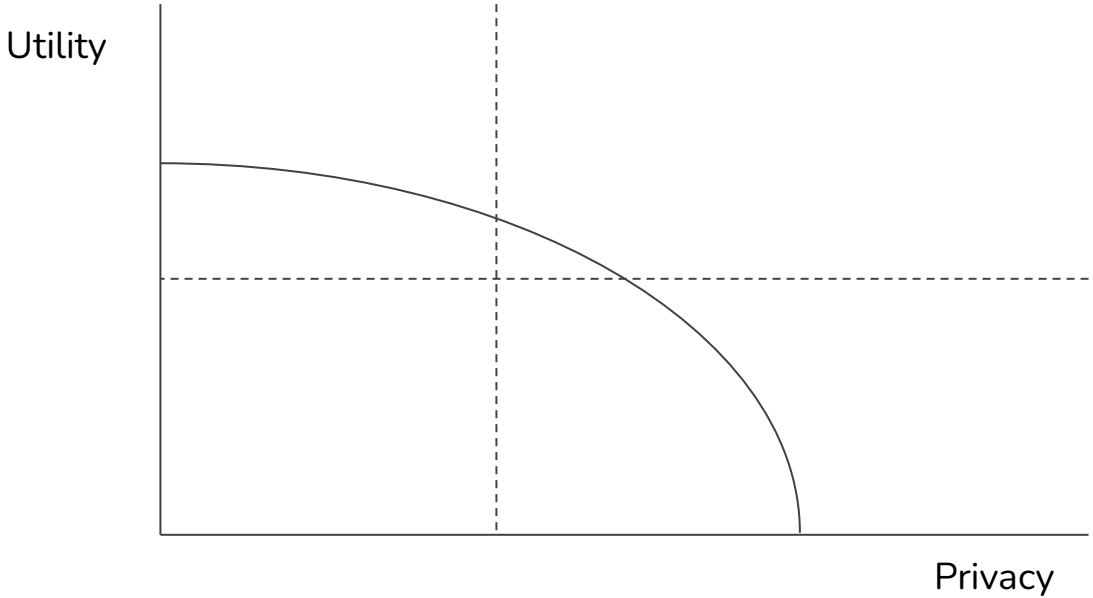
The effect of DP on (ads) decision making

PATCG: Oct 24, 2023

(Subset of) Goals for Private Measurement API

1. Prevent user level tracking
2. Enable aggregate measurement of cross-site behavior, e.g., aggregations based on joining *source events* and *trigger events*
3. Design an API that will be used, e.g., provides enough utility to be useful

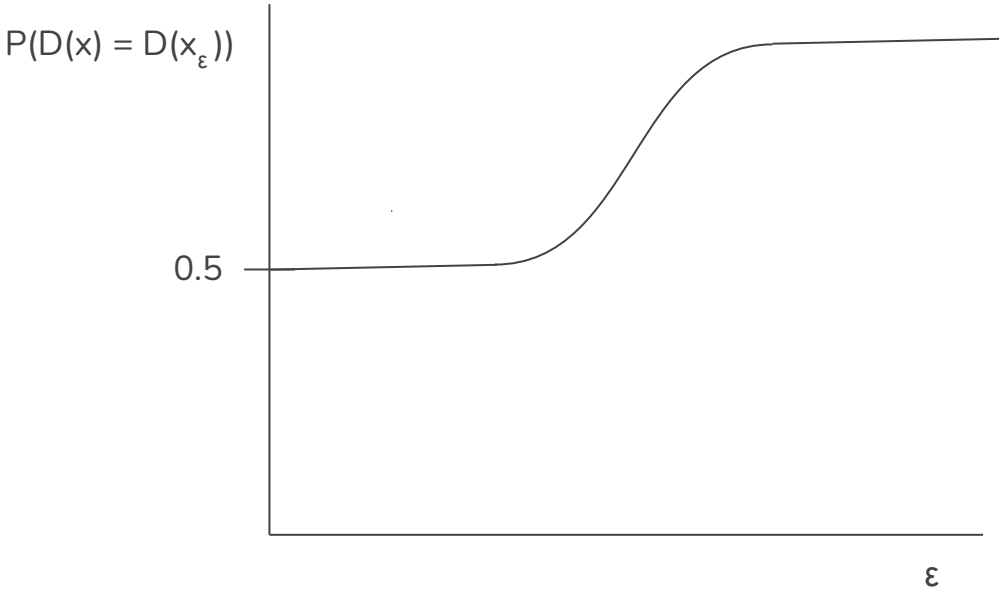
Privacy vs Utility



Goal for this project:
An experiment to estimate a lower bound for useful utility (lower bound on ε.)

