# Stats 210A, Fall 2023
# **Optional** Homework 0

***Not* Due on**: Wednesday, Aug. 30

Lecture 1 included a "whirlwind tour" of measure theory at the heuristic level that we'll be using in class. Problem 1 is meant to give a little more intuition about densities and the others are meant to motivate measure-theoretic probability a bit.

Note this problem set has three problems; a typical problem set will have 5.

**1. Densities**

For a given point $x \in \mathcal{X}$, the *Dirac measure* is defined as

$$\delta_x(A) = 1\{x \in A\} = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}.$$

Essentially, $\delta_x$ is the measure that puts a unit of mass on $x$ and none anywhere else.[1] Integrals wrt $\delta_x$ are defined as $\int f(u)\, \mathrm{d}\delta_x(u) = f(x)$.

Furthermore, suppose $\mu_1$ and $\mu_2$ are both measures on $\mathcal{X}$, and $a_1, a_2 \geq 0$. You may use without proof that the sum $\nu = a_1\mu_1 + a_2\mu_2$ is also a measure, and that for "nice enough" functions,

$$\int f(x)\, \mathrm{d}\nu(x) = a_1 \int f(x)\, \mathrm{d}\mu_1(x) + a_2 \int f(x)\, \mathrm{d}\mu_2(x).$$

(a) Let $x_1, x_2, \ldots, x_n$ be integers (not necessarily all distinct), and define two measures on the set $\mathbb{Z}$ of all integers: the counting measure $\#$ from class, and the *empirical distribution*

$$\widehat{P}_n(A) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}(A).$$

That is, $\widehat{P}_n(A)$ is the fraction of points that fall into the set $A$.

**Note:** if $x_1, \ldots, x_n$ are sampled from some distribution $P$ then $\widehat{P}_n$ is a natural nonparametric estimator of the measure $P$.

Show that $\widehat{P}_n$ is absolutely continuous with respect to $\#$ but not the other way around. What is the density of $\widehat{P}_n$ with respect to $\#$? Is it possible to define a density of $\#$ with respect to $\widehat{P}_n$?

(b) For $\mathcal{X} = [0, \infty)$, define the measure $\mu(A) = \lambda(A) + \delta_0(A)$, where $\lambda$ represents the Lebesgue measure. For fixed $\theta \in \mathbb{R}$, define the random variable

$$X = \max(0, Z) \text{ where } Z \sim N(\theta, 1),$$

what is the density of $X$'s distribution with respect to $\mu$?

(c) Consider two densities $p_1$ and $p_2$ with respect to some common measure $\mu$ on a space $\mathcal{X}$ (not necessarily the same $\mu$ from part (b)). Suppose $p_1$ and $p_2$ both result in the same measure $P$ defined by $P(A) = \int 1_A(x)p_i(x)\, \mathrm{d}\mu(x)$.

[1] In a sense this is defined identically to the indicator function $1_A(x)$, but we think of one as being a function of $x$ with $A$ fixed, and the other as a function of $A$ (a measure) with $x$ fixed.

Define the set $A = \{x : \ p_1(x) \neq p_2(x)\}$, and show that $\mu(A) = 0$ (**Hint:** consider sets like

$$A_n = \left\{ x : \ p_1(x) - p_2(x) \in \left[ \frac{1}{n+1}, \ \frac{1}{n} \right) \right\}$$

for $n = 1, 2, \ldots$.)

Don't worry about whether the measure is well-defined for $A_n$ (i.e., whether these sets are measurable). They are in the sense we need them to be.

**2. A conditional probability paradox**

Let $X, Y \overset{\text{i.i.d.}}{\sim} N(0,1)$. This problem is meant to show that by carelessly conditioning on probability-zero events we can get ourselves into trouble. It is directly inspired by a calculation I personally flubbed a few years ago.

(a) Defining $S = X + Y$ and $D = X - Y$, show $S$ and $D$ are independent and conclude that

$$\mathbb{E}[X^2 + Y^2 \mid D] = D^2/2 + 1$$

(b) Now define the polar parameterization $(R, \Theta)$ with $R = \sqrt{X^2 + Y^2}$ and $\Theta \in [0, 2\pi)$ such that $X = R\cos\Theta$ and $Y = R\sin\Theta$. Show that $R$ is independent of $\Theta$ and conclude that

$$\mathbb{E}[X^2 + Y^2 \mid \Theta] = 2$$

(c) Use (a) and then (b) to find the expectation of $X^2 + Y^2$ conditional on the event $X = Y$. Can you come up with an intuitive explanation for how we could have arrived at two different answers?

**Moral:** Intuition may fail us when we condition on a measure-zero event, and in cases like this the meaning can be ambiguous and give different answers. Conditioning on a random variable, on the other hand, tends to give less ambiguous answers (there are still some ambiguities, similar to those we encounter in defining densities, but they don't really matter).

**3. Non-measurable sets**

This problem goes through a construction of a non-measurable set, meant to motivate measure theory from a real analysis perspective. It concerns the impossibility of defining "volume" for every subset of the unit interval $U = [0, 1)$.

For $x, y \in \mathbb{R}$ define the "wraparound addition" (modulo 1) as the fractional part of their sum:

$$x \oplus y = x + y - \lfloor x + y \rfloor.$$

Recall that for $x \in \mathbb{R}$ and $A \subseteq \mathbb{R}$ we define the set $x + A = \{x + a : \ a \in A\}$. Analogously, we can define

$$x \oplus A = \{x \oplus a : \ a \in A\} \subseteq U$$

Any reasonable definition of "volume" on the interval should have several properties:

(i) Additivity: $\lambda(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \lambda(A_i)$ if all $A_i \subseteq U$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$.

(ii) Translation invariance: $\lambda(x \oplus A) = \lambda(A), \ \forall x \in U, A \subseteq U$.

(iii) Interval length: $\lambda([x, y)) = y - x, \ \forall 0 \leq x \leq y \leq 1$.

Assume that some measure $\lambda$ exists which satisfies (i)–(iii) and which is defined for all subsets of $U$. We will go through several steps to derive a contradiction.

(a) Define the function $A(x)$ mapping elements of $U$ to subsets of $U$, via $A(x) = x \oplus \mathbb{Q}$, where $\mathbb{Q}$ is the set of rational numbers. Show that $\lambda(A(x)) = 0$ for any $x$.

(b) Consider the range $\mathcal{R}_A = \{A(x) : \ x \in U\}$. Show that $\mathcal{R}_A$ is a collection of uncountably many subsets of $U$, all of which are disjoint from each other. That is, show that for any $x, y \in U$, we have either $A(x) = A(y)$ or $A(x) \cap A(y) = \emptyset$.

(c) Now, let $B \subseteq U$ denote a new set, which we construct by selecting a *single element* from each set $R \in \mathcal{R}_A$ (it doesn't matter which element; note this step uses the axiom of choice.)

Define a new function $C(x) = x \oplus B$ and define $\mathcal{R}_C = \{C(x) : \ x \in \mathbb{Q}\}$. Show that $\mathcal{R}_C$ is a collection of *countably* many subsets of $U$, all of which are disjoint from each other, and whose union is $U$.

(d) Show that no matter what value $\lambda(B)$ takes, $\lambda$ will have to violate one of the properties (i)–(iii) (**Hint:** what does the value of $\lambda(B)$ imply about $\lambda(U)$?)

Because the Lebesgue measure satisfies properties (i)–(iii), it follows that $\lambda$ must not be defined for every subset of $U$.

**Moral:** One motivation (but not the only motivation) for the idea of a $\sigma$-field is to exclude pathological counterexamples like this.