HW for Week 1

Mark Schultz

May 29, 2018

1 Probability Pt 1

Exercise 1.1 (Dobrow 2.20)**.** Let $X \sim \text{Bern}(0.5)$ be the r.v. corresponding to the initial (either black or white) counter. The bag then contains $\{W, X\}$. We want to compute:

$$\begin{aligned}\Pr[\text{Second draw W} \mid \text{First draw W}] &= \Pr[\text{Second draw W} \mid \text{First draw W}, X = W]\Pr[X = W]\\ &+ \Pr[\text{Second draw W} \mid \text{First draw W}, X = B]\\ &= (1)\frac{1}{2} + (0)\frac{1}{2} = \frac{1}{2}\end{aligned}$$

Exercise 1.2 (Dobrow 3.2)**.** Let A, B, C be independent, with:

$$\Pr[A] = 1/3, \quad \Pr[B] = 1/4, \quad \Pr[C] = 1/5 \tag{1}$$

Find:

1. $$\Pr[ABC] = \Pr[A]\Pr[B]\Pr[C] = \frac{1}{60}$$

2. $$\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B] + \Pr[C] = \frac{20 + 15 + 12}{60} = \frac{47}{60}$$

3. $$\Pr[AB \mid C] = \frac{\Pr[ABC]}{\Pr[C]} = \frac{1/60}{1/5} = \frac{1}{12}$$

4. $$\Pr[B \mid AC] = \frac{\Pr[ABC]}{\Pr[AC]} = \Pr[B] = \frac{1}{4}$$

5. $$\begin{aligned}\Pr[\text{At most one event occurs}] &= \Pr[A^cB^cC^c] + \Pr[AB^cC^c] + \Pr[A^cBC^c] + \Pr[A^cB^cC]\\ &= \frac{2}{3}\frac{3}{4}\frac{4}{5} + \frac{1}{3}\frac{3}{4}\frac{4}{5} + \frac{2}{3}\frac{1}{4}\frac{4}{5} + \frac{2}{3}\frac{3}{4}\frac{1}{5}\\ &= \frac{1}{60}(24 + 12 + 8 + 6)\\ &= \frac{50}{60} = 5/6\end{aligned}$$

Exercise 1.3 (Dobrow 4.14)**.** Let X, Y, Z be i.i.d. with $X \sim 1 + \text{Bern}(1/2)$. Find the pmf of (X, Y, Z).

We have the pmf will be the product of the original pmf's, which will end up being a uniform probability distribution on the cube with vertices (x_1, x_2, x_3), where $x_i \in \{1, 2\}$.

Exercise 1.4 (Dobrow 4.19)**.** A joint pmf is given by:

$$\Pr[X=1,Y=1]=1/8, \quad \Pr[X=2,Y=1]=1/8, \quad \Pr[X=1,Y=2]=1/4, \quad \Pr[X=2,Y=2]=1/2 \tag{2}$$

1. Find the marginal distributions of X and Y:

$$\Pr[X=x] = \sum_{y \in Y} \Pr[X=x, Y=y] = \begin{cases} 3/8 & x=1 \\ 5/8 & x=2 \end{cases}$$
$$\Pr[Y=y] = \sum_{x \in X} \Pr[X=x, Y=y] = \begin{cases} 1/4 & y=1 \\ 3/4 & y=2 \end{cases}$$

2. Are X and Y independent? We have that:

$$1/8 = \Pr[X=1,Y=1] \neq \Pr[X=1]\Pr[Y=1] = \frac{3}{32} \tag{3}$$

3. Compute $\Pr[XY \leq 3]$ We have that:

$$\Pr[XY \leq 3] = 1 - \Pr[XY \neq 4] = 1 - \Pr[X=2, Y=2] = 1/2$$

4. Compute $\Pr[X+Y > 2]$

$$\Pr[X+Y > 2] = 1 - \Pr[X+Y \leq 2] = 1 - \Pr[X=1, Y=1] = 7/8$$

2 Crypto Pt 1

Exercise 2.1 (K&L 2.3)**.** Prove or refute: Any encryption scheme with message space $\mathcal{M}$ is perfectly secret if and only if, for every probability distribution over $\mathcal{M}$ and every $c_0, c_1 \in \mathcal{C}$, we have $\Pr[C = c_0] = \Pr[C = c_1]$.

False: Consider a perfectly secret encryption scheme with $\mathcal{M} = \mathcal{C} = \mathcal{K} = \{0, 1\}$ (such as a one-time pad). Now, consider a new encryption scheme where $\mathcal{C} = \{0, 1, 2\}$, but Enc and Dec aren't changed at all. Then, we have that:

$$0 = \Pr[\text{Enc}_K(m) = 2] \neq \Pr[\text{Enc}_K(m) = 0] = \frac{1}{2} \tag{4}$$

Exercise 2.2 (K&L 2.6)**.** For each of the following encryption schemes, state whether the scheme is perfectly secret. Justify your answer:

1. $M = \{0, \ldots, 4\}$, *Gen* chooses key uniformly from $\{0, \ldots, 5\}$. Enk returns $[k + m \mod 5]$, dec returns $[c - k \mod 5]$.