Simple example

Suppose we have a binary decision, D(x), and suppose x_ε is a ε-differentially privatized version of x. As $\varepsilon \to 0$, we'd expect the $P(D(x) = D(x_\varepsilon)) \to 0.5$, e.g., random decisions.



Experimental Design

Goal: Observe the causal effect of adding ε-Differential Privacy to ads measurement results.

1. Simulate 10 campaign results (e.g., number of conversions)
2. Copy those results and add Laplace noise for ε
3. Shuffle 20 results
4. Ask binary question (increase or decrease spend) for each result
5. Evaluate how often D(conversions) = D($\text{conversions}_{\varepsilon}$)

A Simple Game

For each of these results, would you increase or decrease spend?

CONVERSIONS	INCREASE/DECREASE SPEND?	
11,715	Decrease	Increase
10,764	Decrease	Increase
5,566	Decrease	Increase
11,716	Decrease	Increase
6,273	Decrease	Increase

Configuration Parameters

1. Campaign Size: 10k - 100M
2. Conversion Rate: 0.5% - 99.5%
3. Induced Variance

Campaign Conversion Count Simulation

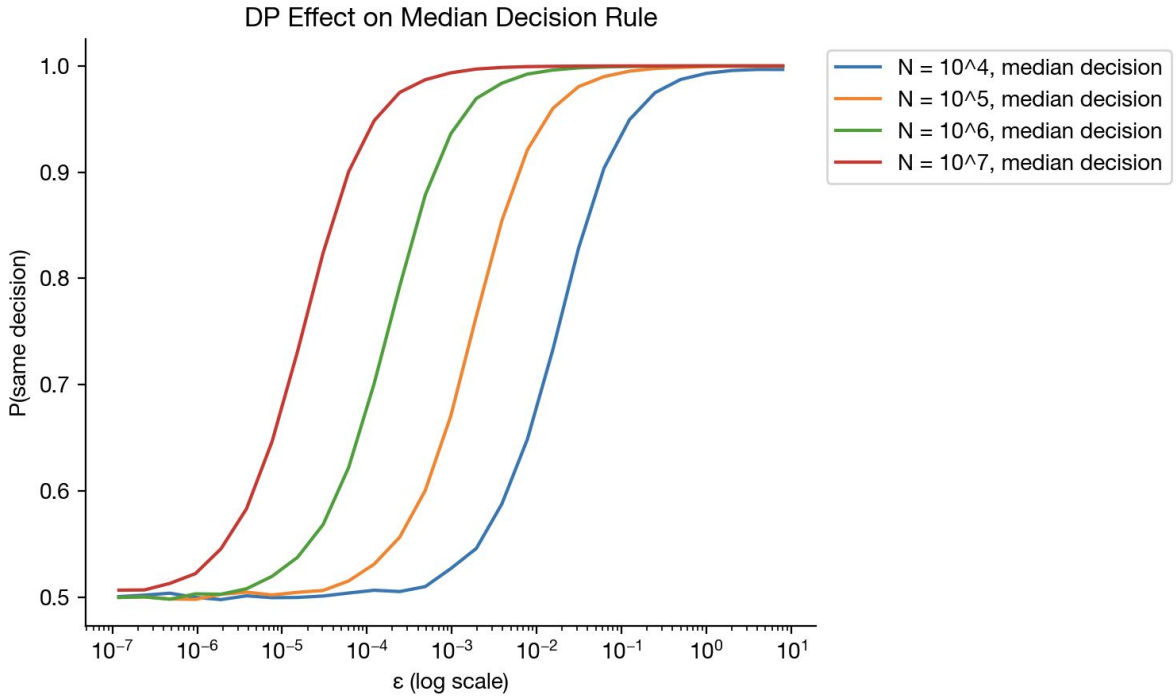
We model these conversion counts using a *beta-binomial process*.

1. Each campaign has an underlying conversion rate, drawn from a *Beta* distribution.
    a. Think of this like the weight of a coin, e.g., P(H) vs P(T).
2. Each campaign then has a conversion count, drawn from a *Binomial* distribution with the conversion rate from /1.
    a. Think of this like flipping a coin N times (for each impression.)

