Privacidade Diferencial

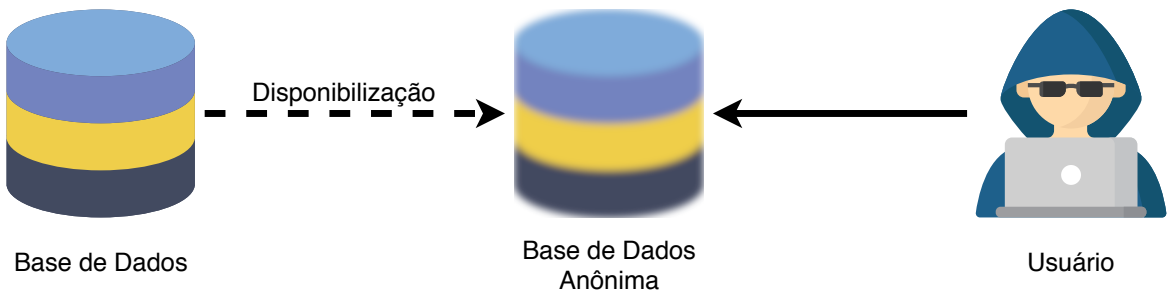
Prof. Javam Machado

LSBD/DC/UFC

Outubro/2019

Até então...

- Modelos de Privacidade Sintáticos
  - Dados são disponibilizados
  - Generalização e/ou supressão (de modo geral)
  - Atendem a uma condição sintática



Problemas com abordagens sintáticas

1. O acesso aos dados anonimizados deve necessariamente conter informações sobre registros individuais (para garantir a utilidade).
2. Operações de generalização/supressão tendem a levar a perda significativa de utilidade dos dados.
3. Um adversário pode ser capaz de descobrir a técnica de anonimização utilizada e os parâmetros adotados.

Privacidade segundo Tore Dalenius, 1977

'Tudo que se pode aprender sobre um indivíduo contido em uma base de dados deve ser possível aprender sem acesso à base de dados.' Estatístico Tore Dalenius, 1977

Vamos pensar um pouco...

"Tudo que se pode aprender sobre um indivíduo contido em uma base de dados deve ser possível aprender sem acesso à base de dados."

Em outras palavras:

- As crenças anteriores e posteriores do adversário sobre um indivíduo não devem ser muito diferentes.
- O acesso ao banco de dados estatístico não deve mudar a opinião do adversário sobre qualquer indivíduo.

Exemplo 1

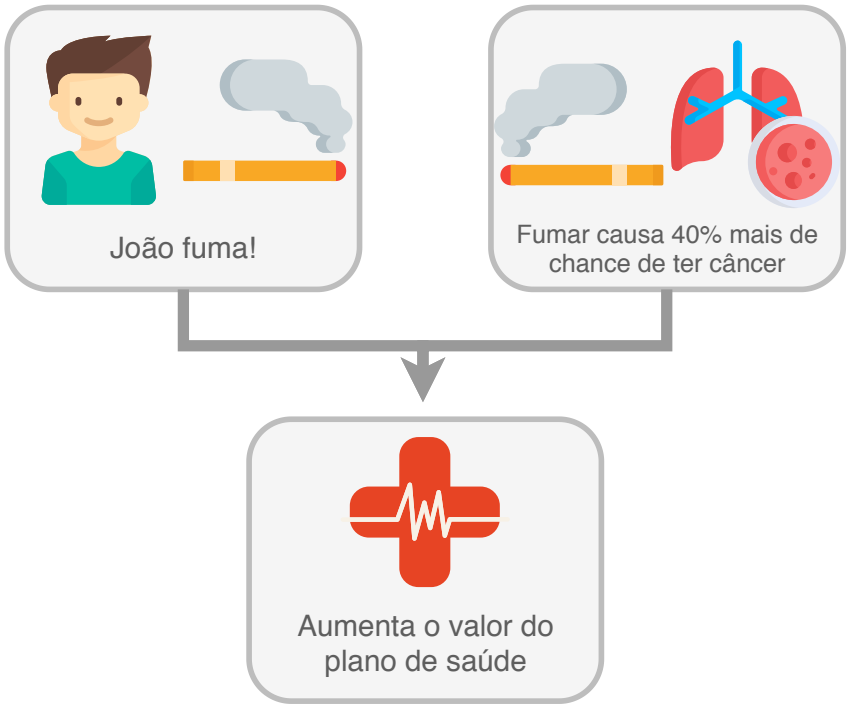
> *"Tudo que se pode aprender sobre um indivíduo contido em uma base de dados deve ser possível aprender sem acesso à base de dados."*