   This will not be perfectly secret, as $k = 0$ and $k' = 5$ being in the same residue class mod 5 will bias the encryption. Specifically, let $m_1 = 0, m_2 = 1$, and $c = 0$. Then, we have that:

$$\Pr[\text{Enc}_K(0) = 0] = \frac{2}{6}, \quad \Pr[\text{Enc}_K(1) = 0] = \frac{1}{6} \tag{5}$$

   (Because $0 + 0 \equiv 0$, and $5 + 0 \equiv 0$, but only $4 + 1 \equiv 0$).

2. This will be perfectly secret. This is because:

$$\begin{aligned}\Pr[\mathrm{Enc}_K(m_1) = c] &= \Pr[(m_1||0) \oplus (K||0) = c] \\ &= \Pr[m_1||0 = (K||0) \oplus c]\end{aligned}$$

We want to know if $\Pr[m_1||0 = (K||0) \oplus c] = \Pr[m_2||0 = (K||0) \oplus c]$. If the last bit of c is 1, then the probability of each event will be zero. If the last bit of c is zero, then the problem reduces to the security of the one-time pad. As the one-time pad is perfectly secure, this would be perfectly secure as well.

Note that $\mathcal{C} = \{0,1\}^\ell$ leads to this being secure as well. The only potential problem is at $c = b||1$, but then:

$$\Pr[\mathrm{Enc}_K(m_1) = c] = \Pr[\mathrm{Enc}_K(m_2) = c] \tag{6}$$

This is either $1/2^{l-1} = 1/2^{l-1}$, or $0 = 0$. Either way, it's fine.

Exercise 2.3 (K&L 2.7)**.** Is the one-time pad where $\mathcal{K} = \left\{x \in \{0,1\}^\ell \mid x \neq 0^\ell\right\}$ perfectly secret?

Nope, it's well-known that perfectly secret schemes have $|\mathcal{K}| \geq |\mathcal{M}|$, which this scheme contradicts.

Exercise 2.4 (K&L 2.13)**.** Perfectly secret for pairs.

1. By exercise 2.2, we have that Enc_K may be made deterministic solely by redefining $\mathcal{K}$.

Now, let $c_1 = c_2$. Then, we have that $\mathrm{Enc}_k(m_1) = \mathrm{Enc}_k(m_2)$. It follows that $\mathrm{Dec}_k(\mathrm{Enc}_k(m_1)) = \mathrm{Dec}_k(\mathrm{Enc}_k(m_2)) \implies m_1 = m_2$. We then have that:

$$\Pr[M_1 = m_1 \wedge M_2 = m_2 \mid C_1 = c_1 \wedge C_2 = c_1] = 0, \quad \Pr[M_1 = m_1 \wedge M_2 = m_2] \neq 0 (\forall m_1, m_2) \tag{7}$$

2. Let $\mathcal{M} = \mathcal{C} = \mathbb{F}_p$ for a prime p. Intuitively, we should expect $|\mathcal{K}| \geq 2|\mathcal{M}|$ (both (m_0, m_1) could be presented to Enc_K at the same time, and if this wasn't true this wouldn't match with our normal notion of perfect security). For this reason, let $\mathcal{K} = \mathbb{F}_p \times \mathbb{F}_p^\times$ be the uniform distribution on all *pairs* of (distinct) elements. We have that $\mathbb{F}_p^\times = \mathbb{F}_p \setminus \{0\}$ is the set of all (multiplicatively) invertible elements of $\mathbb{F}_p$.

Our encryption function will be:

$$\mathrm{Enc}_{(k_0,k_1)}(m) = k_0 + k_1 m \tag{8}$$

We can decrypt (correctly) via:

$$\mathrm{Dec}_{(k_0,k_1)}(c) = \frac{c - k_0}{k_1} \tag{9}$$

Now, we want to show that $\Pr[M_1 = m_1 \wedge M_2 = m_2 \mid C_1 = c_1 \wedge C_2 = c_2] = \Pr[M_1 = m_1 \wedge M_2 = m_2]$. This will come down to showing that $\Pr[C_1 = c_1 \wedge C_2 = c_2 \mid M_1 = m_1 \wedge M_2 = m_2] = \Pr[C_1 = c_1 \wedge C_2 = c_2]$, and then applying Bayes' theorem.

We can examine this conditional probability:

$$\begin{aligned}\Pr[K_0 + K_1 m_1 = c_1 \wedge K_0 + K_1 m_2 = c_2] &= \sum_{(K_0,K_1) \in \mathbb{F}_p \times \mathbb{F}_p^\times} \Pr[k_0 + k_1 m_1 = c_1 \wedge k_0 + k_1 m_2 = c_2] \Pr[(K_0, K_1)] \\ &= \frac{1}{p(p-1)} \sum_{(K_0,K_1)} \Pr[k_0 + k_1 m_1 = c_1 \wedge k_0 + k_1 m_2 = c_2]\end{aligned}$$

The probability this occurs will be precisely $\frac{1}{p(p-1)}$. To see this, note that (K_0, K_1) parametrize *all* (non-constant) lines $f : \mathbb{F}_p \to \mathbb{F}_p$. Any line is completely determined by two points. One line will pass through the "points" (m_1, c_1), (m_2, c_2), and all others will not. So, the event will be true if (K_0, K_1) is the correct line, which happens with the stated probability.

