Liam Hardiman

# 270A - Homework 2

**Problem 1.** The total variation distance between the distributions of random variables $X$ and $Y$ is defined by

$$d_{TV}(X, Y) = \sup_{B \in \mathcal{B}} |\mathbb{P}[X \in B] - \mathbb{P}[Y \in B]|$$

where the supremum is over all Borel subsets $B \subset \mathbb{R}$.

(a) Show that $d_{TV}$ is indeed a metric on the set of distributions.

*Proof.* The absolute value in the definition of $d_{TV}(X, Y)$ ensures that it is symmetric. We also have that if $d_{TV}(X, Y) = 0$, then $\mathbb{P}[X \in B] = \mathbb{P}[Y \in B]$ for all Borel sets, which happens if and only if $X$ is equal to $Y$ in distribution. Finally, for any triple $X, Y, Z$, the triangle inequality follows from the triangle inequality on $|\cdot|$:

$$d_{TV}(X, Z) = \sup_{B \in \mathcal{B}} |\mathbb{P}[X \in B] - \mathbb{P}[Z \in B]|$$

$$\leq \sup_{B \in \mathcal{B}} \left( |\mathbb{P}[X \in B] - \mathbb{P}[Y \in B]| + |\mathbb{P}[Y \in B] - \mathbb{P}[Z \in B]| \right)$$

$$\leq \sup_{B \in \mathcal{B}} |\mathbb{P}[X \in B] - \mathbb{P}[Y \in B]| + \sup_{B \in \mathcal{B}} |\mathbb{P}[Y \in B] - \mathbb{P}[Z \in B]|$$

$$= d_{TV}(X, Y) + d_{TV}(Y, Z).$$

$\square$

(b) Suppose $X$ and $Y$ are integer-valued random variables. Prove that

$$d_{TV}(X, Y) = \frac{1}{2} \sum_{k \in \mathbb{Z}} |\mathbb{P}[X = k] - \mathbb{P}[Y = k]|.$$

*Proof.* First, we claim that in this case, the supremum defining $d_{TV}(X, Y)$ is actually achieved. Let $E$ be the event defined by

$$E = \{k : \mathbb{P}[X = k] \geq \mathbb{P}[Y = k]\}.$$

For any event $A$ we have

$$\mathbb{P}[X \in A] - \mathbb{P}[Y \in A] \leq \mathbb{P}[X \in (A \cap E)] - \mathbb{P}[Y \in (A \cap E)] \tag{1}$$

$$\leq \mathbb{P}[X \in E] - \mathbb{P}[Y \in E].$$

The first inequality follows from the fact that for any $a \in A \setminus E$, $\mathbb{P}[X = a] - \mathbb{P}[Y = a] < 0$. The second inequality follows from the fact that $\mathbb{P}[X = k] - \mathbb{P}[Y = k] \geq 0$ for any $k \in E$. The exact same reasoning shows that

$$\mathbb{P}[Y \in A] - \mathbb{P}[X \in A] \leq \mathbb{P}[Y \in E^c] - \mathbb{P}[X \in E^c]. \tag{2}$$

We quickly see that the right-hand sides of (1) and (2) are equal:

$$(\mathbb{P}[X \in E] - \mathbb{P}[Y \in E]) - (\mathbb{P}[Y \in E^c] - \mathbb{P}[X \in E^c])$$

$$= (\mathbb{P}[X \in E] + \mathbb{P}[X \in E^c]) - (\mathbb{P}[Y \in E] + \mathbb{P}[Y \in E^c])$$

$$= 1 - 1 = 0.$$

Inequalities (1) and (2) are equalities in the case of $A = E$ and $A = E^c$, so this upper bound is actually achieved. We conclude that

$$d_{TV}(X, Y) = \mathbb{P}[X \in E] - \mathbb{P}[Y \in E] = \mathbb{P}[Y \in E^c] - \mathbb{P}[X \in E^c].$$

The desired result follows from summing:

$$\sum_{k \in \mathbb{Z}} \left| \mathbb{P}[X = k] - \mathbb{P}[Y = k] \right| = \sum_{k \in E} \left| \mathbb{P}[X = k] - \mathbb{P}[Y = k] \right| + \sum_{k \in E^c} \left| \mathbb{P}[X = k] - \mathbb{P}[Y = k] \right|$$

$$= 2 \cdot d_{TV}(X, Y).$$

$\square$

**Problem 2.** Let $X$ and $Y$ be independent random variables taking values in the positive integers and having the same distribution given by

$$\mathbb{P}[X = n] = \mathbb{P}[Y = n] = 2^{-n}$$

for all $n \in \mathbb{N}$. Find $\mathbb{P}[X \text{ divides } Y]$.