- Um adversário sabe que João é 10cm mais alto do que a média da população feminina do Brasil.
- Altura é um atributo sensível, neste caso.
- Tendo acesso a base de dados, o adversário pode computar a altura média da população feminina do Brasil e em seguida, computar a altura correta de João.

No exemplo acima, só foi possível aprender sobre a altura de João com o acesso a base de dados.

Exemplo 2

"O acesso ao banco de dados estatístico não deve mudar a opinião do adversário sobre qualquer indivíduo."



No exemplo acima, a opinião do plano de saúde de João mudou com o acesso a base de dados.

Deve-se então garantir que...

- As visões sobre um indivíduo antes e depois de ter acesso ao banco de dados não devem ser muito diferentes.
- O acesso ao banco de dados não deve mudar substancialmente o conhecimento que o adversário possui sobre um indivíduo.

Privacidade Diferencial

Privacidade Diferencial

É um **modelo matemático** e não uma condição sintática.

$$log\left(\frac{Pr(M(D_1) = O)}{Pr(M(D_2) = O)}\right) \leq \varepsilon$$

- A diferença entre as probabilidades de uma consulta retornar o mesmo resultado em dois conjuntos de dados é limitada pelo parâmetro ε.
- Qualquer elemento único do *dataset* deve ter apenas um impacto limitado no resultado de uma consulta.

Conjunto de dados vizinhos

- Pares de entradas que diferem em apenas uma tupla
  - Ausência ou presença
  - Valor

ID	Peso (Kg)	Altura (m)
1	87,2	1,7
2	81,2	1,62
3	74,2	1,75
4	60	1,61
5	78,5	1,58

Tabela: D_1

ID	Peso (Kg)	Altura (m)
1	87,2	1,7
2	81,2	1,62
4	60	1,61
5	78,5	1,58

Tabela: D_2

Privacidade Diferencial

- Disponibiliza, de maneira geral, informações estatísticas sobre conjuntos de dados



- Hospital
- Pesquisas médicas

- Censo
- Economistas

- Google
- Sistemas de recomendação

Privacidade Diferencial

- Disponibiliza, de maneira geral, informações estatísticas sobre conjuntos de dados



