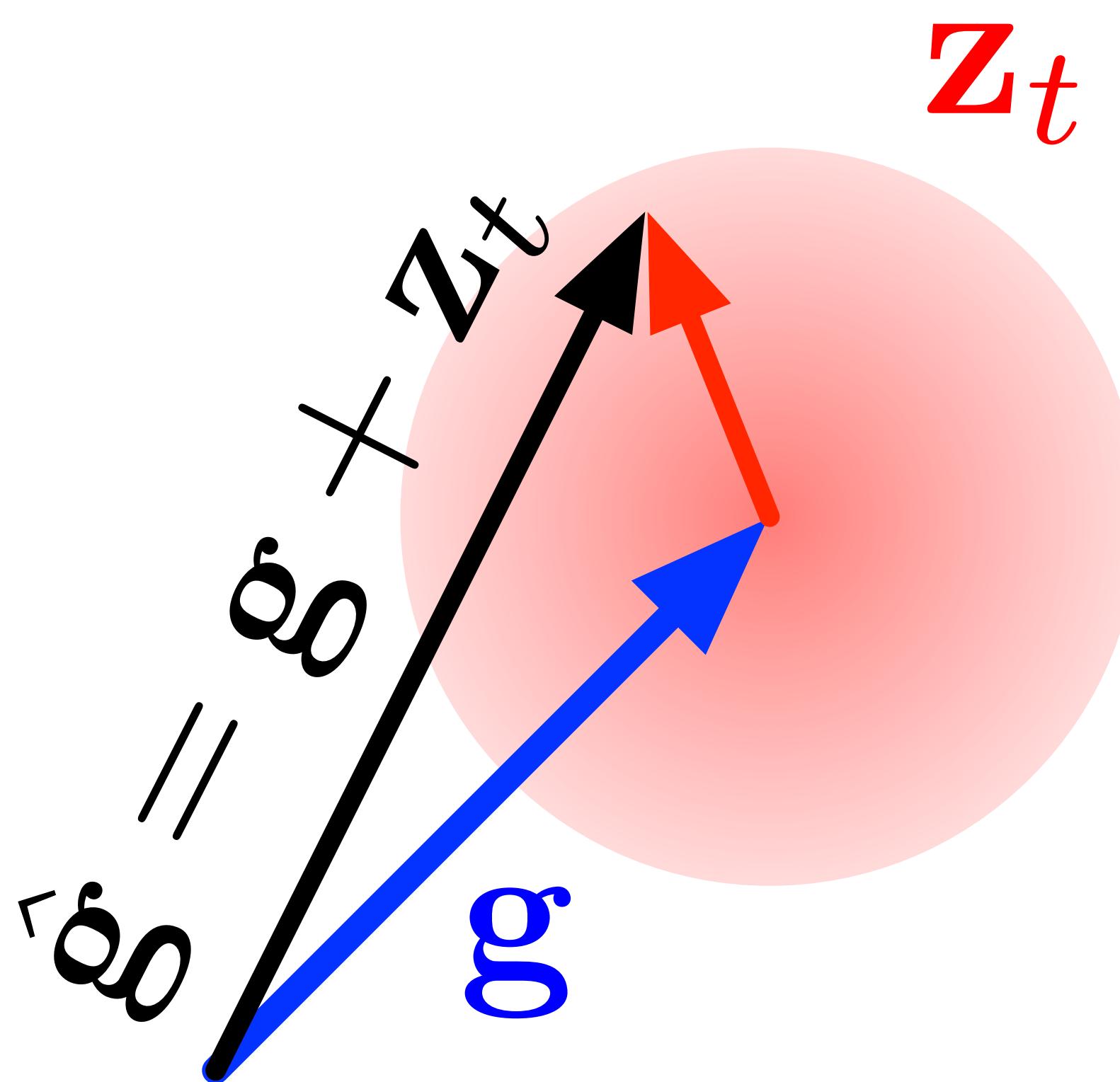
$$\mathbf{w}^* = \operatorname*{argmin}_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^n \ell(\mathbf{w}, \mathbf{x}_i).$$

We can use DP to approximate this in a number of ways:

- “**Output perturbation**”: compute the minimizer and add noise.
- “**Objective perturbation**”: Add a random term to the objective function and minimize it.
- “**Functional mechanism**”: Add noise to an approximation of the loss function  $\ell(\cdot)$ .

