Measure Theory

Mark Dean

Lecture Notes for Fall 2015 PhD Class in Decision Theory - Brown University

1 Introduction

Next, we have an extremely rapid introduction to measure theory. Given the short time that we have to spend on this, we are really only going to be able to introduce the relevant concepts, and try to give an idea of why they are important. For more information, Efe Ok has an as-yet unpublished book available online here https://files.nyu.edu/eo1/public/ that covers the basics, and points the way to many other classic texts. For those of you who are thinking of taking decision theory more seriously, then a taking a course in measure theory is probably advisable.

Take some underlying set X. Measure theory is the study of functions that map subsets of X into the real line, with the interpretation that this number is the 'measure' or the 'size' or the 'volume' of that set. Of course, not every function f defined on a subset of 2^X is going to fit into our intuitive notion of how a measure should behave. In particular, we would like a measure to have at least the following two properties:

1. $f(A)\backslash \geq 0$
2. $f(\cup_{I=1}^{\infty} A_i) = \sum_{i=1}^{\infty} A_i$ For any pairwise disjoint sets A_i

The first property says that we can't have negatively measured sets, the second says, if we look at the measure of the union of disjoint sets should be equal to the sum of the measure of those sets. Both these properties should seem appealing if we think about what we intuitively mean by 'measure'.

Primarily, we are going to be interested in measure theory as a basis for probability. We will think of X as describing the states of the world, and the 'measure' of a set as the probability of an event in this set occuring. However, measure theory is much more general than that. For example, if we think about intervals on the real line, the natural measure is the length of those intervals (i.e. , for $[a, b]$, the measure is $b - a$.). The measure that results from this proposition is called the Lesbesgue measure, and is one of the ways we can make formal concepts such as integration.

2 $\sigma-$Algebras

The first question we need to ask is the following: what domain do we want to apply our measures to? In other words, what subsets of 2^X would we like to allow into the domain of our function f. In the case of probability measures, then we might find an arbitrary collection of 2^X a bit unsatisfactory. In particular, we might want the following properties to hold (where X is some non empty set, and $\emptyset \notin \Sigma \subset 2^X$)

1. if $A \in \Sigma$, then $X/A \in 2^X$
2. if $A, B \in \Sigma$ then $A \cup B \in \Sigma$

The first property allows us to say that, for any event A to which we will assign a measure, we can also assign the probability of NOT A. The second says that, for any two events A and B to which we can assign a measure, we can assign a measure to A and B.

Any collection Σ that obeys these properties is called an algebra. In fact, we are going to require slightly more, specifically that, for any countable collection $\{A_i\}_{i=1}^{\infty} \in \sum^{\infty}$, we would like $\cup_{i=1}^{\infty} A_i \in \Sigma$. In other words, we would like Σ to be closed under countable unions. If Σ has this property, then we call it a $\sigma-$algebra.

This may sound a little too much like unnecessary hard work: Why don't we just demand that our measure is defined on 2^X for any set? It turns out that this can lead us into difficulties. For example, it is not possible to apply the Lesbegue measure to every subset of the real line: that is, there is no function that satisfies the properties listed above, is defined on $2^{\mathbb{R}}$ and is equal to the length of intervals. Given that we have to give up on something, the most natural thing is to allow for the possibility of non-measureable sets - i.e. to have our measure defined on something smaller than the power set.[1]

Two properties that s$\otimes$em immediately from the definition of a $\sigma-$algebra Σ on a set X are as follows:

1. $\emptyset \in \Sigma$ and $X \in \Sigma$. The latter property comes from the fact that Σ is non empty, so contains some set A, and therefore contains X/A, and so X. This in turn implies that $\emptyset \in \sum$
2. Σ is closed under countable interstections. This follows from the fact that $A_1 \cap A_2 = X/(X/A_1 \cup X/A_2)$