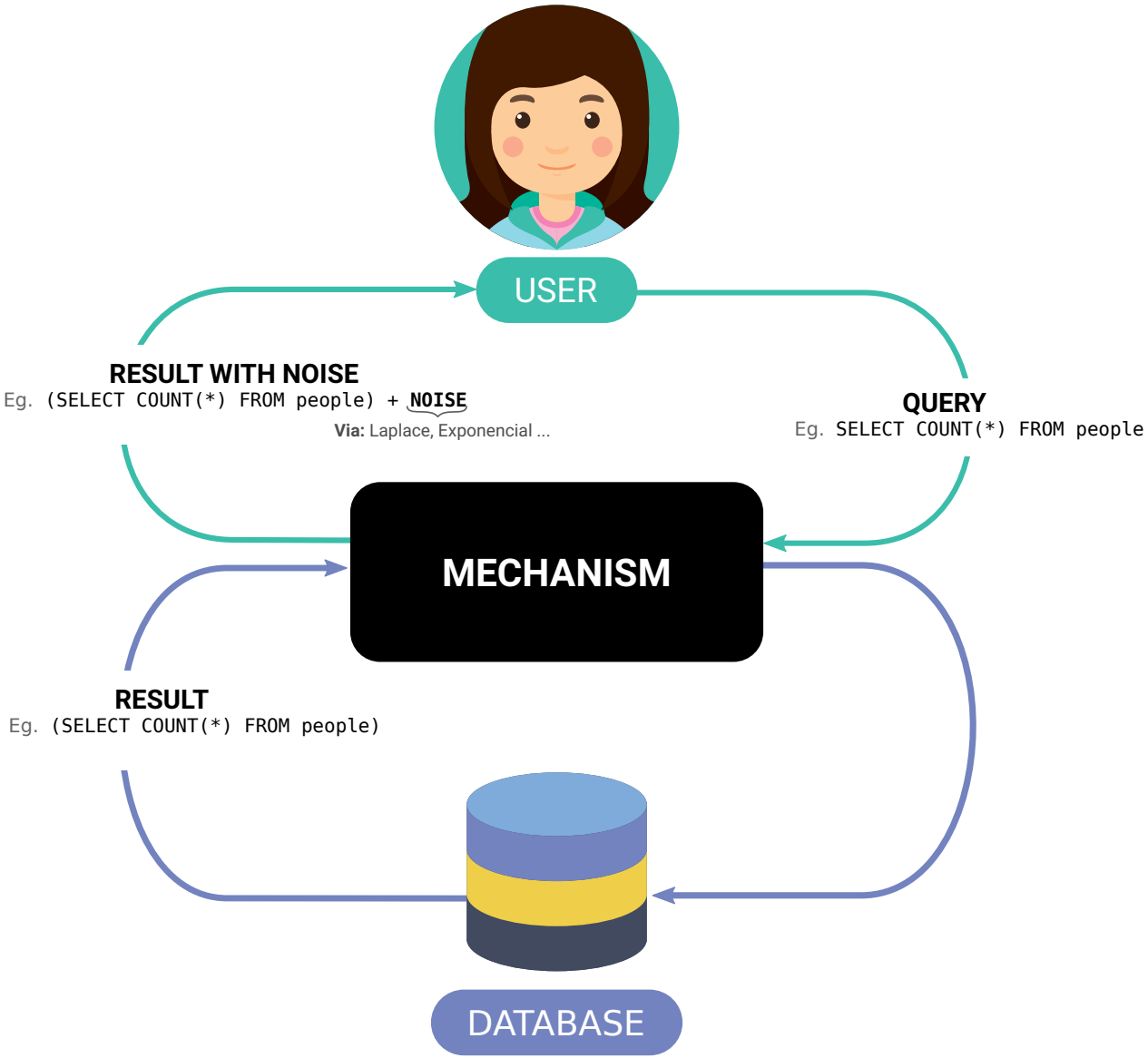
- Hospital
- Pesquisas médicas

- Censo
- Economistas

- Google
- Sistemas de recomendação

Informações estatísticas de consultas de agregação são suficientes para os processos de mineração acima.

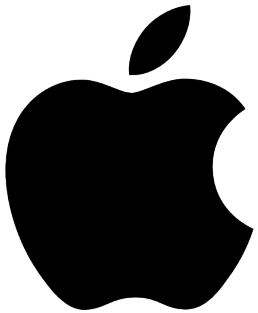
Privacidade Diferencial



Privacidade Diferencial – Em resumo

Um adversário não deve ser capaz de aprender nada sobre um indivíduo específico que ele já não poderia ter aprendido antes sem acesso ao conjunto de dados.

Quem utiliza Privacidade Diferencial?



