Statistics 210A Lecture 1 Notes

Daniel Raban

August 26, 2021

1 Basics of Measure Theory

1.1 Motivation for measure theory

Suppose $X \sim N(0,1)$ has a Normal distribution. We have a probability distribution $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. We can evaluate the expectation of a function by

$$\mathbb{E}[f(X)] = \int f(x)\phi(x)\,dx.$$

If we have a Binomial random variable $Y \sim \text{Binom}(n,p)$, we can write the expectation as

$$\mathbb{E}[f(Y)] = \sum_{y=0}^{n} f(y)q(y).$$

If we have $Z = \max(X,0)$ and we want to calculate the expectation, we could do

$$\mathbb{E}[f(Z)] = \frac{1}{2} f(0) + \int_0^\infty f(z)\phi(z)\,dz.$$

In general, we could have probability distributions on other sets, such as orthogonal matrices. We want a consistent notation for all of these situations. We want to be able to write something like $\int f(x)\,dP(x)$.

Especially in applied contexts, you may never need the full power of measure theory, but measure theory helps us understand what it means to, for example, condition on an event (and whether we can do so with probability 0 events).

1.2 Measures

Measure theory is a rigorous grounding for probability theory.

Definition 1.1. Let $\mathcal{X}$ be a set. A **measure** μ maps subsets $A \subseteq \mathcal{X}$ to non-negative numbers: $\mu(A) \in [0,\infty]$.

Example 1.1. Let $\mathcal{X}$ be countable (e.g. $\mathcal{X} = \mathbb{Z}$). The **counting measure** is $\#(A) = \#$ points in A.

Example 1.2. Let $\mathcal{X} = \mathbb{R}^n$. **Lebesgue measure** is the usual notion of volume: $\lambda(A) = \int \cdots \int_A dx_1 \cdots dx_n$. It is not actually possible to assign a measure to every subset of $\mathbb{R}^n$; it is possible to construct such a non-measurable set using the axiom of choice.

Example 1.3. The standard Gaussian distribution is a measure. If $Z \sim N(0, 1)$ and $X = \mathbb{R}$, we can make a measure via $P(A) = \mathbb{P}(Z \in A) = \int_A \phi(x)\, dx$.

In general, the domain of a measure μ is a collection of subsets of $\mathcal{X}$: $\mathcal{F} \subseteq 2^{\mathcal{X}}$. Such a collection is called a **σ-field** and satisfies certain properties.

Example 1.4. If $\mathcal{X}$ is countable, such as with counting measure, then we can take $\mathcal{F} = 2^{\mathcal{X}}$.

Example 1.5. If $\mathcal{X} = \mathbb{R}^n$, we can take $\mathcal{F}$ to be the Borel σ-field, which is the σ-field you get if you want to be able to measure all the open subsets of $\mathbb{R}^n$.

Definition 1.2. Given a **measurable space** $(X, \mathcal{F})$, a **measure** is a map $\mu : \mathcal{F} \to [0, \infty]$ with $\mu(\bigcup_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i)$ for disjoint A_i.

Notice that measures can take infinite values, such as $\#(\mathbb{Z}) = \infty$.

Definition 1.3. A **probability measure** is a measure μ with $\mu(\mathcal{X}) = 1$.

1.3 Integration with respect to measures

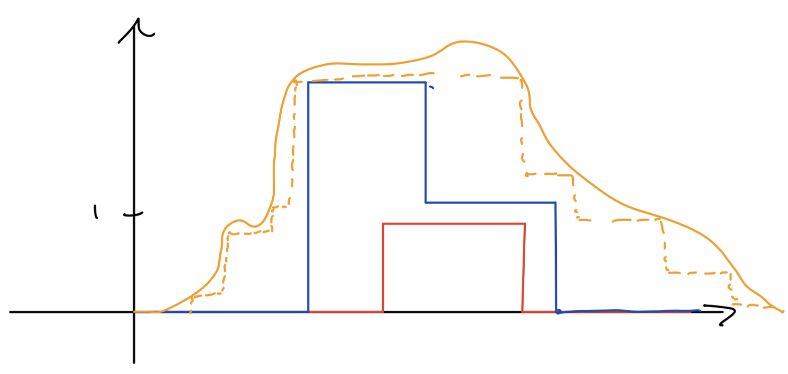
We want to be able to talk about what $\int f(x)\, d\mu(x)$ means for a nice enough function f. Define

$$\int \mathbb{1}_{\{x \in A\}}\, d\mu(x) = \mu(A).$$

Extend this definition to other f by linearity and limits:

$$\int \sum_{j=1}^n c_i \mathbb{1}_{\{x \in A_i\}}\, d\mu(x) = \sum_{j=1}^n c_i \mu(A_i).$$

With functions like this, we can approximate a wide class of functions:



Example 1.6. With counting measure,

$$\int f \, d\# = \sum_{x \in \mathcal{X}} f(x).$$

Example 1.7. With Lebesgue measure,

$$\int f \, d\lambda = \int \cdots \int f(x) \, dx_1 \cdots dx_n.$$

Example 1.8. With the Gaussian distribution,

$$\int f \, dP = \int f \phi \, dx.$$

With a discrete distribution, we cannot write it as a density with respect to Lebesgue measure. When do we have a density?

1.4 Densities

Definition 1.4. Given $(X, \mathcal{F})$ and two measures P, μ, we say P is **absolutely continuous with respect to** μ (denoted $P \ll \mu$) if $\mu(A) = 0 \implies P(A) = 0$. We also say that μ **dominates** P.

If $P \ll \mu$, then (under mild conditions) we can define the density function $p : \mathcal{X} \to [0.\infty)$ with

$$P(A) = \int_A p(x) \, d\mu(x) = \int \mathbb{1}_{\{x \in A\}} p(x) \, d\mu(x).$$

We can also write

$$\int f(x) \, dP(x) = \int f(x) p(x) \, d\mu(x).$$

The theorem which allows us to do this is the *Radon-Nikodym theorem*. The common notation is $p(x) = \frac{dP}{d\mu}(x)$, where the density is referred to as a **Radon-Nikodym derivative**.

Example 1.9. If P is a probability distribution and μ is Lebesgue measure, we call $p(x)$ a **probability density function (pdf)**.

Example 1.10. If P is a probability distribution and μ is counting measure, we call $p(x)$ a **probability mass function (pmf)**.

1.5 Probability spaces and random variables

How would we talk about an expression like $\mathbb{P}(\frac{X_1 \cdots X_n}{Y} \geq W_{\widehat{\Theta}_{\text{MLE}}})$?

Denote Ω as the **outcome space** and $\omega \in \Omega$ as an **outcome variable**. $A \subseteq \Omega$ is called an **event**, and $\mathbb{P}(A)$ is the *probability of* A. All of these elements together are called a **probability space** $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.5. A **random variable** is a function $X : \Omega \to \mathcal{X}$.

With this definition, we can say things like $X(\omega) = 5$. So Ω contains all the randomness, and if we knew ω, then we would know the value of X.

Definition 1.6. We say that X has **distribution** Q $(X \sim Q)$ if

$$\mathbb{P}(X \in B) = \mathbb{P}(\{\omega : X(\omega) \in B\}) = Q(B).$$

The **expectation** is then

$$\mathbb{E}[f(X, Y)] = \int_\Omega f(X(\omega), Y(\omega))\, d\mathbb{P}(\omega).$$

In practice, we will still calculate expectations as usual.