So, we get that:

$$\begin{aligned}\Pr[K_0 + K_1 m_1 = c_1 \wedge K_0 + K_1 m_2 = c_2] &= \frac{1}{p(p-1)} \sum_{(K_0,K_1)} \frac{1}{p(p-1)} \\ &= \frac{p(p-1)}{p^2(p-1)^2} = \frac{1}{p(p-1)}\end{aligned}$$

This is precisely the number of elements in $\mathbb{F}_p$.

If you didn't want to be restricted to prime-power fields, the above argument could be repeated with $\mathbb{F}_q$ where $q = p^k$. The only difference is that $q = p^k$ would be inserted everywhere.

3 Probability Pt 2

Exercise 3.1 (Derive Dobrow Eq. 4.14)**.**

Exercise 3.2. Evaluate the following:

$$X \perp\!\!\!\perp Y \iff \operatorname{Cov}(X,Y) = 0 \tag{10}$$

$[\implies]$.

Assume $X \perp\!\!\!\perp Y$. Then, we have that:

$$\begin{aligned}\operatorname{Cov}(X,Y) &= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\ &= \sum_{x,y}(x-\mu_X)(y-\mu_Y)p_{X,Y}(X=x, Y=y) \\ &= \sum_{x,y}(xy - (\mu_Y x + \mu_X y) + \mu_X\mu_Y)p_X(X=x)p_Y(Y=y), \quad X \perp\!\!\!\perp Y \\ &= \sum_{x,y} xyp_X(X=x)p_Y(Y=y) - \sum_{x,y}(\mu_Y x + \mu_X y)p_X(X=x)p_Y(Y=y) \\ &+ \mu_X\mu_Y \sum_{x,y} p_X(X=x)p_Y(Y=y) \\ &= \sum_{x,y} xyp_X(X=x)p_Y(Y=y) - \sum_x \mu_X p_X(X=x) \sum_y yp_Y(Y=y) \\ &- \sum_y \mu_Y p_Y(Y=y) \sum_x xp_X(X=x) + \mu_X\mu_Y \\ &= \sum_{x,y} xyp_X(X=x)p_Y(Y=y) - \mu_X\mu_Y \sum_x p_X(X=x) - \mu_X\mu_Y \sum_y p_Y(Y=y) + \mu_X\mu_Y \\ &= \sum_x xp_X(X=x) \sum_y yp_Y(X=x) - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y \\ &= \sum_x xp_X(X=x)\mu_Y - \mu_X\mu_Y \\ &= \mu_Y\mu_X - \mu_X\mu_Y \\ &= 0\end{aligned}$$

$[\impliedby]$ The other direction is false, because $\operatorname{Cov}(X,Y)$ is a measure of *linear* dependence, but $X \perp\!\!\!\perp Y$ requires that there is *no* dependence of any sort.

We can try to find a counterexample. Let $X \sim \text{Unif}\{-1,0,1\}$, and let $Y \sim f(X)$, for some to-be-determined function f. Then, we have that:

$$\begin{aligned}\text{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \sum_{x=-1}^{1}\sum_{y\in A} xyp_{X,Y}(X=x, Y=y) - 0 = \sum_{x=-1}^{1} xf(x)p_X(X=x) = -f(-1) + 0 + f(1)\end{aligned}$$

So, we just need some f such that $f(1) = f(-1)$, or some *even* $f(x)$. Choosing $f(x) = x^2$ works, so we get that $\text{Cov}(X, X^2) = 0$, but clearly X^2 is dependent on X.

It seems like we could generalize this for X being any symmetric, mean 0 distribution, and f being any even function. If we remember that sums are just integrals with respect to the counting measure μ, this becomes "easy" to show, as we have that:

$$\mathbb{E}[Xf(X)] = \int xf(x)p_X(x)\mathrm{d}\mu \tag{11}$$

As $f(x)$ and $p_X(x)$ are both "even", the integrand will be odd, so if we integrate it over a origin-symmetric domain, it will be 0.

Exercise 3.3 (4.27)**.** Variance of the sum of n tetrahedronal dice rolls.

Tetrahedron die rolls will be distributed Unif$\{1,\dots,4\}$. Let X_i be distributed as this (and i.i.d.), and $Y = \sum_{i=1}^{n} X_i$. Then, we have that:

$$\text{Var}(Y) = n\text{Var}(X) \tag{12}$$

Now, we have that:

$$\mathbb{E}[X] = \frac{1+2+3+4}{4} = 5/2 \tag{13}$$

We then have that:

$$\mathbb{E}[X^2] = \frac{1+4+9+16}{4} = 15/2 \tag{14}$$

It follows that:

$$\text{Var}(X) = 15/2 - (5/2)^2 = 30/4 - 25/4 = 5/4 \implies \text{Var}(Y) = 5n/4 \tag{15}$$

Exercise 3.4 (4.37)**.** Suppose $X \perp\!\!\!\perp Y$, with $\text{Var}[X] = \sigma_X^2$ and $\text{Var}[Y] = \sigma_Y^2$. Let $Z = wX + (1-w)Y$ with $w \in (0,1)$. Find the variance of Z. What value of w is this minimized for?