Apples's Differential Privacy is about collection your data – but not your data

Privacy integrated queries: an extensible platform for privacy-preserving data analysis

Rappor: Randomized aggregatable privacy-preserving ordinal response

Mas nem todos utilizam da forma adequada

A privacidade diferencial da Apple pode não ser tão diferencial assim, como mostra este estudo

A teoria sobre privacidade diferencial afirma que o coeficiente épsilon considerado ideal — ou seja, que fornece ao receptor de dados a quantidade de informação suficiente para que a pesquisa seja relevante mas, ao mesmo tempo, permite que o usuário permaneça totalmente anônimo, sem chances de identificação posterior — fica em torno de 1. A pesquisa realizada nos sistemas da Apple, entretanto, mostrou uma realidade muito diferente: enquanto o macOS ficou com um coeficiente 6, por si só já considerado ruim, o iOS 10 apresentou coeficiente 14. Uma beta do iOS 11, por sua vez, conquistou um quase desprezível coeficiente 43 que, segundo os especialistas, significa que seus dados basicamente não são protegidos — mas eles próprios notam que esta é uma característica das versões de testes dos sistemas, que sempre são corrigidas antes do lançamento para o grande público.

Figura: Fonte: https://macmagazine.uol.com.br/post/2017/09/18/a-privacidade-diferencial-da-apple-pode-nao-ser-tao-diferencial-assim-como-mostra-este-estudo/

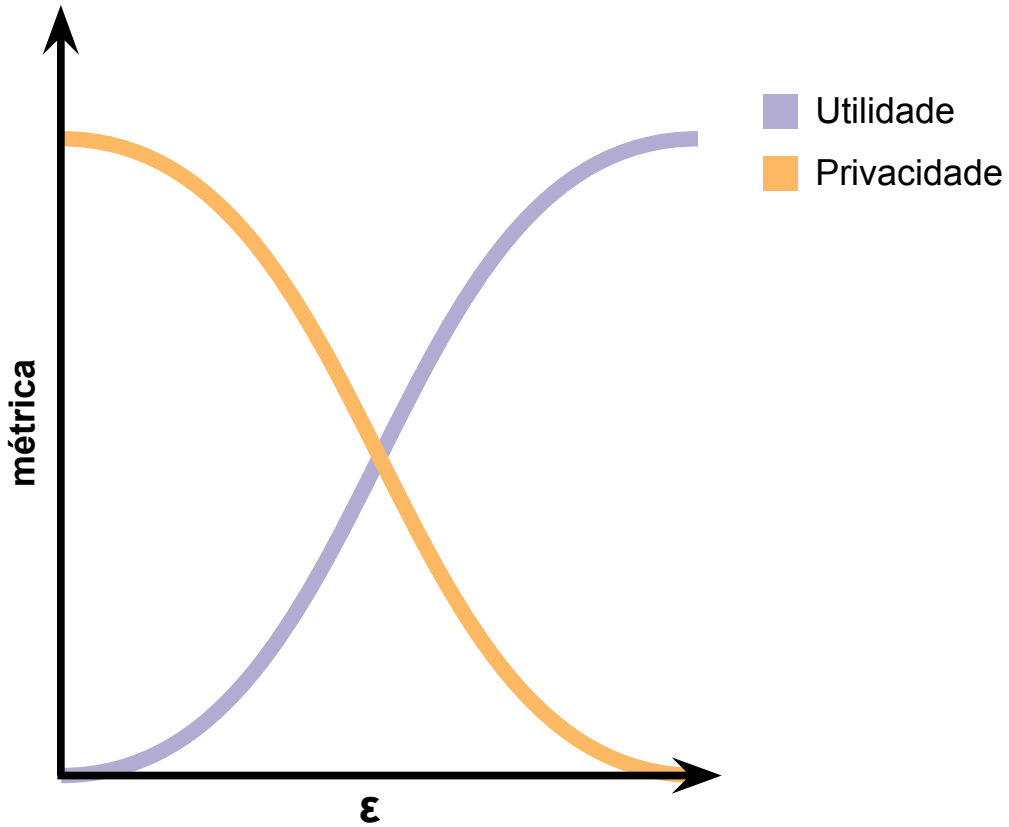
Limite de privacidade ε

Limite de privacidade ε

- Parâmetro: ε-privacidade diferencial
- $log\left(\frac{Pr(M(D)=O)}{Pr(M(D')=O)}\right) \leq \varepsilon$
- Controla o grau de indistinguibilidade
- Quanto menor o ε, menor deve ser a diferença de probabilidades
- Dependente da consulta
- Valores recomendados: 0.01, 0.1, $ln2$ e $ln3$