*Solution.* Let $f(x, y) = \mathbb{P}[X = x, \ Y = y]$ be the joint probability mass function and let $E$ be the event that $X$ divides $Y$. By independence and Fubini's theorem we have

$$\mathbb{P}[X \text{ divides } Y] = \sum_{x,y} f(x, y) \mathbb{1}_E(x, y)$$

$$= \sum_{x=1}^{\infty} \sum_{x|y} 2^{-x} \cdot 2^{-y}$$

$$= \sum_{x=1}^{\infty} \sum_{k=1}^{\infty} 2^{-x} \cdot 2^{-kx}$$

$$= \sum_{k=1}^{\infty} \sum_{x=1}^{\infty} 2^{-x(k+1)}$$

$$= \sum_{k=1}^{\infty} \frac{1}{2^{k+1} - 1}.$$

$\square$

**Problem 3.** A total of $n$ bar magnets are placed end to end in a line with random independent orientations. Adjacent like poles repel, ends with opposite polarities join to form blocks. Find the expected number of blocks of joint magnets.

*Solution.* Let $E_k$ be the event that there are $k$ blocks.

$$\mathbb{E}[\text{number of blocks}] = \sum_{k=1}^{n} k \cdot \mathbb{P}[\text{there are } k \text{ blocks}]$$

$$= \sum_{k=1}^{n} k \cdot \mathbb{P}[E_k].$$

Let's compute $\mathbb{P}[E_k]$. In how many ways can we arrange $k$ blocks? Such an arrangement consists of a sequence of blocks whose lengths are positive integers that sum to $n$. We can then identify block-length sequences with $k$-tuples of positive integers that sum to $n$. Each block-length sequence corresponds to exactly two magnet configurations, where each magnet in one configuration is oriented opposite to the corresponding magnet in the other.

Assuming that each magnet is oriented according to a uniform distribution, any particular configuration occurs with probability $2^{-n}$. Since each block-length sequence corresponds to exactly two magnet configurations, we have

$$\mathbb{P}[E_k] = 2^{1-n} \cdot n_k,$$

where $n_k$ is the number of block-length sequences of length $k$.

Now to find $n_k$. This is essentially the stars and bars problem. We want to count the number of ways to arrange our $n$ magnets into $k$ blocks so that each block has at least one magnet in it. This is the same as counting the number of ways to place $k-1$ "dividers" in the $n-1$ gaps between consecutive magnets. This gives

$$n_k = \binom{n-1}{k-1}.$$

Putting it all together, we have

$$\mathbb{E}[\text{number of blocks}] = 2^{1-n} \cdot \sum_{k=1}^{n} k \cdot \binom{n-1}{k-1}.$$

$\square$

**Problem 4.** Let $X$ and $Y$ be random variables with mean 0, variance 1, and covariance $\text{Cov}(X,Y) = \rho$. Show that

$$\mathbb{E}[\max(X^2, Y^2)] \leq 1 + \sqrt{1-\rho^2}.$$

*Proof.* Since $\max(u, v) = \frac{1}{2}(u + v) + \frac{1}{2}|u - v|$, we have

$$\mathbb{E}[\max(X^2, Y^2)] = \mathbb{E}\left[\frac{1}{2}(X^2 + Y^2) + \frac{1}{2}|X^2 - Y^2|\right]$$

$$= 1 + \frac{1}{2}\mathbb{E}[|X^2 - Y^2|].$$

By Cauchy-Schwarz, we have

$$\mathbb{E}[|X^2 - Y^2|] \leq \|X - Y\|_{L^2} \cdot \|X + Y\|_{L^2}$$

$$= \sqrt{2 - 2\rho} \cdot \sqrt{2 + 2\rho}$$

$$= 2\sqrt{1 - \rho^2}.$$

Substituting into the above equality, we obtain

$$\mathbb{E}[\max(X^2, Y^2)] \leq 1 + \sqrt{1 - \rho^2}.$$

Note that this is well-defined, since $|\rho| = |\text{Cov}(X, Y)| \leq \sqrt{\text{Var}[X] \cdot \text{Var}[Y]} = 1$. $\qquad\square$