$$\text{Var}(Z) = w^2\sigma_X^2 + (1-w)^2\sigma_Y^2 = w^2\sigma_X^2 + (1 - 2w + w^2)\sigma_Y^2$$

(Note no covariance term due to independence). We then have that:

$$\frac{\partial}{\partial w}\text{Var}[Z] = 2w\sigma_X^2 - 2\sigma_Y^2 + 2w\sigma_Y^2 = 0 \implies w = \frac{\sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}$$

Exercise 3.5 (4.48)**.** Let X be the first of two fair die rolls (so $X \sim \text{Unif}\{1,\dots,6\}$. Also, let Y be the second die roll). Let M be the maximum of the two rolls.

- Find the conditional pmf of M given $X = x$.

  Given that the first die roll has occurred, we have that:

$$\begin{aligned}\Pr[M = m \mid X = x] &= \Pr[\max(Y, x) = m \mid X = x] \\ &= \begin{cases} 0 & x > m \\ \Pr[Y \le m] & x = m \\ \Pr[Y = m] & x < m \end{cases} \\ &= \begin{cases} 0 & x > m \\ m/6 & x = m \\ 1/6 & x < m \end{cases}\end{aligned}$$

- Find $\mathbb{E}[M \mid X = x]$

$$\begin{aligned}\mathbb{E}[M \mid X = x] &= \sum_{m=1}^{6} m \Pr[M = m \mid X = x] \\ &= \sum_{m<x} m \Pr[M = m \mid X = x] + x \Pr[M = x \mid X = x] + \sum_{m>x} m \Pr[M = m \mid X = x] \\ &= 0 + x(x/6) + (6 - x + 1)(1/6) = \frac{x^2 - x + 7}{6}\end{aligned}$$

- Find the joint PMF of X and M:

$$\begin{aligned}\Pr[M = m, X = x] &= \Pr[M = m \mid X = x] \Pr[X = x] \\ &= \Pr[M = m \mid X = x] \frac{1}{6} \\ &= \begin{cases} 0 & x > m \\ m/36 & x = m \\ 1/36 & x < m \end{cases}\end{aligned}$$

4 Diff Privacy Pt 1

Let X be a database of reed students, where we have their name, their year at Reed, and their major.

- The sensitivity of this is 1, as each student has one major.

  As the sensitivity is only one, this can be acchieved by standard query counting techniques. Specifically, count the number of students, then add Lap(1/.7) = Lap(1.428) noise to each count. Specifically, make a single vector-valued query $q : X \to \mathbb{R}^k$ where there are k majors. As each student impacts at most one count in this query, $\Delta q = 1$, so this will be $\epsilon = 0.7$ differentially-private.

- Yes, as it's just post-processing

- We could do this with two vector-valued queries (one as before, another $q : X \to \mathbb{R}^4$ where each axis corresponds to Freshman, Sophomore, Junior, Senior). Then, we could allocate $\epsilon/2$ privacy to each (or any $a, b > 0$ privacy with $a + b = \epsilon$) to get the result with $(\epsilon, 0)$-DP.

  We could alternatively do this with one query, but the function will now have sensitivity $\Delta f = 2$ (each person can change the count of both the majors, and the years). As the noise is Lap$(\Delta f/\epsilon)$, we have that:

$$\frac{\Delta f}{(\epsilon/2)} = \frac{2\Delta f}{\epsilon} \tag{16}$$

  So these both involve adding the same amount of noise.

- This could be done with the exponential mechanism. The utility here should be:

$$u(x, \text{major}) = \#\{\text{people in } x \text{ with the given major}\} \tag{17}$$

  We can now investigate how accurate this is.

  The exponential mechanism has the "guarantee" that it is highly unlikely that the returned element r has a utility score that is less than $OPT_u(x)$ by more than an additive factor of:

$$O((\Delta u/\epsilon) \log |\mathcal{R}|) \tag{18}$$

More specifically, the probability it differs by more than:

$$\frac{2\Delta u}{\epsilon}(\ln|\mathcal{R}|+t) \tag{19}$$

is less than e^{-t}. Here, $\mathcal{R}$ is the list of all majors at Reed. Ignoring ad-hoc majors (which is fine, as they'll never be most populous anyway), it appears there are 40 official majors offered. This means our additive bound is bounded by:

$$\frac{2}{\epsilon}(3.7+t)=\frac{7.4+2t}{\epsilon} \tag{20}$$

We therefore have that, with probability bounded above by e^{-t}, that:

$$|OPT_u(x)-\text{Output}|\leq\frac{7.4+2t}{\epsilon} \tag{21}$$

The accuracy of the Laplace mechanism can be given by Theorem 3.8 in the book:

Let $f:\mathbb{N}^{|\mathcal{X}|}\to\mathbb{R}^k$ and let $y=\mathcal{M}_L(x,f(\cdot),\epsilon)$. Then, for all $\delta\in(0,1]$:

$$\Pr\left[\|f(x)-y\|_\infty\geq\ln\left(\frac{k}{\delta}\right)\left(\frac{\Delta f}{\epsilon}\right)\right]\leq\delta \tag{22}$$