Limite de privacidade ε – *Trade-off*

- $log\left(\frac{Pr(M(D)=O)}{Pr(M(D')=O)}\right) \leq \varepsilon$
- $\Downarrow$ budget[1] $\varepsilon \implies \Uparrow$ privacidade

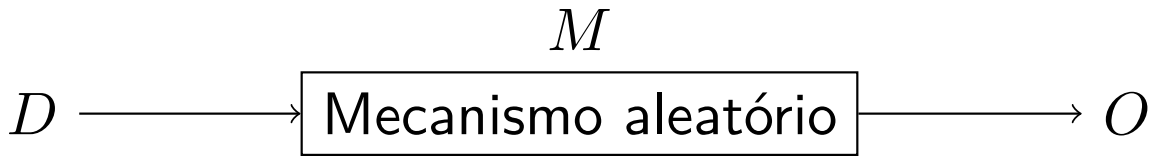


[1] Também chamado de limite de privacidade

Mecanismo

Mecanismo

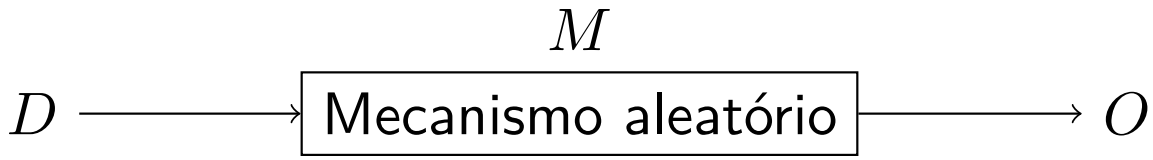
Ideia geral:



- Qualquer saída O de M é produzida com **quase** a mesma probabilidade, não importando se um indivíduo específico está na base de dados D.

Mecanismo

Ideia geral:

