Notes 3 for Honors Probability and Statistics

Ernie Croot

September 12, 2003

1 Measure Theory and Integration

Now that we have a measure on $B(\mathbb{R})$, namely the Borel-Lebesgue measure, we can try to see what measure it assigns to certain sets. The main tool we will use is the Monotone Convergence Theorem for Sets in one of the following two forms:

MCT1. Suppose that Σ is a σ-algebra on a set S, and that $\mu : \Sigma \to [0, \infty]$ is a measure. If $A_1, A_2, ...$ are subsets in Σ satisfying $A_1 \subseteq A_2 \subseteq \cdots$ then we have that if $A = \cup_j A_j$,

$$\lim_{k\to\infty} \mu(A_k) \ = \ \mu(A).$$

Note: We had proved this before for probability measures, but it holds for measures in general.

MCT2. Suppose S, Σ, and μ are as above. Given $A_1 \supseteq A_2 \supseteq A_3 \cdots$, all lying in Σ, let A be their intersection. Then, we have

$$\lim_{k\to\infty} \mu(A_k) \ = \ \mu(A).$$

Note: This is a homework problem you were asked to solve, except that here μ is a general measure, not just a probability measure.

Now, suppose that μ is the Lebesgue measure on $B(\mathbb{R})$. Then, we have the following two basic observations:

I. For any $a \in \mathbb{R}$, $\mu(\{a\}) = 0$. To see this, we apply MCT2: Let A_j be the open interval $(a - 1/j, a + 1/j)$. Then, $\{a\} = \cap_j A_j$. So,

$$\mu(\{a\}) \ = \ \lim_{j\to\infty} \mu(A_j) \ = \ \lim_{j\to\infty} \frac{2}{j} \ = \ 0.$$

II. As a corollary of I we get that any countable subset of $\mathbb{R}$ has measure 0 (why?).

Here is a question that leads to a deeper appreciation of just how special the Lebesgue measure is:

Question: Must we have that if λ is a measure on $B(\mathbb{R})$ then $\lambda(\{a\}) = 0$?

It turns out that the answer is NO; that is, there do exist strange measures on $B(\mathbb{R})$ which assign singletons non-zero measures. A good example of such a measure is the following one: Suppose that $A \in B(\mathbb{R})$. Then we have that

$$\lambda(A) \;=\; \begin{cases} 1, & \text{if } 0 \in A; \\ 0, & \text{if } 0 \notin A. \end{cases}$$

It is easy to check that this is a measure on $B(\mathbb{R})$, and it is obvious that $\lambda(\{0\}) = 1$.

The natrual progression of our discussion of the Lebesgue measure at this point would be to define the Lebesgue integral, and then state and prove various theorems about it. However, to do this task properly would take too much time, and since we only need the most basic properties of the Lebesgue integral for this course, I will only say here a few things about it. First, let me say that another way of writing the Lebesgue measure of a subset $A \in B(\mathbb{R})$ is as

$$\mu(A) \;=\; \int_A \mu(dx) \;=\; \int_{\mathbb{R}} 1_A(x)dx, \tag{1}$$

where $1_A(x)$ takes the value 1 if $x \in A$, and takes the value 0 if $x \notin A$. This function $1_A(x)$ is sometimes called the indicator function for the set A. I should say here that rather than thinking of this as another way of writing $\mu(A)$, you should think of it as the "defining property" of Lebesgue integrals over indicator functions. The Lebesgue integral over general functions is then defined as a certain limit of integrals over finite linear combinations of indicator functions.

The right-most integral may be interpreted as

$$\int_{\mathbb{R}} 1_A(x)dx \;=\; \int_{-\infty}^{\infty} 1_A(x)dx. \tag{2}$$

Now, if A is just a finite union of intervals, then we could evaluate the integral on the right hand side here quite easily using the standard Riemann integral from Calculus. For example, suppose that $A = (0,1) \cup (2,3)$. Then, the right-most integral would be

$$\int_0^3 1_A(x)dx \;=\; \int_0^1 1dx \;+\; \int_2^3 1dx \;=\; 1+1 \;=\; 2.$$

If, on the other hand, A is a really gnarly set, constructed using infinite set operations, then we may not be able to evaluate it using the Riemann integral. A classic example is when we let A be the set of rationals in $[0,1]$. Then, by manipulating (1) and (2) on a purely formal level we should have that

$$\int_{-\infty}^{\infty} 1_A(x)dx \;=\; \int_0^1 1_A(x)dx \;=\; \mu(A) \;=\; 0, \tag{3}$$

since A is countable. This equation in some ways reflects our common sense notion that the rationals are a "very thin", albeit "dense", subset of the reals in $[0,1]$.