$\sigma-$Algebras are tricky beasts when one is dealing with infinite base sets (you should convince yourself that for finite base sets they are the same as algebras). In fact, it is generally not possible to precicely characterise what sets are in an algebra, and which are not. For this reason, we tend to start by thinking about the events that we would like to measure, and simply define the σ algebra generated by these events in the following sense:

Definition 1 *Let X be a non-empty set, and $\mathcal{A}$ be a non-empty subset of 2^X. We call Σ the $\sigma-$algebra generated by $\mathcal{A}$ if*

1. $\mathcal{A} \subset \Sigma$

2. For any other $\sigma-$algebra Σ' such that $\mathcal{A} \subset \Sigma'$, we have $\Sigma' \subset \Sigma$

we write $\Sigma(\mathcal{A})$ to denote such an algebra

It is not obvious that every collection $\mathcal{A}$ should generate a $\sigma-$algebra, but in fact it is true. You can prove this yourself using tricks that we have seen before (hint - is an arbitrary intersection of $\sigma-$algebras itself a $\sigma-$algebra?)

One very common $\sigma-$algebra for us to work with is that generated by all the open sets in some metric space. This is called the **Borel** $\sigma-$algebra. Let X be some metric space and $\mathcal{O}_X$ be all the

[1] The Banach-Tarski Paradox that we discussed before is actually an example of the impossibility of measuring every set in three dimensional Euclidian spaces.

open sets on X, we denote the Borel $\sigma-$algebra on X as $\mathcal{B}(X) = \Sigma(\mathcal{O}_X)$. Note, that, if we are working the real line, there are many alternative ways of characterizing the the Borel sets. The following collections of $2^{\mathbb{R}}$ all generate the same $\sigma-$algebras

1. $\mathcal{O}_{\mathbb{R}}$
2. All closed and bounded intervals on $\mathbb{R}$
3. The set of all closed sets on $\mathbb{R}$
4. The set of all open and bounded intervals on $\mathbb{R}$

These equivalences also point out the difference between an algebra and a σ-algebra: If we think of all the right-closed intervals on $\mathbb{R}$, then (a, b) would have to be contained in any $\sigma-$algerbra that contains these sets, but not necessarily in an algebra that contains them.

Given our previous discussion, it should come as no suprise that, while the Borel $\sigma-$algebra on $\mathbb{R}$ is 'large', it is not equivalent to $2^{\mathbb{R}}$. This means that there are some subsets of $\mathbb{R}$ that are not Borel sets. (there are examples, but without taking a lot more time, they won't give you much intuition) Notice that any measure that is defined on the Borel sets would not take a value for sets that are not included in the Borel $\sigma-$algebra. For this reason, these sets are called non-Borel-measureable (under the Borel $\sigma-$algebra). More generally, note that the property of measurability or otherwise is defined relative to an underlying $\sigma-$algebra.

3 Probability Measures

Now that we have defined our domain, we are in position to define what we mean by a probability measure: this is a function that is going to assign probabilities to each of the events in our $\sigma-$algebra. Let X be a non-empty set, Σ be a $\sigma-$algebra on X and $p : \Sigma \to \mathbb{R}$. Here are some definitions:

Definition 2 *Here are the definitions of some properties of p*

1. *If, for any $\{A_i\}_{i=1}^n \in \Sigma^n$ that is pairwise disjoint, $p(\cup_{i=1}^n A_i) = \sum_{i=1}^n p(A_i)$ then we say p is finitely additive*

2. If, for any $\{A_i\}_{i=1}^{\infty} \in \Sigma^{\infty}$ that is pairwise disjoint, $p(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} p(A_i)$ then we say p is $\sigma-$additive

Definition 3 *If $p : \Sigma \to \bar{\mathbb{R}}_+$ is $\sigma-$additive, and $p(\emptyset)$ then it is called a measure.*

1. If p is a measure such that $p(X) < \infty$, then it is a finite measure

2. If $p(X) = 1$ then it is a probability measure

You should convince yourself that the properties of a probability measure are enough to make it *behave* like we would like a probability measure to behave. For example, you should convince yourself of the fact that that a probability measure has the property of monotonicity:

Claim 1 *Let $\{X, \Sigma, p\}$ be a probability space and $A, B \in \Sigma$, such that $A \subseteq B$. Then $p(A) \leq p(B)$*

Does this imply that, if A is a strict subset of B that $p(A) < p(B)$? Does it imply that the the probability of the subset of a zero probability event is necessarily zero?