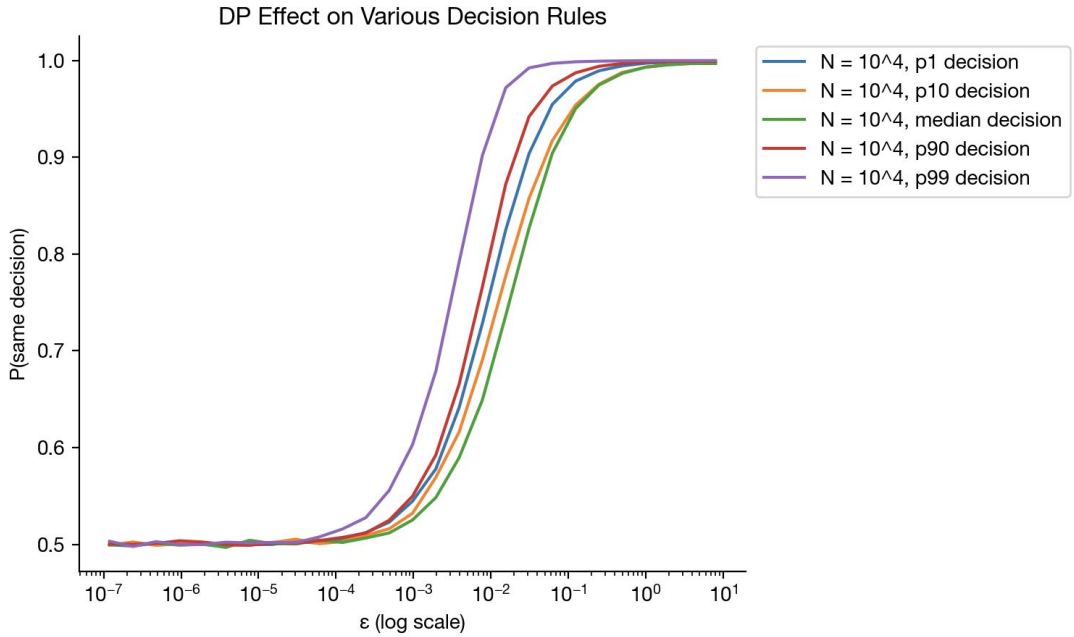
The variance of the *Beta* distribution is induced, by showing the user quantiles at p01, p10, p50, p90, and p99, and allowing them to increase or decrease.

We started with a distribution fit to one of the open source Criteo datasets, and then adjust according to user input. (For small conversion rates, we also make some more modifications.)

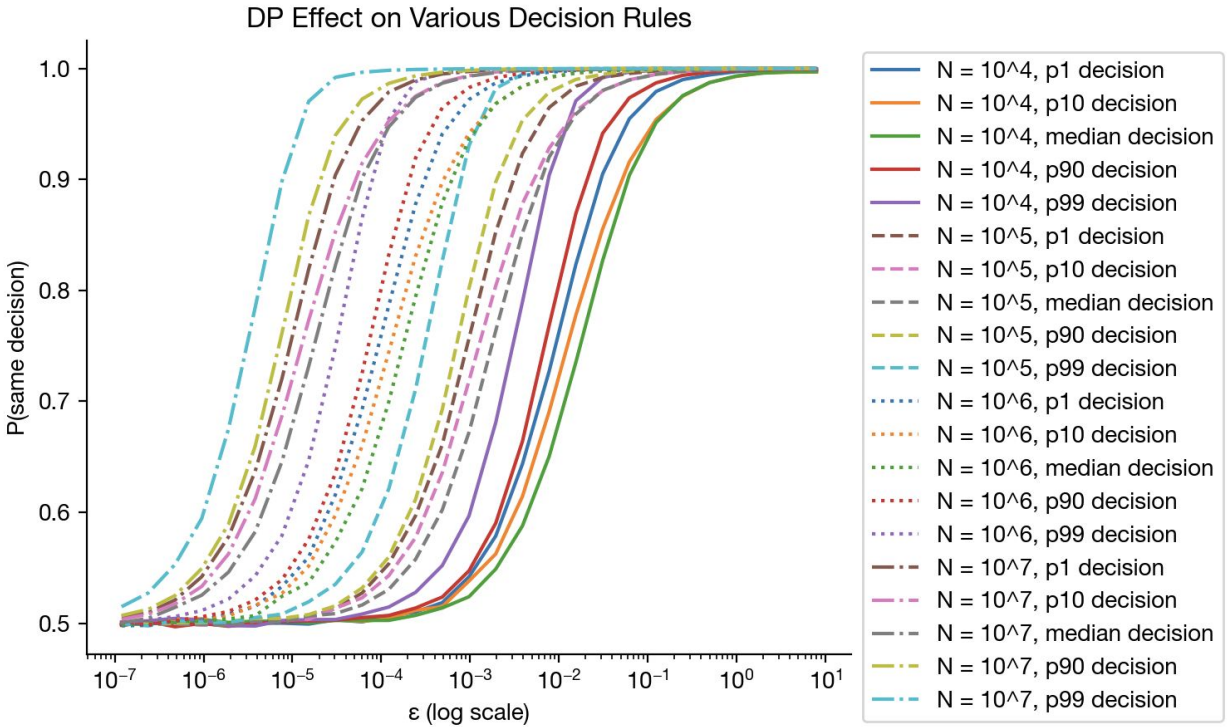
Too Simple?


