A Simpler Example of the Packing Lower Bound

CS 229r: Mathematical Approaches to Data Privacy, Fall 2012

February 19, 2013

We recommend reading this example in place of (or at least prior to) Theorem 113 in the Dwork–Roth text.

Theorem 1. *Let our data universe be $\mathcal{X} = \{0,1\}^k$. Let $M : \mathcal{X}^n \to \mathbb{R}^k$ be an ε-differentially private mechanism such that for every database $x \in \mathcal{X}^n$, with probability at least $1/2$, $M(x)$ outputs all of the 1-way marginals of x with error smaller than $n/2$. (That is, for each $j = 1, \ldots, k$, the j'th component of $M(x)$ should approximately equal the number of rows of x whose j'th bit is 1, up to an error smaller than $n/2$.) Then $n = \Omega(k/\varepsilon)$.*

Note that we are switching back to non-normalized summation queries, for consistency with Dwork–Roth Chapter 8. Note that the bound $n = \Omega(k/\varepsilon)$ is tight to within a constant factor (why?), and separates ε differential privacy and (ε, δ) differential privacy (why?).

Proof: For every string $w \in \{0,1\}^k$, consider the database x_w that consists of n rows, all of which equal w. Let $B_w \subseteq \mathbb{R}^k$ consist of all tuples of numbers that provide answers to the 1-way marginals on x_w with error $< n/2$. That is, $B_w = \{(a_1, \ldots, a_k) \in \mathbb{R}^k : \forall i | a_i - n \cdot w_i| < n/2\}$. (Put differently B_w is the open ℓ_∞ ball of radius $n/2$ around $n \cdot w \in \{0, n\}^k$.) Notice that the sets B_w are disjoint from each each other.

If M is an accurate mechanism for answering 1-way marginals, then for every w, $\Pr[M(x_w) \in B_w] \geq 1/2$. Thus, setting $\Delta = n$ and $s = k$ in Lemma 110, we have $\varepsilon \geq (\ln 2) \cdot (k-1)/n$. □

For those who look at the proof of Theorem 111 in Dwork–Roth: it gives a similar lower bound of $n = \Omega(k/\varepsilon)$ for answering k low-sensitivity queries (but not counting queries), which at first seems like a weaker result than the above. The advantage of Thm. 111 is that the lower bound only requires a data universe of size $d = k$ for getting the lower bound,[1] whereas the above lower bound for 1-way marginals uses a data universe of size $d = 2^k$. We need to use a large data universe to get such a lower bound for answering counting queries (why?). Also, it is known that with (ε, δ) differential privacy, one can accurately answer $k = \exp(n^{\Omega(1)})$ arbitrary low-sensitivity queries (even over data universes of size $d = \exp(n^{\Omega(1)})$), so Theorem 111 gives an exponential separation between pure ε and (ε, δ) differential privacy. On hw2, you may see another dramatic separation between ε and (ε, δ) differential privacy.

[1] In fact, if we view databases as ordered tuples in $\mathcal{X}^n$ rather than multisets, the data universe only needs to be of size $d = 2$.