Corr. 3.12 said something similar for the Exponential mechanism. Specifically, it said that:

$$\Pr\left[|u(\mathcal{M}_E(x,u,\mathcal{R}))-\text{OPT}_U(x)|\leq\frac{2\Delta u}{\epsilon}(\ln|\mathcal{R}|)+t)\right]\leq e^{-t} \tag{23}$$

So try to make these look similar, set $\delta=e^{-t}$ (now, $\delta\in[0,\infty)$). This gives the expression:

$$\Pr\left[\|f(x)-y\|_\infty\geq\frac{\Delta f}{\epsilon}(\ln(k)+t)\right]\leq e^{-t} \tag{24}$$

- The Laplace mechanism seems inappropriate for this (specializing to think about a single major, we could represent years as $1,2,3,4$, but if there are 100 1's, and 50 4's, and nothing else, we could easily output 2 in [major] error).

  Instead, thinking about this via the exponential mechanism. To do this, we'll have to define a utility function. For now, specialize to the problem for a *single major.*

  Let $u_m(x,r)$ be the utility (for major m) of r given data x. Then, $u_m(x,r)$ should give utility to each (potentially outputted) year somehow following the true rankings of the years. This means, if $\#F\leq\#J\leq\#\text{Sophomore}\leq\#\text{Senior}$, then the utilities of each result should follow this inequality as well.

  An easy choice is $u_m(x,r)=\#\{\text{students in that year (with that major)}\}$.

  We then need to define an *overall* utility (or query all 40 utilities with privacy $\epsilon/40$, and then post-process the results. This seems bad at face value, but I haven't investigated it more). We can view our intermediate result as $(u_{m_1}(x,r),u_{m_2}(x,r),\ldots,u_{m_{40}}(x,r))\in\mathbb{R}^{40}$. We then somehow need to combine these all into a single real number. Here, we have some choices to make:

  - To combine them, we could use some sort of norm. An ℓ_1-norm may be appropriate (especially since it makes the Δu analysis much easier than other ℓ_p norms).
  - Before we combine them, potentially with an ℓ_p norm, do we want to *normalize* the various utilities? Doing so will mean that we can potentially be *very* inaccurate for large departments, but not doing so means we can potentially be *very* inaccurate for small departments.

If we don't normalize things, we have that:

$$u(x,r) = \sqrt[p]{\sum_{i=1}^{40} |\#\{\text{Students in major } m_i \text{ with year } r_i\}|^p}$$

We then have that:

$$\Delta u = \max_r \max_{\text{neigh. } x,y} |u(x,r) - u(y,r)| \tag{25}$$

Adding a new student can impact at most one major, so it will impact at most one summand in the norm (regardless of r). As a result, this computation will boil down to[1]:

$$\Delta u \leq \sqrt[p]{x+1} - \sqrt[p]{x}$$

Worst case scenario is $x = 0$, so $\Delta u = 1$.

Even if $x \neq 0$ (so our database actually has records), we get the same worst-case bound. This is because p-th roots are concave, so $\sqrt[p]{x+y} \geq \sqrt[p]{x} + \sqrt[p]{y}$.

It follows the "biggest" Δu will occur when all records are in a single major, in a single year, and then the computation becomes:

$$\Delta u \leq |D| + 1 - |D| = 1 \tag{26}$$

Exercise 4.1. 1. I don't think this is a fair statement. As an easy example, if they smoke, and it's discovered there's some strong correlation between that and something else (say cancer), people will learn there's a decent chance that person has that other thing.

2. This seems fairer. Specifically:

$$\Pr[\text{Participation causes something to be learned}] \leq e^\epsilon \Pr[\text{Something learned without participation}] \tag{27}$$

Here, $e^\epsilon = e^{.2} \approx 1.22$. So, people will have a 20% higher chance of learning something about that person if they participate. I'd tell them this is the precise meaning of "learn much", and let them decide if this is worth it.

3. This is essentially the previous question, but "learning much about you" means "learning whether you satisfy the predicate 'This person participated in the study"'.

Exercise 4.2. 1. Deterministic, non-constant functions can't be differentially private.

2. This isn't differentially private. Let x be a database with a yes responses, and y be one with $a+1$ yes responses. Then, we would need that:

$$\Pr[M(x) \in \{a-1, a+1\}] \not\leq e^\epsilon \Pr[M(y) \in \{a-1, a+1\}]$$

The LHS of this is 1, and the RHS of this is 0. So, this can't be differentially private.

3. This also won't be differentially private. Let x and y be as before. Then:

$$\Pr[M(x) \in \{a\}] \not\leq e^\epsilon \Pr[M(y) \in \{a\}]$$

The LHS of this 1/2, but the RHS of this is 0.

4. The mechanism here is:

   (a) Flip a coin. If heads, release a.

   (b) If tails, flip another coin. If heads, release $a+1$, if tails, release $a-1$.

[1] There will be an application of triangle inequality first, but this is omitted due to how "easy" it is.