DP Effect on Median Decision Rule
P(same decision)
1.0
0.9
0.8
0.7
0.6
0.5
10^{-7}
10^{-6}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
ε (log scale)
N = 10^4, median decision
N = 10^5, median decision
N = 10^6, median decision
N = 10^7, median decision

Difference across decision boundaries



Difference across decision boundaries



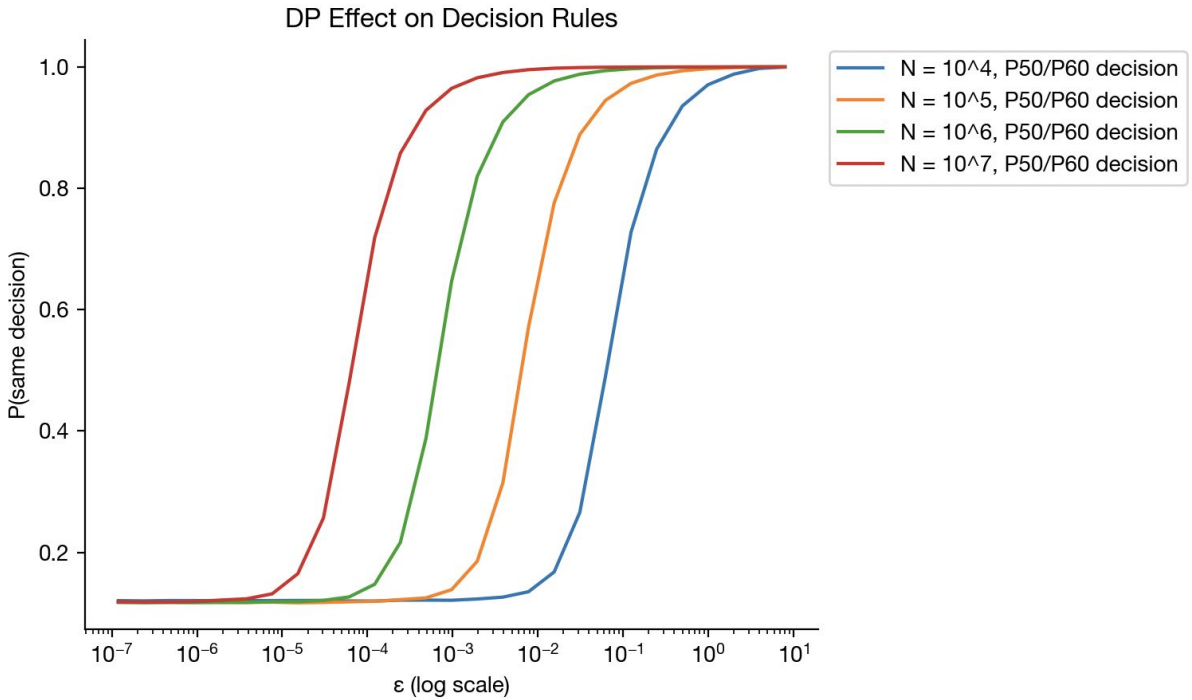
Ideas for extensions

Add "maintain" or "not enough data" #24

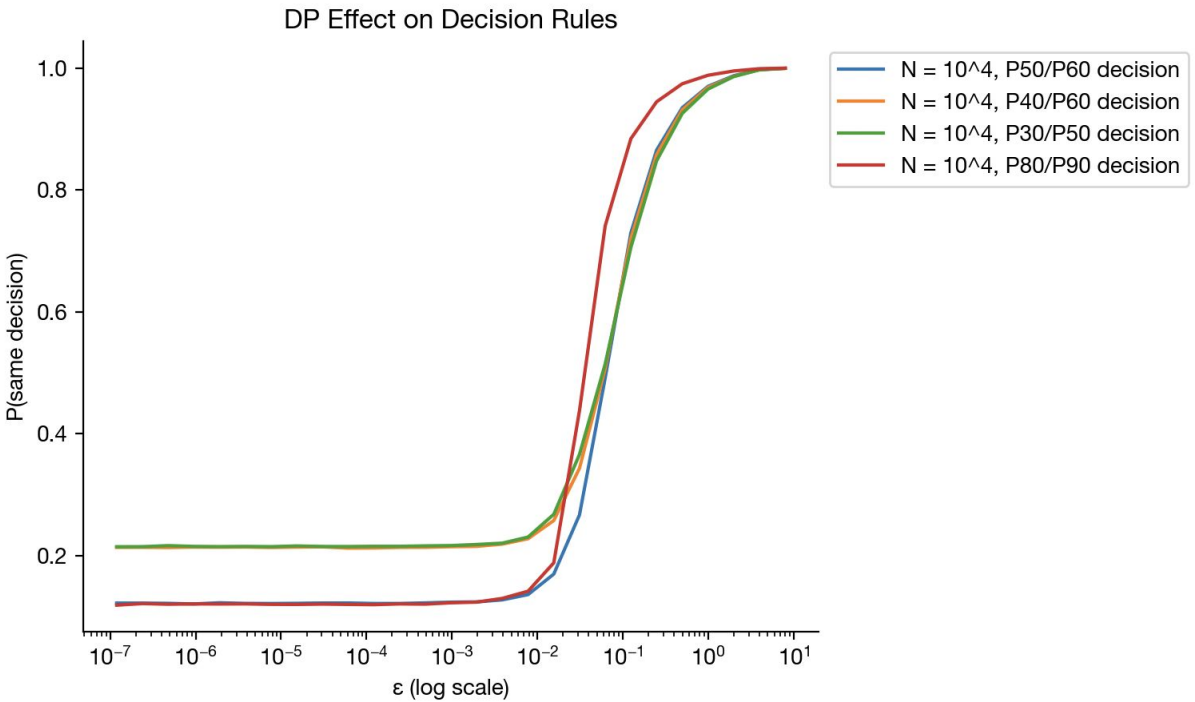
Open **alexWhitworth** opened this issue 5 days ago · 2 comments

- Add a new option “maintain”
- Separate out non-noised and noised results into two pages
- Add a *confidence interval** around noised values
  - *CI would only be with respect to DP noise
- Open question: How to measure accuracy?
  - Naive solution: Keep it the same, i.e., only count (increase, increase), (decrease, decrease), and (maintain, maintain). This is a bit weird as both (increase, maintain) and (increase, decrease) are both “wrong” but it seems like the latter should be more wrong.