You will play around with some of the properties of probability measures for homework. However, one result that you should know is that probability measures satisfy Boole's inequality

Boole's Inequality For any probability space $\{X, \Sigma, p\}$

$$p\left\{\cup_{i=1}^{\infty} A_i\right\} \leq \sum_{i=1}^{\infty} p(A_i) \text{ for any } \{A_m\} \in \Sigma^{\infty}$$

4 Constructing Probability Spaces

As we mentioned above, σ-algebras are tricky things to work with, while algebras are much simpler. Luckily there is an extremely powerful result that allows us to use algebras to construct our probability measures, then assume the *existance* of an extenision to that probability measure to the $\sigma-$algebra generated by that algebra. This is Carathedory's Extension Theorem.

Theorem 2 (Carathedory) *Let $\mathcal{A}$ be an algebra on some non-empty X and $q : \mathcal{A} \to \mathbb{R}_+$. IF q is $\sigma-$additive on $\mathcal{A}$, then there exists a measure p on $\Sigma(\mathcal{A})$ such that $p(A) = q(A)$ for all $\mathcal{A}$. Moreover, if q is finite, then p is unique*

Thus, we can uniquely identify a probability measure on a $\sigma-$algebra by describing behavior on an algebra that generates that σ-algebra.

We can see the power of these result when defining the Lebesgue-Stieltjes Probability measure on $\mathbb{R}$. In order to do so, we first need to define a distribution function:

Definition 4 *A map $F : \mathbb{R} \to [0, 1]$ is said to be a distribution function if it is increasing, right continuous and $F(-\infty) = 0 = 1 - F(\infty)$*

You have been dealing with distribution functions for a long time: these are just the CDF functions standard in statistics. One question we might want to know is: what is the relationship between distribution functions and probability measures? In fact, it turns out that, using Carathedory, we can show that each distribution function induces a unique probability measure on the Borel σ-algebra of $\mathbb{R}$. To see this, let F be a distribution function, $\mathcal{A}$ be the algebra generated by the right-closed intervals on $\mathbb{R}$ and define the following q on $\mathcal{A}$:

$$\begin{aligned}
\text{if } -\infty &\leq a \leq b < \infty, \text{ then } q((a, b]) = F(b) - F(a) \\
\text{if } -\infty &\leq a, \text{ then } q((a, \infty)) = 1 - F(a) \\
&\text{if } A_1 ... A_m \text{ are finitely many disjoint intervals in } \mathcal{A} \text{ then} \\
q\left(\cup_{i=1}^{m} A_i\right) &= \sum_{i=1}^{m} q(A_i)
\end{aligned}$$

Thus we can use F to define a probability measure on $\mathcal{A}$. But what if we want to extend this measure to the Borel sets of $\mathbb{R}$. How do we know that such an extension exist? And if there exists more than one extension, how do we know which one to choose? Luckily, Carathedory means that we don't have to worry about these things: a unique extension exists *as long as we can show that q is $\sigma-$additive on $\mathcal{A}$*. Luckily it is, which you can show for homework.

So any distribution function defines a unique probability measure on the borel sets of $\mathbb{R}$. Interestingly, the converse is true: any probabilitty measure p on the borel sets of $\mathbb{R}$ defines a probability measure as $f(x) = p((-\infty, a])$. Thus, there is a tight relationship between probability measures and distribution functions.

