Review of *Local Private Hypothesis Testing:* χ^2 tests.

May 31, 2018

1 The BibTex

First, this has the following BibTex:

```
@ARTICLE{2017arXiv170907155G,
    author = {{Gaboardi}, M. and {Rogers}, R.},
     title = "{Local Private Hypothesis Testing: Chi-Square Tests}",
   journal = {ArXiv e-prints},
archivePrefix = "arXiv",
    eprint = {1709.07155},
 primaryClass = "math.ST",
  keywords = {Mathematics - Statistics Theory, Computer Science - Cryptography an
      year = 2017,
     month = sep,
    adsurl = {http://adsabs.harvard.edu/abs/2017arXiv170907155G},
   adsnote = {Provided by the SAO/NASA Astrophysics Data System}
}
```

2 The Paper

We'll try to answer the following questions:

1. *What new background is needed for this article*?

   This uses the concept of *concentrated differential privacy.* This is a variant of differential privacy where the "defining inequality" doesn't hold for *all events*. Recall that for (ϵ, δ) differential privacy, we say that for any two neighboring databases that:

   $$\Pr[M(x) \in S] \leq e^{\epsilon} \Pr[M(y) \in S] + \delta \tag{1}$$

   must hold for *all* subsets $S \subseteq \mathrm{Range}(M(x))$.

Concentrated differential privacy instead says that:

$$\mathbb{E}_{y\sim\mathcal{M}(x)}[\exp(t\ln(\frac{\Pr[M(x)=y}{\Pr[M(y)=y})-\rho)] \leq e^{t^2\rho}, \quad \forall t \geq 0 \tag{2}$$

Note that the $\ln\frac{f(x)}{f(y)} - \rho$ quantity is *essentially* the same thing as ϵ in the definition of (ϵ,ρ) differential privacy. If $M(x)$ was (ϵ,ρ) differentially private, we'd have that this inequality is just (bounded by):

$$\mathbb{E}[\exp(t\epsilon)] \leq \exp(t^2\rho) \tag{3}$$

We can view the Advanced Composition Theorem as a *concentration inequality*, saying that *privacy loss* is concentrated about the mean.

Concentrated Differential privacy makes a *new* (incomparable) definition of differential privacy, with the intent that it satisfies a concentration inequality under composition. The tradeoffs are that in (ϵ,δ) DP, with probability δ privacy loss can be *arbitrarily bad.* Concentrated differential privacy essentially says that the *privacy loss* is small mean, and subgaussian.

2. *What new techniques do the authors apply?*

   While I don't understand it super well, it seems like for 2 of their three cases they can derive distributions for $\tilde{p}$.

   Their computations of the asymptotic distribution of everything seems to utilize heavily some stuff from a multivariable version of probability, which I've never seen before.

3. *What new results do the authors get?*

   They develop three different χ^2 hypothesis test for *local, concentrated* differential privacy. These are:

   (a) A statistic guaranteed to converge to a χ^2 distribution under H_0

   (b) A statistic guaranteed to converge to a χ^2 distribution under H_0 when a private value is chosen from each participant via the exponential mechanism

   (c) A statistic that converges to χ^2 when private data is chosen via a bit flipping mechanism.

   They also develop the corresponding independence tests. Finally, they show experimental evidence that no single one of the developed tests is superior (in power) in all cases.