# Deep Learning and DP

## Privacy for neural networks



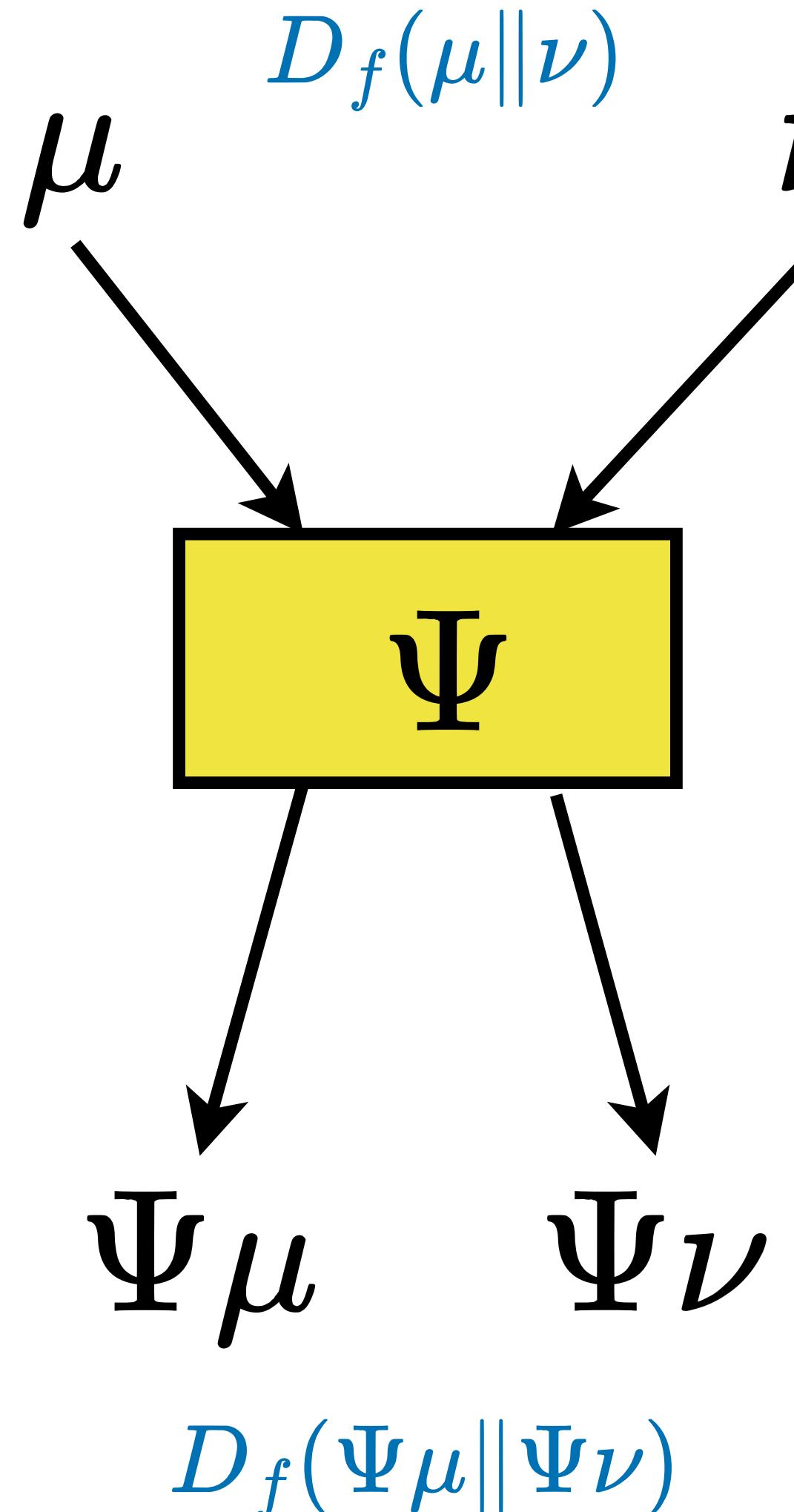
Deep neural networks (DNNs) use stochastic optimization algorithms in training: we can make stochastic gradient descent (SGD) differentially private by **adding noise to the gradients**.

- For high-dimensional problems, Gaussian noise is very effective.
- SGD has many iterations so we are potentially leaking a lot of information.