Let x be a database with a yes responses, and y be a database with $a+1$ yes responses. Then, we have that:

$$\Pr[M(x) = k] = \begin{cases} 1/2 & k = a \\ 1/4 & k \in \{a-1, a+1\} \end{cases} \tag{28}$$

We also have that:

$$\Pr[M(y) = k] = \begin{cases} 1/2 & k = a+1 \\ 1/4 & k \in \{a, a+2\} \end{cases} \tag{29}$$

This motivates looking at $S = \{a-1\}$. Then, we have that:

$$\Pr[M(x) = a-1] \leq e^{\epsilon} \Pr[M(y) = a-1] \tag{30}$$

The LHS of this is 1/4, but the RHS of this is 0, so this is a contradiction.

We'll let Y denote the random variable that we release. Let $x = (x_1, x_2, \ldots, x_{500})$. let $B_i \sim \text{Bern}(p)$ be i.i.d. Bernoulli r.v.s. Then, we have that:

$$Y(x) = \sum_{i=1}^{500} x_i \oplus B_i \tag{31}$$

Let x be a database with a yes responses, and y be a database with $a+1$ yes responses.

Given x, we can attempt to analyze $Y(x)$ by noting that $Y(x)$ is the sum of a Bern(p) r.v.'s, and $500-a$ Bern($1-p$) r.v.'s. We can more simply say that:

$$Y(x) \sim \text{Binom}(a, p) + \text{Binom}(500-a, 1-p) \tag{32}$$

We can use the convolution formula to find the distribution of this. Specifically, if $Z \sim X + Y$, then:

$$p_Z(t) = \sum_{\text{y} \in \text{Supp}(Y)} p_X(t-y) p_Y(t)$$

Here, let $X \sim \text{Binom}(a, p)$, and $Y \sim \text{Binom}(500-a, 1-p)$. Then, we have that:

$$\begin{aligned} p_Z(t) &= \sum_{y=0}^{500-a} p_X(t-y) p_Y(t) \\ &= \sum_{y=0}^{500-a} \left(\binom{a}{t-y} (p)^{t-y} (1-p)^{a-(t-y)} \right) \left(\binom{500-a}{t} (1-p)^t p^{500-a-t} \right) \\ &= \sum_{y=0}^{500-a} \binom{a}{t-y} \binom{500-a}{t} p^{500-a-y} (1-p)^{a+y} \\ &= \sum_{y=0}^{500-a} \frac{a!}{(t-y)!(a-t+y)!} \frac{(500-a)!}{t!(500-a-t)!} p^{500-a-y} (1-p)^{a+y} \end{aligned}$$

Now, if $p = 0$ or $p = 1$, this will deterministically be either a or $500-a$ respectively, which will therefore not be private.

If $p = 0.5$, we get that $Y(x) \sim \text{Binom}(a, p) + \text{Binom}(500-a, p) = \text{Binom}(500, p)$. Here, the output of is only dependent on the size of the database, and not the contents of it at all.

If $p = 0.1$, we have that $Y(x) \sim \text{Binom}(a, 0.1) + \text{Binom}(500-a, 0.9)$. To show this is ϵ-differentially private, we need to show that for all $\mathcal{S} \subseteq \mathbb{Z}$, that if x, y are neighboring databases, then:

$$\Pr[M(x) \in S] \leq e^{\epsilon} \Pr[M(y) \in S] \tag{33}$$

Note that $\Pr[M(x) \in \{s_1, s_2, \ldots, s_k\}] = \sum_{j=1}^{k} \Pr[M(x) = s_j]$, so we need to compute $\Pr[M(x) = s_j]$. We have that (given x has a yes responses):

$$\Pr[M(x) = k] = \sum_{y=0}^{500-a} \binom{a}{t-y} \binom{500-a}{t} (0.1)^{500-a-y} (0.9)^{a+y} \tag{34}$$

5 Finding PDFS

Exercise 5.1. Let $X \sim \text{Exp}(\lambda)$, and $Y \sim \text{Exp}(\lambda)$. Let $W = X + Y$. What's f_W?

$$\begin{aligned}
F_W(w) &= \Pr[W \leq w] \\
&= \Pr[X + Y \leq w] \\
&= \iint_{(x,y):X+Y \leq w} f_{X,Y}(x, y) \mathrm{d}x\mathrm{d}y \\
&= \int_{x=0}^{w} \int_{y=0}^{w-x} f_X(x) f_Y(y) \mathrm{d}y\mathrm{d}x, \qquad X \perp\!\!\!\perp Y \\
&= \int_{x=0}^{w} \int_{y=0}^{w-x} \lambda e^{-\lambda x} \lambda e^{-\lambda y} \mathrm{d}y\mathrm{d}x \\
&= \lambda^2 \int_{x=0}^{w} \int_{y=0}^{w-x} e^{-\lambda x - \lambda y} \mathrm{d}y\mathrm{d}x \\
&= \lambda^2 \int_{x=0}^{w} e^{-\lambda x} \int_{y=0}^{w-x} e^{-\lambda y} \mathrm{d}y\mathrm{d}x \\
&= \lambda^2 \int_{x=0}^{w} e^{-\lambda x} \left(-\frac{1}{\lambda} e^{-\lambda y} \Big|_0^{w-x} \right. \mathrm{d}x \\
&= \lambda^2 \int_{x=0}^{w} e^{-\lambda x} \left(-\frac{e^{-\lambda(w-x)}}{\lambda} + \frac{1}{\lambda} \right) \mathrm{d}x \\
&= \lambda \int_{x=0}^{w} -e^{-\lambda w} + e^{-\lambda x} \mathrm{d}x \\
&= -\lambda w e^{-\lambda w} - e^{-\lambda w} + 1 \\
f_W(w) &= -\lambda e^{-\lambda w} + \lambda^2 w e^{-\lambda w} + \lambda e^{-\lambda w} \\
&= \lambda^2 w e^{-\lambda w}
\end{aligned}$$

Exercise 5.2. Let $Z \sim \mathcal{N}(0, 1)$. Find the pdf of $X = Z^2$.