We can use this method (with a few technical tweeks) to define the Lesbegue measure on the real lines. This is a measure l that assigns to each interval the length of that interval, so $l((a, b]) = b - a$.

This is the standard way of measuring the real line. It is worth noting that the lesbegue measure has some potentially surprising properties. First, (unsurprsingly), the measure of any singleton is zero, as

$$l(\{a\}) = l\left(\cap_{i=1}^{\infty}(a - \frac{1}{m}, a]\right) = \lim_{m\to\infty} l((a - \frac{1}{m}, a]) = \lim_{m\to\infty} \frac{1}{m} = 0$$

Perhaps more surprisingly, the σ-additivity of l therefore implies that any countable set has measure zero (as any countable set is the countable union of singletons). Thus, for example, $\mathbb{Q}$ is measure-zero. (Note, this doesn't mean that there is a 1-1 relationship between countability and non-zero measure-ness: there are measure zero uncountable sets as well, though they are weird).

5 Random Variables and Expectations

Next, we define the concept of a random variable, and through it, the concept of an expectation.

Consider the experiment of rolling two dice. Imagine further that what you are interested in is the sum of the numbers on the two dice. How could we talk about the probabilities of various different sums? One way would be to construct a probability space $\{X, \Sigma, p\}$, wher X is the natural number between 2 and 12 and p is the probabilities of events in this probability space generated by rolling the two die. However, it seems that this is somewhat ineffcient. After all, the underlying event here is rolling the two dice. Surely it would be nicer to use this underlying probability space which then *generates* probabilites over the numbers $\{2, ...12\}$. Apart from anything else, this would save us from having to generate a new probability space for every different way that we would like to combine the numbers on the two dice (for example, the product of the two numbers, or the number on die 1 minus the number on die 2). It is this excerice that leads us to the concept of a random variable.

Definition 5 *Let (X, Σ) be a measure space, and Y be a metric space. Let x be a map from X to Y such that $x^{-1}(B) \in \Sigma$ for any Borel set B in Y. Then we say that x is a $Y-$valued random function.*

We also describe x as a Σ-measurable function Note the requirement that is being made here: If we look at any event A in the Borel sets in Y (for example, any open set in Y) then the set of elements in x that map to Y must be measurable according to Σ. In fact, it turns out that there are some short cuts that can help us check measurability

Remark 1 *Let (X, Σ) be a measure space and $x : X \to Y$ for some metric space Y. Then x is a Y-valued random variable if and only if*

1. *$x^{-1}(O) \in \Sigma$ for any open set O*
2. *$x^{-1}(S) \in \Sigma$ for any closed set S*
3. *If $Y = \mathbb{R}$, then $\{w \in X | x(w) \leq \alpha\} \in \Sigma$ for any $\alpha \in \mathbb{R}$*

Before going further we need the following definition:

Definition 6 *Let Y be a metric space. A $Y-$valued random variable is called discrete if its range is a countable set, and simple if its range is finite*

As an example, consider the family of indicator functions.

Example 1 *Let (X, Σ) be a measurable space. For any event $S \in \Sigma$, we define the indicator function 1_s as*

$$\begin{aligned} 1_s(w) &= 1 \textit{ if } w \in S \\ &= 0 \textit{ otherwise} \end{aligned}$$

Clearly any indicator function is measurable (why?). Moreover, the set of simple random variables on (X, Σ, p) is identicle to the set of functions defined by

$$a_1 1_{A_1} + ... + a_n 1_{A_n}$$

for sequences $\{a_i\}_{i=1}^n \in \mathbb{R}^n$ and $\{A_i\}_{i=1}^n \in \Sigma^n$

It is worth noting that there is a one way relationship between measurability and continuity, in that continuous functions are measureable, but not necessarily visa versa:

Lemma 1 *Let X and Y be two metric spaces, and $x : X \rightarrow Y$ be continuous at all but countably many points. The x is a Y-valued random variable on the borel sets of X. However, there are also functions that are continuous nowhere which are also measurable*

