CSC473: Advanced Algorithm Design Winter 2020

Worksheet: Couplings

Aleksandar Nikolov

In this worksheet you will:

- Learn the concept of *couplings*.
- Learn about measuring distanced between probability distributions (via *total variation distance*).
- Learn how to bound total variation distance via couplings.

Coupling. A coupling of two random variables X and Y taking values in Ω is a random variable $Z = (Z_1, Z_2)$ taking values in $\Omega \times \Omega$, such that Z_1 has the same probability distribution as X, and Z_2 has the same probability distribution as Y. I.e. for any $x, y \in \Omega$:

$$\mathbb{P}(X = x) = \sum_{z \in \Omega} \mathbb{P}(Z_1 = x, Z_2 = z) \qquad \mathbb{P}(Y = y) = \sum_{z \in \Omega} \mathbb{P}(Z_1 = z, Z_2 = y).$$

Exercise 1. *As an example, consider X which is distributed uniformly in $\{1, 2, 3\}$, and Y such that $\mathbb{P}(Y = 1) = 1/2$, and $\mathbb{P}(Y = 2) = \mathbb{P}(Y = 3) = 1/4$.*

***a**. Suppose $Z = (Z_1, Z_2)$ is a coupling of X and Y such that Z_1 and Z_2 are independent. Fill out the table below, where rows indicate values for Z_1, and columns values for Z_2, and each entry is the probability of the corresponding pair of values.*

$Z_1 \backslash Z_2$	*1*	*2*	*3*
1			
2			
3			

***b**. Suppose $Z = (Z_1, Z_2)$ is chosen as follows. First, we sample Z_1 uniformly in $\{1, 2, 3\}$. Then, with probability $3/4$, we set $Z_2 = Z_1$, and with probability $1/4$, we set $Z_2 = 1$. Fill out the table for this coupling, and verify that it is a coupling of X and Y.*

$Z_1 \backslash Z_2$	*1*	*2*	*3*
1			
2			
3			

***c**. Now fill out the table for a coupling $Z = (Z_1, Z_2)$ of X and Y such that $\mathbb{P}(Z_1 \neq Z_2) = 1/6$.*

$Z_1 \backslash Z_2$	*1*	*2*	*3*
1			
2			
3			

Note: Usually we do not explicitly write Z but instead just say there is a coupling (X, Y) of X and Y, or that we have coupled X and Y, and write just X instead of Z_1 and Y instead of Z_2.

Total Variation Distance. Given two random variables X and Y, taking values in the same finite set Ω, their total variation distance is

$$d_{tv}(X, Y) = \max_{S \subseteq \Omega} |\mathbb{P}(X \in S) - \mathbb{P}(Y \in S)|.$$

I.e. this is the biggest difference in the probability assigned to any event by the distribution of X and the distribution of Y.

The following equation is often useful:

$$d_{tv}(X, Y) = \frac{1}{2} \sum_{\omega \in \Omega} |\mathbb{P}(X = \omega) - \mathbb{P}(Y = \omega)|.$$

Lemma 1. *Two random variables X and Y taking values in a finite set Ω have $d_{tv}(X, Y) \leq \alpha$ if and only if there exists a coupling (X, Y) such that $\mathbb{P}(X \neq Y) \leq \alpha$.*

Exercise 2. *Let's prove a special case of the "only if" direction. Suppose that X and Y take values in $\Omega = \{0, 1\}$, and $\mathbb{P}(X = 1) = p$, $\mathbb{P}(Y = 1) = q$, $p > q$. Give a coupling of X and Y so that $\mathbb{P}(X = Y) = p - q$.*

$X \setminus Y$	*0*	*1*
0		
1		