We have that:

$$\begin{aligned}
F_X(x) &= \Pr[X \leq x] \\
&= \Pr\left[Z^2 \leq x\right] \\
&= \Pr\left[-\sqrt{x} \leq Z \leq \sqrt{x}\right] \\
&= F_Z(\sqrt{x}) - F_Z(-\sqrt{x}) \\
f_X(x) &= \frac{1}{2\sqrt{x}} f_Z(\sqrt{x}) + \frac{1}{2\sqrt{x}} f_Z(-\sqrt{x}) \\
&= \frac{1}{2\sqrt{x}} \frac{1}{\sqrt{2\pi}} \left(e^{-x/2} + e^{-x/2} \right) \\
&= \frac{e^{-x/2}}{\sqrt{2\pi x}}
\end{aligned}$$

Exercise 5.3. Prepare talk on something — Advanced Composition Theorems.

6 What is Composition?

We want composition to mean *at least* two things:

1. Querying the *same* database with *multiple* differentially private algorithms. So querying $Q(x), H(x), G(x)$, or even just querying $Q(x)$, $Q(x)$, $Q(x)$.

   It's important to note that:

   $$(Q \circ H \circ G)(x) \tag{35}$$

   is *not* composition — after we query $G(x)$ first, the rest is just post-processing.

2. Querying *multiple* databases that all contain the same person, while keeping their *cumulative* privacy loss low. This is harder because it's intrinsically *adaptive* — the following situation is plausible:

   - An adversary queries $Q(x)$
   - After seeing $Q(x)$, the adversary tries to make a new database y that somehow *depends* on the information in $Q(x)$
   - The adversary queries $H(y)$, and repeats this process.

First, we'll introduce an alternative definition of differential privacy that ends up being equivalent to our initial definition. For two random variables Y, Z taking values in the same domain, we define the *max divergence*[2] to be:

$$D_\infty(Y||Z) = \max_{S \subseteq \mathrm{Supp}(Y)} \left(\ln \left(\frac{\Pr[Y \in S]}{\Pr[Z \in S]} \right) \right) \tag{36}$$

Then, $\mathcal{M}$ is differentially private if and only if, for any two neighboring databases x, y we have that:

$$D_\infty(\mathcal{M}(x)||\mathcal{M}(y)) \leq \epsilon \tag{37}$$

(and similarly with x and y permuted).

We can see this because:

$$\begin{aligned}
\exp(D_\infty(\mathcal{M}(x)||\mathcal{M}(y))) &\leq e^\epsilon \\
\exp\left(\max_{S \subseteq \mathrm{Supp}(\mathcal{M}(x))} \left(\ln \left(\frac{\Pr[\mathcal{M}(x) \in S]}{\Pr[\mathcal{M}(y) \in S]} \right) \right) \right) &\leq e^\epsilon \\
\max_{S \subseteq \mathrm{Supp}(\mathcal{M}(x))} \left(\frac{\Pr[\mathcal{M}(x) \in S]}{\Pr[\mathcal{M}(y) \in S]} \right) &\leq e^\epsilon, \quad \exp(\cdot) \text{ is strictly monotonically increasing} \\
\implies \forall S \subseteq \mathrm{Supp}(\mathcal{M}(x)) \qquad \frac{\Pr[\mathcal{M}(x) \in S]}{\Pr[\mathcal{M}(y) \in S]} &\leq e^\epsilon
\end{aligned}$$

We'll also need a version of D_∞ that is equivalent to (ϵ, δ) differential privacy. This will be:

$$D_\infty^\delta(Y||Z) = \max_{S \subseteq \mathrm{Supp}(Y) : \Pr[Y \in S] > \delta} \left(\ln \left(\frac{\Pr[Y \in S] - \delta}{\Pr[Z \in S]} \right) \right) \tag{38}$$

Then, we have that (ϵ, δ) differential privacy is equivalent to $D_\infty^\delta(\mathcal{M}(x)||\mathcal{M}(y)) \leq \epsilon$ for all neighboring databases.

Now, how do we model adaptive composition? Let $\mathcal{F}$ be the "a family of database access mechanisms" (it could be the set of all ϵ-differentially private mechanisms).

We define two experiments (0 and 1) as follows. We'll let A be an "adversary", which will mean a "stateful randomized algorithm".

[2] The ∞ is there because this is a special case of the Renyi Divergence of parameter ∞.