This too can be simulated…



But maybe that's OK



Another Idea for extensions

1. Ask user for more configuration
    a. Number of campaign subcategories to be measured (e.g., different creatives, different audiences, etc.)
    b. Length of campaign and frequency of measurement (e.g., 1 week / daily)
2. Make multiple decisions at each measurement
    a. Split spend across length of campaign
    b. Make adjustments in spend decisions for each measurement (e.g., each day)
3. Open Questions
    a. This would provide an estimate of the delta between total conversions in each scenario.
    b. This is a bit different than finding the point where decisions are random, and less clear at which point this delta is unacceptable.

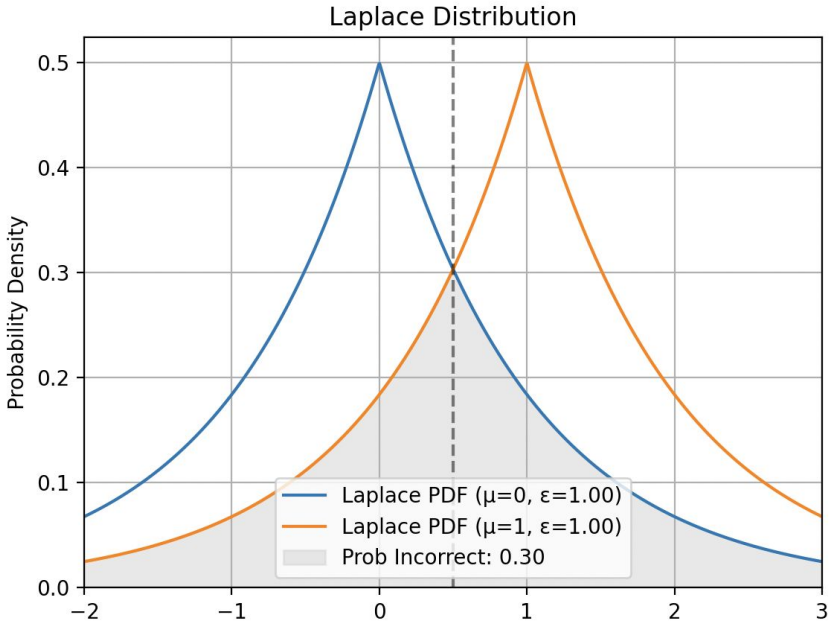
Running the Experiment

Our hope is to recruit people to “play” this game, collect the results, and analyze the results to share with the group.

