CS208: Applied Privacy for Data Science
Membership Attacks

School of Engineering & Applied Sciences
Harvard University

February 3, 2022

Public Access to Genome-Wide Data: Five Views…

Perusall Comments:

- What is a DNA mixture?
- Paywall?/exclusion to access
- Privacy of means vs. medians and other aggregates
- Researchers are human (“sloppy”) and human factors are real vulnerability
- “I would hope the next 12 months would produce greater clarity and time to produce a proportionate long-term response.” **12 months?**
- Fear of scientific abuse, particularly with genetics

Hypothesis Tests

A null hypothesis is a conjectured model of the world with observable implications.

Often it is a simplified model, for which there is some informational value if it can be refuted.

Null Distributions

If t is a function of the data, it has a sampling distribution. The distribution that t would obtain if the null hypothesis were true is called the *null distribution*.

- If we use the value of t to draw an inference about the null hypothesis, we call t a **test statistic**.
- We observe t^* in some observed dataset $\mathbf{X}^*$ and reason whether it could have been a draw from the null distribution.
- If t^* is unlikely to have come from the null distribution, we **reject the null** hypothesis.
- If t^* could have been obtained from the null distribution, we **fail to reject the null**.
- Failing to reject the null, does not prove the null to be true.

Inferential Errors

Reasoning from known data to about an unknown hypothesis is called inference. Inferential errors are commonly labelled by type:

	Null True	Null False
Fail to Reject Null		
Reject Null		

Inferential Errors

Reasoning from known data to about an unknown hypothesis is called inference. Inferential errors are commonly labelled by type:

	Null True	Null False
Fail to Reject Null	Correct	
Reject Null		Correct

Inferential Errors

Reasoning from known data to about an unknown hypothesis is called inference. Inferential errors are commonly labelled by type:

	Null True	Null False
Fail to Reject Null	Correct	Error (Type II)
Reject Null	Error (Type I)	Correct

Inferential Errors

Reasoning from known data to about an unknown hypothesis is called inference. Inferential errors are commonly labelled by type:

	Null True	Null False
Fail to Reject Null	Correct	Error (Type II) Sensitivity
Reject Null	Error (Type I)	Correct Sensitivity

Inferential Errors

Reasoning from known data to about an unknown hypothesis is called inference. Inferential errors are commonly labelled by type:

	Null True	Null False
Fail to Reject Null	Correct Specificity	Error (Type II) Sensitivity
Reject Null	Error (Type I) Specificity	Correct Sensitivity

We parameterize our hypothesis test by choice of δ which results in a **critical value**, c, which divides the null distribution into the rejection regions.

Inferential Errors

Reasoning from known data to about an unknown hypothesis is called inference. Inferential errors are commonly labelled by type:

	Null True	Null False
Fail to Reject Null	Correct Specificity $1 - \delta$	Error (Type II) Sensitivity
Reject Null	Error (Type I) Specificity δ	Correct Sensitivity

Example

H_0 : K-dimensional random variables $\mathbf{x}$ and $\mathbf{z}$ are both drawn from a standard Normal distribution with the same mean, $\mathcal{N}(\vec{\mu}, 1)$.

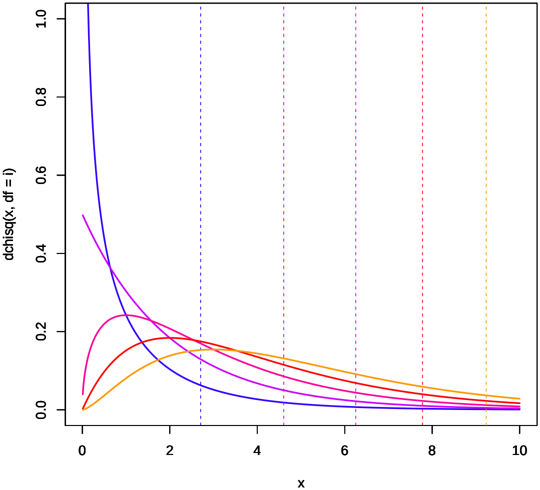
Then one test statistic is:

$$t(\mathbf{x}, \mathbf{z}) = ||\mathbf{x} - \mathbf{z}||_2 = \sqrt{\sum_{i=1}^{K} (x_i - z_i)^2}$$

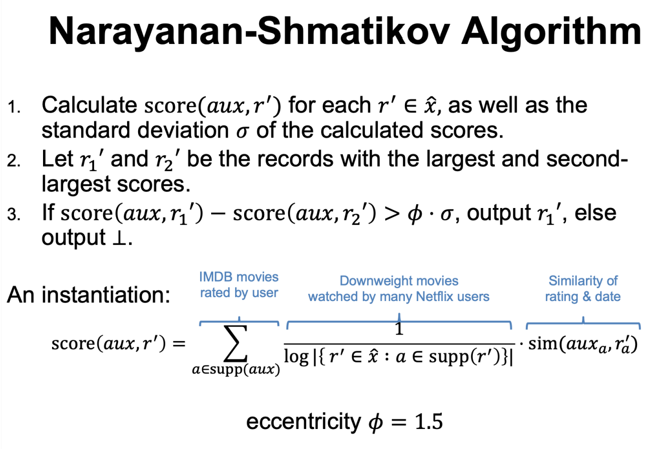
Which has null distribution $\chi^2(K)$.