1. For $i = 1 \ldots k$:
    (a) A outputs two adjacent databases, x_i^0, x_i^1, a mechanism $\mathcal{M}_i$, and parameters w_i
    (b) A receives $y_i \leftarrow \mathcal{M}_i(w_i, x_i^b)$

Now, it will be useful to mention the *view* of A throughout the execution of this experiment. This is essentially "the information A has access to". As an example, for "Experiment 0", we'll have that $b = 0$, but A *won't know* that $b = 0$. What A *will* know is view(A), and is:

1. The randomness that A uses (as a randomized algorithm)
2. All of the $(y_1, \ldots, y_k)$ that are output by $\mathcal{M}_i$
3. All of the (x_i^0, x_i^1)
4. All of the $\mathcal{M}_i$
5. All of the w_i

Technically, only the first two are necessary to state, as the rest can be derived from them.

Intuitively, if we have x_i^0 always contain Bob's data, and x_i^1 always not contain Bob's data, then experiment 0 is somehow analogous to "Bob's data being in database x^0", and experiment 1 is "Bob's data not being in database x^1". We can then define a notion of differential privacy for a *family* of database access mechanisms.

Defn: We say that a family $\mathcal{F}$ of database access mechanisms satisfies (ϵ, δ)-differential privacy under k-fold adaptive composition if, for every adversary A, we have that $D_\infty^\delta(V^0 || V^1) \leq \epsilon$, where V^i is the view of A in experiment i.

For ϵ-differential privacy, just note that $D_\infty^0(X||y) = D_\infty(X||Y)$.

Thm: For all $\epsilon, \delta, \delta' > 0$, the class of (ϵ, δ)-differentially private mechanisms satisfies $(\epsilon', k\delta + \delta')$-differential privacy under k-fold adaptive composition for:

$$\epsilon' = \sqrt{2k \ln(1/\delta')\epsilon} + k\epsilon(e^\epsilon - 1) \tag{39}$$

Proof: The view of the adversary A is $v = (r, y_1, \ldots, y_k)$, where r are the coins, and the y_i are from the $\mathcal{M}_i$.

We'll define $V^i = (R^i, Y_1^i, \ldots, Y_k^i)$ denote the random variable corresponding to the view of A in experiment i.

Let:

$$B = \Big\{ v : \Pr\big[V^0 = v\big] > e^{\epsilon'} \Pr\big[V^1 = v\big] \Big\} \tag{40}$$

This is intuitively the "set of events that would violate $(\epsilon, 0)$ differential privacy". We want to show that:

$$\Pr\big[V^0 \in B\big] \leq \delta \tag{41}$$

This means that for any set (of views) S, we'll have[3]:

$$\Pr\big[V^0 \in S\big] \leq \Pr\big[V^0 \in S \setminus B\big] + \Pr\big[V^0 \in B\big] \leq e^{\epsilon'} \Pr\big[V^1 \in S\big] + \delta \tag{42}$$

This is equivalent to showing $D_\infty^\delta(V^0||V^1) \leq \epsilon'$. So, all we have to show is that $\Pr\big[V^0 \in B\big] \leq \delta$.

Now, note that we have that:

$$\ln\left(\frac{\Pr\big[V^0 = v\big]}{\Pr[V^1 = v]}\right) = \ln\left(\frac{\Pr\big[R^0 = r\big]}{\Pr[R^1 = r]} \prod_{k=1}^{k} \frac{\Pr\big[Y_i^0 = y_i \mid R^0 = r, Y_1^0 = y_1, \ldots, Y_{i-1}^0 = y_{i-1}\big]}{\Pr\big[Y_i^1 = y_i \mid R^1 = r, Y_1^1 = y_1, \ldots, Y_{i-1}^1 = y_{i-1}\big]}\right)$$

[3] By the Union Bound

Here, all that's happening is that we're rewriting:

$$\Pr[ABC] = \Pr[B \mid AB] \Pr[A \mid B] \Pr[B] \tag{43}$$

Assuming the coins are picked uniformly randomly, we'll have that:

$$\frac{\Pr\left[R^0 = r\right]}{\Pr\left[R^1 = r\right]} = 1 \tag{44}$$

We can drop this term, and use that $\ln(\prod_i A_i) = \sum_i \ln(A_i)$.

We can then use properties of logarithms to get:

$$\begin{aligned} * &= \sum_{i=1}^{k} \ln\left(\frac{\Pr\left[Y_i^0 = y_i \mid R^0 = r, Y_1^0 = y_1, \ldots, Y_{i-1}^0 = y_{i-1}\right]}{\Pr\left[Y_i^1 = y_i \mid R^1 = r, Y_1^1 = y_1, \ldots, Y_{i-1}^1 = y_{i-1}\right]}\right) \\ &\overset{\text{def}}{=} \sum_{i=1}^{k} c_i(r, y_1, \ldots, y_k) \end{aligned}$$

To analyze $c_i(r, y_1, \ldots, y_k)$, note that we're conditioning on *everything* but Y_i^0. We can think of this meaning that the prefix $(r, y_1, \ldots, y_{i-1})$ is fixed

7 Statistics 1

Exercise 7.1. Let $X \sim \mathcal{N}(\mu, \sigma^2)$.