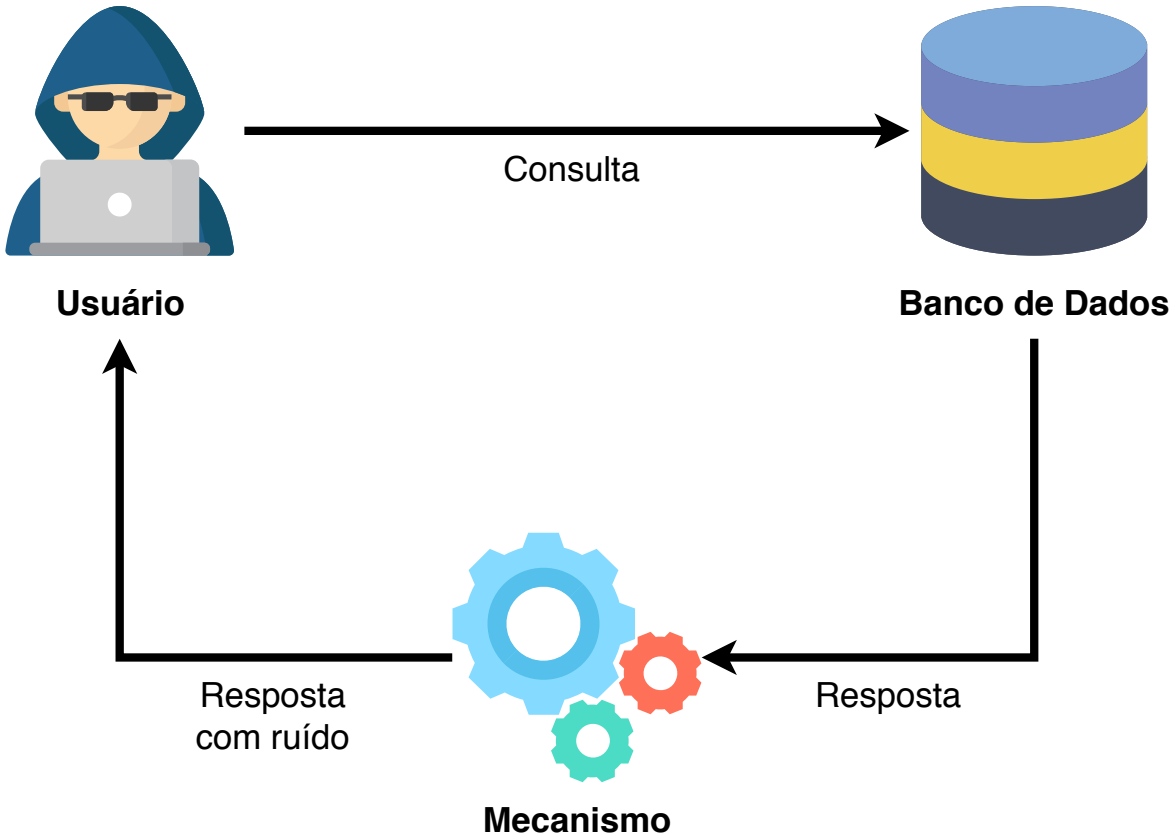


- Qualquer saída O de M é produzida com **quase** a mesma probabilidade, não importando se um indivíduo específico está na base de dados D.
- Um mecanismo M satisfaz ε-privacidade diferencial sse:

$$log\left(\frac{Pr(M(D) = O)}{Pr(M(D') = O)}\right) \leq \varepsilon$$

- Para quaisquer dois datasets D e D' vizinhos e todas possíveis saídas O

Mecanismo – *Overview* no fluxo da PD



- Consultas: count, sum, avg, min, max

Mecanismo

A definição:

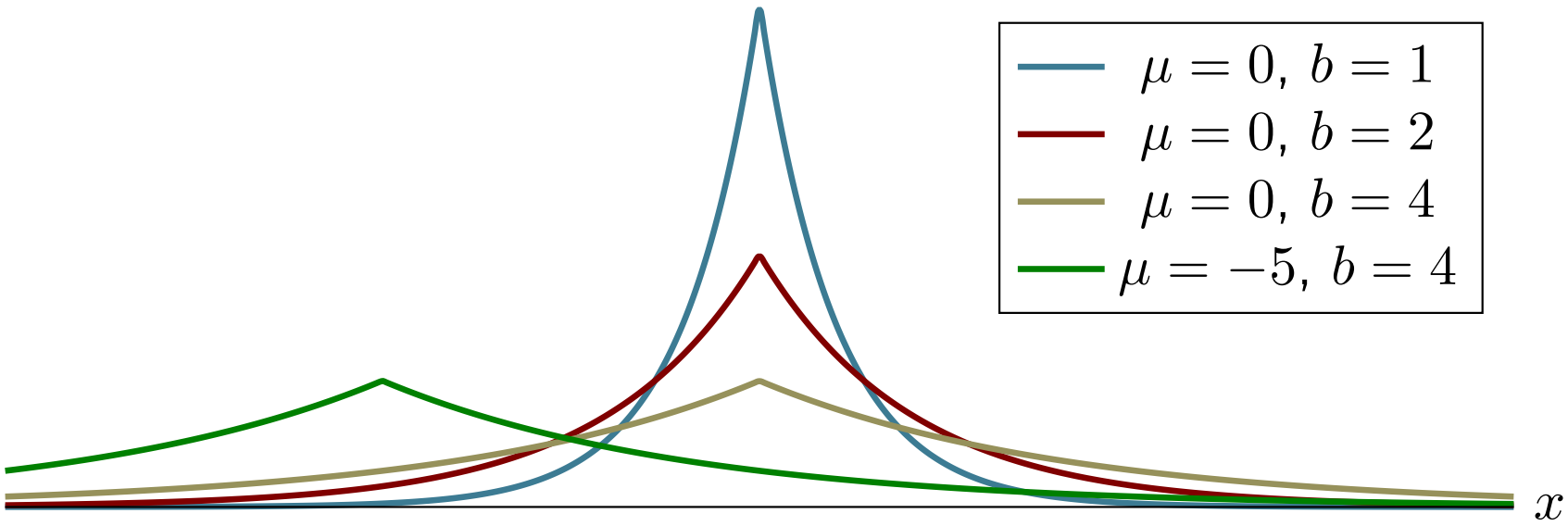
$$log\left(\frac{Pr(M(D) = O)}{Pr(M(D') = O)}\right) \leq \varepsilon$$