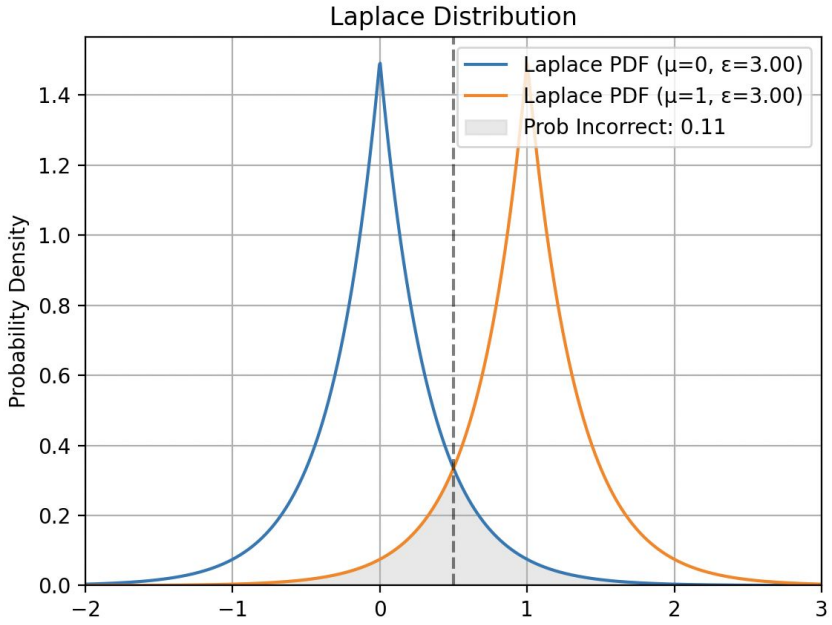
1. Who should we recruit to play the game?
2. How many should we recruit? (How many can we practically recruit?)
    a. Variance across people is currently unknown, so can’t do power calculation.
3. Would anyone want to help recruit participants?

Extras

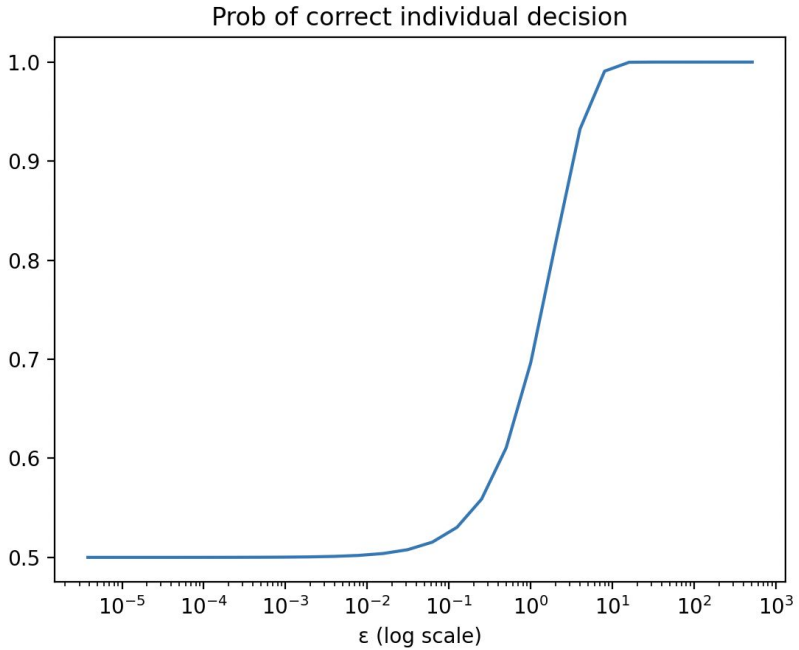
Effect of Noise on Individual Event



Effect of Noise on Individual Event



Effect of Noise on Individual Event



Effect of Noise on Individual Event