**Problem 5.** A *median* of a random variable $X$ is a number $m \in \mathbb{R}$ such that

$$\mathbb{P}[X \leq m] \geq \frac{1}{2} \quad \text{and} \quad \mathbb{P}[X \geq m] \geq \frac{1}{2}.$$

(a) Prove that a median exists for every random variable $X$.

*Proof.* The idea is that continuity of measure should ensure that $\mathbb{P}[X \leq t]$ eventually "flips" from being at most $\frac{1}{2}$ to being at least $\frac{1}{2}$. To this end, let $a$ and $b$ be defined by

$$a = \inf\left\{t : \mathbb{P}[X \leq t] \geq \frac{1}{2}\right\}, \quad b = \sup\left\{t : \mathbb{P}[X \geq t] \geq \frac{1}{2}\right\}.$$

That $a$ and $b$ are finite follows from the fact that $\lim_{t \to -\infty} \mathbb{P}[X \leq t] = 0$ and $\lim_{t \to \infty} \mathbb{P}[X \geq t] = 0$. First, we claim that $a \leq b$. If there were some $c$ with $b < c < a$, then we'd have $\mathbb{P}[X \leq c] < 1/2$ since $c < a$ and $\mathbb{P}[X \geq c] < 1/2$. We'd then have $\mathbb{P}[X \leq c] + \mathbb{P}[X \geq c] < 1$, which can't happen.

Finally, we have that any $m \in [a, b]$ is a median, since

$$m \geq a \implies \mathbb{P}[X \leq m] \geq \frac{1}{2},$$

$$m \leq b \implies \mathbb{P}[X \geq m] \geq \frac{1}{2}.$$

$\qquad\square$

(b) Show that the mean $\mu$, median $m$, and variance $\sigma^2$ of a random variable $X$ satisfy

$$|\mu - m| \leq C\sigma$$

for some absolute constant $C$.

4

*Proof.* We can prove a not-so-tight version with Chebyshev's inequality. For any $C > \sqrt{2}$, we have by Chebyshev,

$$\mathbb{P}[|X - \mu| \geq C\sigma] \leq \frac{1}{C^2} < \frac{1}{2}.$$

Fix a median $m$. Then since $\mathbb{P}[X \geq m] \geq 1/2$ and $\mathbb{P}[X \geq \mu + C\sigma] < 1/2$, we must have $\mu + k\sigma \geq m$. Similarly, since $\mathbb{P}[X \leq m] \geq 1/2$ and $\mathbb{P}[X \leq \mu - k\sigma] < 1/2$, we must have $m \geq \mu - k\sigma$. Putting these together gives

$$|\mu - m| \leq C\sigma$$

for any $C > \sqrt{2}$. Taking the infimum over $C > \sqrt{2}$ gives $|\mu - m| \leq \sqrt{2}\sigma$.

$\square$

**Problem 6.** Let $X_1, \ldots, X_n$ be independent, identically distributed random variables for which $\mathbb{E}[1/X_i] < \infty$. Consider the partial sums

$$S_m = X_1 + \cdots + X_m.$$

Show that

$$\mathbb{E}[S_m/S_n] = m/n \quad \text{for all } m \leq n.$$

*Proof.*

$$
\begin{aligned}
\mathbb{E}\left[\frac{S_m}{S_n}\right] &= \sum_{i=1}^{m} \mathbb{E}\left[\frac{X_i}{X_1 + \cdots + X_n}\right] \\
&= m \cdot \mathbb{E}\left[\frac{X_1}{X_1 + \cdots + X_n}\right] \quad \text{(since the } X_k\text{'s are iid)} \\
&= \frac{m}{n} \sum_{i=1}^{n} \mathbb{E}\left[\frac{X_1 + \cdots + X_n}{X_1 + \cdots + X_n}\right] \\
&= \frac{m}{n}.
\end{aligned}
$$

$\square$

**Problem 7.** Suppose the joint density $f(x_1, \ldots, x_n)$ of random variables $X_1, \ldots, X_n$ can be represented as

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

for some measurable functions $f_i : \mathbb{R} \to [0, \infty)$. Prove that the $X_1, \ldots, X_n$ are independent.