Também é comumente apresentada da seguinte forma:

$$Pr(M(D) = O) \leq exp(\varepsilon)Pr(M(D') = O)$$

A diferença entre as probabilidades de uma consulta retornar o mesmo resultado em dois conjuntos de dados é limitada pelo parâmetro ε.

Mecanismo – Laplace



μ é o parâmetro de localização e
$b > 0$ é o parâmetro de escala ou diversidade

Mecanismo – Laplace

- Utilizado para consultas numéricas
- Definições:

$$Lap(x|b,\mu) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)$$

$$M_L(x,f,\varepsilon) = f(x) + (Y_1,\ldots,Y_k)$$

- Onde Y_i são variáveis aleatórias i.i.d.[2] $Lap(\Delta f/\varepsilon, 0)$.
- Δf é a sensibilidade dos dados.

[2] Independente e identicamente distribuída

Laplace – Sensibilidade Δf

- Mede a maior diferença que um usuário faz ao ser removido do conjunto de dados na resposta da função de consulta
- Quanto maior o valor de Δf, mais ruído terá de ser adicionado à resposta do mecanismo para mascarar a remoção de um indivíduo

$$\Delta f = \max_{D, D' \in \mathbb{N}^{|\mathcal{X}|}} ||f(D) - f(D')||$$

- $\mathbb{N}^{|\mathcal{X}|}$ representa o universo de todos conjuntos possíveis de D.

Laplace – Prova

$$
\begin{aligned}
\frac{Pr(D = O)}{Pr(D' = O)} &= \frac{\frac{1}{2b}\exp\left(-\frac{|f(D)-O|}{b}\right)}{\frac{1}{2b}\exp\left(-\frac{|f(D')-O|}{b}\right)} \\
&= \exp\left(\frac{|f(D') - O| - |f(D) - O|}{b}\right) \\
&\leq \exp\left(\frac{|f(D') - f(D)|}{b}\right) \\
&\leq \exp(\varepsilon)
\end{aligned}
$$

Fonte: The Algorithmic Foundations of Differential Privacy; Cynthia Dwork Aaron Roth; 2014

Laplace – Exemplo

- Conjunto de dados da Receita Federal
  - Número de imóveis que cada indivíduo declarou
  - Consulta: **SUM**
  - Mecanismo de Laplace

ID	Nome	N. Imoveis
1	Zé	4
2	Sá	2
3	Gil	7
4	Lia	1

  - Resultado original: 14 imóveis

Laplace – Exemplo

ID	Nome	N. Imoveis
2	Sá	2
3	Gil	7
4	Lia	1

$$f(D_1) = 2 + 7 + 1 = 10$$

ID	Nome	N. Imoveis
1	Zé	4
3	Gil	7
4	Lia	1

$$f(D_2) = 4 + 7 + 1 = 12$$

ID	Nome	N. Imoveis
1	Zé	4
2	Sá	2
4	Lia	1

$$f(D_3) = 4 + 2 + 1 = 7$$

ID	Nome	N. Imoveis
1	Zé	4
2	Sá	2
3	Gil	7

$$f(D_4) = 4 + 2 + 7 = 13$$

Laplace – Exemplo

Sensibilidade local:

- $|14 - 7| = 7$

Parâmetro:

- $b = \frac{\Delta f}{\varepsilon} = \frac{7}{\varepsilon}$
- $\varepsilon = 1 \implies b = 7$
- $Lap(0, 7)$

Ruído	$f' = f(D) +$ ruído	$Pr(f')\%$
-4.58	9.42	3.70
-0.15	13.85	6.98
12.15	26.15	1.25
-6.43	7.57	2.85
2.89	16.89	4.72