Randomised response

Adding noise to the mix

Imagine being asked:

- Have you ever cheated on your spouse?
- Have you ever committed a crime?

Answering truthfully, even if guaranteed about the data being later anonymised might be threatening to the respondent:

- Extrinsically, e.g., if certain responses carry the risk of sanctions (illegal behaviour)
- Intrinsically, if the questions are personal, stressful, or make the respondent "look bad"

Imperial College London

Adding noise to the mix

Each respondent is asked a sensitive Yes/No question, but their answer is randomly perturbed (Lensvelt-Mulders et al., 2005):

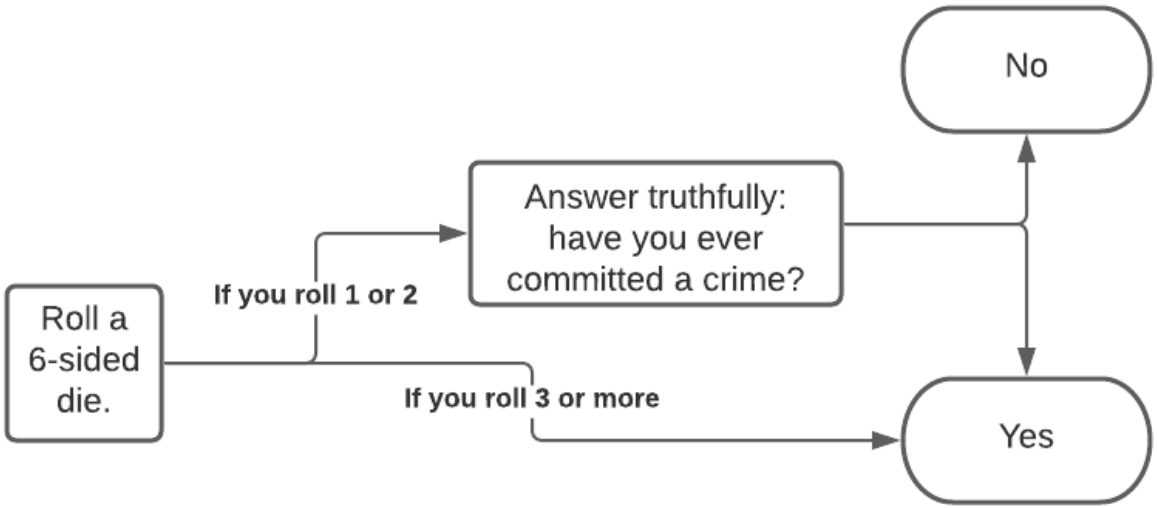

Figure 7: Randomized response trades off privacy for precision.

Would this protocol protect the user if we flip Yes and No?

Adding noise vs other techniques

How does this compare to redaction / coarsening / aggregation?

- Every patient record is simultaneously protected
- Can recover true aggregates with pre-specified accuracy
- Can resemble coarsening if the noise is local (say, Gaussian)
- At the extreme, randomisation approaches redaction

In this course, we are big fans of randomisation.

responder.id	gender	age	cheated	cheated.rr
pid 1	0	24	1	1
pid 2	0	27	1	1
pid 3	0	51	0	1
pid 4	0	41	1	1
pid 5	0	30	0	1
pid 6	0	48	0	0

Table 4: Randomisation can also preserve group aggregates, e.g., percentage of women aged under 40 that have cheated on their spouse.

A real-world deployment: Google and Apple

Your phone reports information back continuously to help improve services, some of which is hard to obfuscate, e.g.,

- Which websites consume more energy than average?
- Your location

URLs and locations are harder but still possible to randomise, e.g., Google's RAPPOR (Erlingsson, Pihur, and Korolova, 2014).

RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response

Úlfar Erlingsson
Google, Inc.
ulfar@google.com

Vasyl Pihur
Google, Inc.
vpihur@google.com

Aleksandra Korolova
University of Southern California
korolova@usc.edu