*Proof.* Let $S = \{i_1, \ldots, i_m\} \subset [n]$ and let $B_1, \ldots, B_m$ be arbitrary Borel sets. Set $E_S = A_1 \times \cdots \times A_n$ where $A_{i_j} = B_j$ if $i_j \in S$ and $A_i = \mathbb{R}$ otherwise. We then have

$$\mathbb{P}[X_{i_1} \in B_1, \ldots, X_{i_m} \in B_m] = \int_{E_S} f(x_1, \ldots, x_n) \, dx$$

$$= \left(\int_{B_1} f_{i_1}(x_{i_1}) \, dx_{i_1}\right) \cdots \left(\int_{B_m} f_{i_m}(x_{i_m}) \, dx_{i_m}\right) \prod_{j \notin S} \left(\int f_j(x_j) \, dx_j\right). \tag{3}$$

On the other hand, if we let $E_{i_j} = A_1 \times \cdots \times A_n$ where $A_{i_j} = B_j$ and $A_i = \mathbb{R}$ otherwise, we have

$$\mathbb{P}[X_{i_1} \in B_1] \cdots \mathbb{P}[X_{i_m} \in B_m] = \left( \int_{E_{i_1}} f(x_1, \ldots, x_n) \, dx \right) \cdots \left( \int_{E_{i_m}} f(x_1, \ldots, x_n) \, dx \right)$$

$$= \left( \int_{B_1} f_{i_1}(x_{i_1}) \, dx_{i_1} \right) \left( \prod_{j \neq i_1} \int f_j(x_j) \, dx_j \right) \tag{4}$$

$$\cdots \left( \int_{B_m} f_{i_m}(x_{i_m}) \, dx_{i_m} \right) \left( \prod_{j \neq i_m} \int f_j(x_j) \, dx_j \right).$$

Now to compare the bottom-most expressions in (3) and (4). Both contain the same $\int_{B_j} f_{i_j}(x_{i_j}) \, dx_{i_j}$ terms, so it's a matter of matching up the other terms. Since $f$ is a probability distribution, we have

$$1 = \int f(x_1, \ldots, x_n) \, dx = \prod_{i=1}^{n} \left( \int f_i(x_i) \, dx_i \right).$$

Using this, the relevant terms in the last expression in (4) becomes

$$\left( \prod_{j \neq i_1} \int f_j(x_j) \, dx_j \right) \cdots \left( \prod_{j \neq i_m} \int f_j(x_j) \, dx_j \right) = \left( \prod_{j=1}^{n} \int f_j(x_j) \, dx_j \right)^{m-1} \left( \prod_{j \notin S} \int f_j(x_j) \, dx_j \right)$$

$$= \left( \prod_{j \notin S} \int f_j(x_j) \, dx_j \right),$$

which is what appears in the last expression in (3). □

**Problem 8.** Let $X, Y$ be independent random variables taking nonnegative values. Express the density of $XY$ in terms of the densities of $X$ and $Y$.

*Solution.* Let $f_X$ and $f_Y$ be the densities of $X$ and $Y$, respectively. Let's look at the distribution $\mathbb{P}[XY \leq t]$. Since $f_X$ and $f_Y$ are both integrable, we have by Fubini's theorem

$$\mathbb{P}[XY \leq t] = \int_{xy \leq t} f_X(x) f_Y(y) \, dy dx$$

$$= \int_0^\infty \int_0^{t/x} f_X(x) f_Y(y) \, dy dx.$$

Since $f_X$ and $f_Y$ are integrable, the function $\int_0^s f_X(x) f_Y(y) \, dy$ is absolutely continuous and we can apply the fundamental theorem of calculus – provided that we can differentiate under the integral with respect to $x$. We obtain the density

$$\frac{d}{dt} \mathbb{P}[XY \leq t] = \int_0^\infty \frac{1}{x} f_X(x) f_Y(t/x) \, dx.$$

□

6