Now, it is obvious that $1_A(x)$ is not Riemann-integrable, because the upper and lower sum estimates don't converge to one another. What do I mean here? Suppose we partition the interval $[0,1]$ into n subintervals $I_j = [(j-1)/n, j/n]$, $j = 1,2,...,n$, and we let m_j be the minimum value of $1_A(x)$ on I_j and let M_j be the maximum value of $1_A(x)$ on I_j. Then, if $1_A(x)$ *were* Riemann-integrable we would have to have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} m_j \;=\; \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} M_j;$$

however, since every interval I_j contains infinitely many rationals and irrationals (non-rationals), we have that $m_j = 0$ and $M_j = 1$. So, the sum on the left-hand-side is always 0 and the sum on the right-hand-side is always 1, no matter how large we take n to be. Thus, the two limits are not equal, and $1_A(x)$ is therefore *not* Riemann-integrable.

In everything we'll ever do, we will not need to evaluate such integrals (via Lebesgue integration), so henceforth we will only make use of the Riemann integral. That said, there is one more result that I want to mention before we leave integration theory altogether:

Radon-Nikodym Theorem. Suppose that α is a measure on $B(\mathbb{R})$, and let μ be the Lebesgue measure on $B(\mathbb{R})$. Then, we have that if the following condition holds

$$\text{for every } A \in B(\mathbb{R}), \quad \mu(A) = 0 \text{ if and only if } \alpha(A) = 0,$$

then there exists a Lebesgue-integrable function $f(x) \geq 0$ (actually, $f(x)$ needs to also be what is called an L^1 function, but we will not worry about this extra assumption) such that α can be represented as

$$\alpha(B) \ = \ \int_B f(x)\mu(dx) \ = \ \int_{-\infty}^{\infty} f(x) 1_B(x) dx.$$

NOTE: There are more more general formulations of the Radon-Nikodym theorem than what I stated here.

The first thing I want to point out that the condition which says that α and μ "agree on measure zero sets" is actually necessary; that is, it is easy to produce examples where the conclusion of the theorem is false if we don't assume this condition. One example is our measure λ stated previously. This measure clearly doesn't agree with μ on measure 0 sets, since for example,

$$\lambda(\{0\}) \ = \ 1 \ \neq \ 0 \ = \ \mu(\{0\}).$$

The measure λ also fails to satisfy the conclusion of the theorem; of course, to properly see this we would need to understand how the Lebesgue integral works, which I said I wasn't going to explain. Basically, one can see that if there were such a function $f(x) \geq 0$, then for $B = \{0\}$ we would have to have

$$\begin{aligned} 1 \ &= \ \lambda(\{0\}) \ = \ \int_{-\infty}^{\infty} f(x) 1_B(x) dx \\ &= \ \int_0^0 f(x) dx \ = \ 0. \end{aligned}$$

The second thing I want to point out here is that the theorem kind of suggests that it might be possible to construct other measures by starting with a function $f(x) \geq 0$ which is Lebesgue-integrable, and then defining the measure through integration. This brings me to a question which was asked in class, namely:

Question: Suppose that one selects a real number at random. What is the probability that the number is positive?

The problem with this question comes in when you try to decide "with respect to which probability measure?" The sort of measure you would *want* to have here, namely one which weights all length L intervals $(a, a+L)$ equally, doesn't exist. That is to say

> For any $L > 0$ there does not exist a probability measure on $B(\mathbb{R})$ which assigns all the intervals $(a, a+L)$ the same value $P((0, L))$.

So, the question is ill-posed. However, if we impose a genuine probability measure at the outset, then it can be solved: First, suppose that

$$f(x) \;=\; \frac{e^{-x^2}}{\sqrt{\pi}}.$$

Then, a standard fact is that

$$\int_{-\infty}^{\infty} f(x)dx \;=\; 1.$$

Now, using Lebesgue integration we can define a probability measure α on $B(\mathbb{R})$ as follows:

$$\alpha(A) \;=\; \int_A f(x)\mu(dx) \;=\; \int_{-\infty}^{\infty} f(x)1_A(x)dx.$$

Notice here that $\alpha(\mathbb{R}) = 1$, which is required in order for α to be a probability measure. If you knew more about the Lebesgue integral you could check that α is actually a measure. Now, let us consider the following variation on the above question:

Question. Suppose that one picks a random number in $\mathbb{R}$ with respect to the measure α. What is the probability that that number was positive?

The event we are asking for the probability measure of is $\mathbb{R}^+$. So, we are asking for $\alpha(\mathbb{R}^+)$, which is

$$\int_{-\infty}^{\infty} f(x)1_{\mathbb{R}^+}(x)dx \;=\; \int_0^{\infty} f(x)dx \;=\; \frac{1}{2}.$$

So, the answer is $1/2$.

Using the language of measures of events in $B(\mathbb{R})$ to describe random events or processes may seem a little wierd, and later we will show how to recast all these questions in terms of "random variables", "distribution functions", and "density functions", which is a more natural context for these types of problems.