Example

$\chi^2(K)$ Distribution with critical values for $\delta = 0.1$, for i in 1 to 5:



Netflix Challenge (from last week)



Narayanan-Shmatikov Algorithm

1. Calculate $\mathrm{score}(aux, r')$ for each $r' \in \hat{x}$, as well as the standard deviation σ of the calculated scores.
2. Let $r_1{}'$ and $r_2{}'$ be the records with the largest and second-largest scores.
3. If $\mathrm{score}(aux, r_1{}') - \mathrm{score}(aux, r_2{}') > \phi \cdot \sigma$, output $r_1{}'$, else output $\perp$.

An instantiation:

$$\mathrm{score}(aux, r') = \overbrace{\sum_{a \in \mathrm{supp}(aux)}}^{\text{IMDB movies rated by user}} \overbrace{\frac{1}{\log|\{r' \in \hat{x} : a \in \mathrm{supp}(r')\}|}}^{\text{Downweight movies watched by many Netflix users}} \cdot \overbrace{\mathrm{sim}(aux_a, r'_a)}^{\text{Similarity of rating \& date}}$$

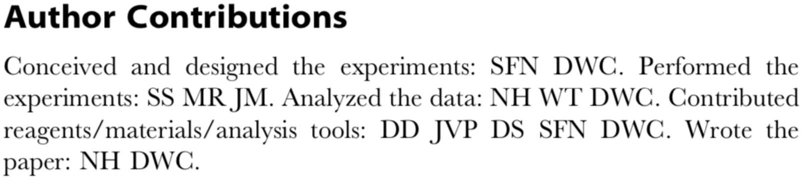
eccentricity $\phi = 1.5$

Homer , Szelinger, Redman, Duggan, Tembe, Muehling, Pearson, Stephan, Nelson, & Craig (2008)

Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays.

Homer , Szelinger, Redman, Duggan, Tembe, Muehling, Pearson, Stephan, Nelson, & Craig (2008)

Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays.



Author Contributions

Conceived and designed the experiments: SFN DWC. Performed the experiments: SS MR JM. Analyzed the data: NH WT DWC. Contributed reagents/materials/analysis tools: DD JVP DS SFN DWC. Wrote the paper: NH DWC.

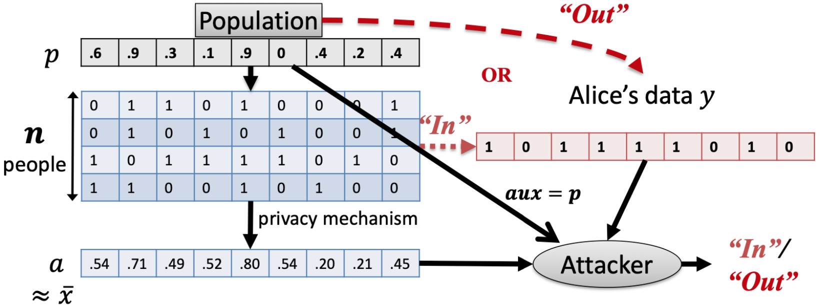
Homer *et al. (2008)*

Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays.

- Membership attack on individual's inclusion in sample/dataset with published summary statistics (means).
- Membership can violate privacy if membership betrays an implicit variable.
- *"Their [Braun et al.] work showed high specificity for the test statistic of Homer et al., but with possibility of low sensitivity."* Bruce Weir's Viewpoint: Individual Genotyping in Forensics and GWAS Contexts

Dwork, Smith, Steinke, Ullman, Vadhan (2015)

The Attacker



p	.6	.9	.3	.1	.9	0	.4	.2	.4

0	1	1	0	1	0	0	0	1
0	1	0	1	0	1	0	0	1
1	0	1	1	1	1	0	1	0
1	1	0	0	1	0	1	0	0

a	.54	.71	.49	.52	.80	.54	.20	.21	.45

1	0	1	1	1	1	0	1	0

$$A(y,a,p) = \begin{cases} \text{IN} & \text{if } \langle y-p, a-p\rangle > T \\ \text{OUT} & \text{if } \langle y-p, a-p\rangle \le T \end{cases}$$

$$T = T_{p,a} = O\left(\sqrt{d\log(1/\delta)}\right)$$