**Privacy accounting** lets us track the overall privacy loss.

# Strong data processing inequalities

## Quantifying the privacy gain from post-processing



The (Dobrushin) **contraction coefficient** of a channel  $\Psi$  for a divergence  $D_f$  is

$$\eta_f(\Psi) = \sup_{\mu, \nu: D_f(\mu \parallel \nu) \neq 0} \frac{D_f(\Psi\mu \parallel \Psi\nu)}{D_f(\mu \parallel \nu)}.$$

This quantifies the guaranteed gap (if it exists) in the **data processing inequality (DPI)**:

$$D_f(\Psi\mu \parallel \Psi\nu) \leq D_f(\mu \parallel \nu).$$

If  $\eta_f(\Psi) > 0$  this is a **strong data processing inequality (SDPI)**.

Dobrushin (1956), Ahlswede, Gács (1976)

# Contraction for the divergence

## Applications to DP-SGD and LDP

The **contraction coefficient** for the  $E_\gamma$  divergence admits a 2-point characterization:

$$\eta_f(\Psi) = \sup_{w,w'} E_\gamma(\Psi(w) \| \Psi(w')).$$

This is very similar to Dobrushin's characterization for total variation:

$$\eta_{\text{TV}}(\Psi) = \sup_{w,w'} \text{TV}(\Psi(w), \Psi(w')).$$

# Application to noisy SGD

Iterations just tack on more Markov kernels

In noisy SGD with iterates  $\{W_t\}$  compute clipped gradient updates  $g_t(W_{t-1})$ :

$$W_t \leftarrow \Pi_{\mathcal{W}}(g_t(W_{t-1}) + \sigma Z_t)$$

At each iteration, take  $\mu, \nu$  to be distributions of  $W_{t-1}$  under the two hypotheses and  $\Psi_t \mu, \Psi_t \nu$  to be distributions of  $W_t$ . We have two Markov chains:

$$\Psi_T \Psi_{T-1} \dots \Psi_1 \mu_0 \quad \Psi_T \Psi_{T-1} \dots \Psi_1 \nu_0$$

Idea: analyze privacy for the *last* iterate by using the contraction for the  $E_\gamma$  divergence.

# Privacy amplification by iteration

## An abbreviated timeline

If you hide the iterates, the privacy leakage converges (instead of increasing with the number of iterations).

- Feldman, Mironov, Talwar, Thakurta (2018) - convex, smooth
- Balle, Barthe, Gaboardi, Geumlek (2019) - strongly convex
- Chourasia, Ye, Shokri (2021/2022) Ryffel, Bach, Pointcheval (2022) - strongly convex, smooth for minibatch noisy SGD
- Altschuler, Talwar (2022)/Altschuler, Bok, Talwar (2024) - projected, convex, bounded
- Asoodeh, Diaz (2024) - use data processing inequalities to remove convexity and smoothness assumptions for projected DP-SGD and regularized DP-SGD.

# Contraction and Bayesian estimation

## Focusing on the local model

Suppose we have  $X_1^n$  i.i.d.  $\sim P_{X|\theta}$  with prior  $\theta \sim P_\Theta$  and privatized version  $Z_1^n$  with  $Z_i = \Psi_{\varepsilon,\delta}(X_i)$  (local DP). Then the **Bayes risk**

$$R(\Theta, \varepsilon, \delta) = \inf_{\Psi_{\varepsilon,\delta}} \inf_{\hat{\theta}} \mathbb{E}[\ell(\theta, \hat{\theta}(Y_1^n))]$$

can be lower bounded in terms of an  $E_\gamma$ -**mutual information**. In the language of **quantitative information flow**:

$\theta$  is a secret, the loss  $\ell$  is a negative gain, and we look for the maximally leaky channel subject to an  $(\varepsilon, \delta)$  constraint... (Is this right?)



**Morning After a Snowfall  
at Koishikawa**

礪川雪の旦

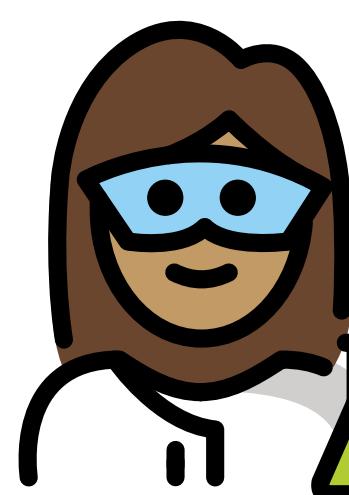
Koishikawa yuki no  
ashita

other destinations

# Where do we go from where we started?

## Wrapping up the end of the tour

Sara



$$s \in \{0, 1\}$$



$$Y \sim P_{Y|S=s}$$



$$\hat{s} \in \{0, 1\}$$



Benito

We started out with a simple story: **protecting a single bit**.

- Differential privacy both is and is not just as simple as **hypothesis testing**.
- Taking an **information-theoretic view** opens the door to better analyses.
- The **gap between algorithms and analysis** is shrinking.
- The **gap between algorithms and applications** is still large.

# The gap between theory and practice

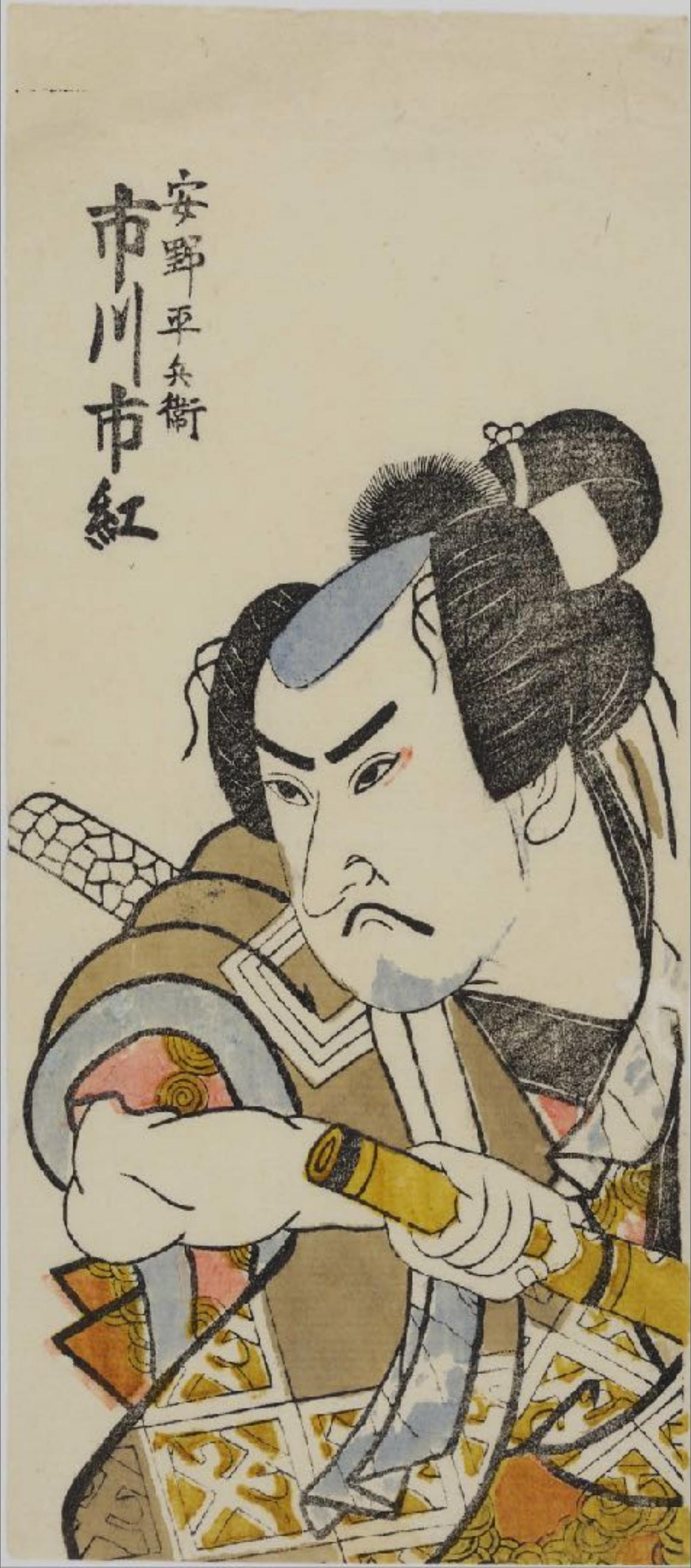
It's wider than you might think

There are lots of issues with implementing differential privacy in practice:

- Approximate versus exact sampling (and side channels)
- Approximate versus exact optimization
- “Privacy amplification” and its implementation
- Numerical precision and floating points
- Managing privacy budgets



Several interesting  
challenges left for:  
  
mathematics  
computational statistics  
engineering  
human-computer interaction  
technology policy





The Great Wave off  
Kanagawa

神奈川沖浪裏

Kanagawa oki nami ura

¡Gracias por su atención!