One further useful property of measurability is that it is preserved by the act of continuously combining random variables

Remark 2 *Let Y be a seperable metric space, Z be a metric space and x and y to $Y-$valued random variables on (X, Σ). then, if $\theta : Y \times Y \rightarrow Z$ is a continuous map, then the function defined by*

$$\begin{aligned} \bar{\theta} &: X \rightarrow Z \\ \textit{such that } \bar{\theta}(t) &= \theta(x(t), y(t)) \end{aligned}$$

is a random variable on (X, Σ).

Thus, the sum, product, max and min of random variables are also themselves random variables.

With these interesting asides out of the way, we can now define the distribution of a random variable x on (X, Σ). By the distribution, we mean the probability that a random variable falls in a particular range S Obviously, what we would like to do is to assign the probability of the underlying events that give rise to S, i.e. $p(x^{-1}(S))$. The fact that we demand that the random variable be measurable is exactly the condition that we can do this for any Borel set S.

Definition 7 *Let Y be a a metric space.. A $Y-$valued random variable on a probability space (X, Σ, p) induces a Borel probability measure on Y as follows:*

$$p_x(S) = p(x^{-1}(S) \textit{ for every } S \in B(Y)$$

This is the distribution of x. If $Y = \mathbb{R}$, then we call this the distribution function of x.

It is easy to check that $(Y, B(Y), p_x)$ form a probability space

We can now extend the equivalence result that we stated earlier:

Remark 3 *There is a one to one correspondance between the following concepts*

1. *Borel probability measures on* $\mathbb{R}$

2. *Distribution functions*

3. *random variables on* $((0,1), B(0,1), l)$

One further useful definition is the concept of two random variables being 'almost surely' equal. Consider an experiment X that has three outcomes:

1. A coin lands heads

2. A coin lands tails

3. The coin rolls in a pattern which, if recorded, would be a proof of the Reimann hypothesis in ancient Aramaic.

and conider the random variables

$$\begin{aligned} y(t) &= 1 \text{ if } x = 1 \\ &= 2 \text{ if } x = 2 \\ &= 3 \text{ if } x = 3 \end{aligned}$$

and

$$\begin{aligned} z(t) &= 1 \text{ if } x = 1 \\ &= 2 \text{ if } x = 2 \\ &= 3 \text{ if } x = 100 \end{aligned}$$

These two variables are clearly not identical, but, if we think the probability of 3 is zero, then they are clearly not importantly different in some sense. This is the concept of almost sure equality:

Definition 8 *Two random variables* x *and* y *on a probability space* (X, Σ, p) *are said to be equal almost surely if*

$$p\left\{w \in X | x(w) = y(w)\right\} = 1$$

We write $x =_{as} y$. The concept of $x \geq_{as} y$ is defined analogously - i.e. the sets for which $x\,(t) \geq y(t)$ are measure 1.

Next we will define the concept of an expectation. Formally, this is just the weighted average of a random variable, with the weights determined by its probabilities. While we want to describe the expectation operator for any random variable, we will start simply, with simple random variables. As we have already shown, if (X, Σ, p) is a simple random variable, we can define any simple random variable as

$$\sum_{a \in x(X)} a 1_{x^{-1}(a)}$$

In such cases, we can define the idea of expectation relatively simply

$$E(x) = \sum_{a \in x(X)} ap(1_{x^{-1}(a)})$$

You should check, but the expectations operator defined in this way has all the nice properties that we would expect, such as:

1. Linearity, i.e.$E(\alpha x + y) = \alpha E(x) + E(y)$
2. $E(x) \geq E(y)$ if $x \geq y$ almost surely

Of course, this is not particularly helpful on its own. In order to extend this definition to non-negative random variables, we do the following:

Definition 9 *Let x be an $\bar{\mathbb{R}}$ valued random variable on (X, Σ, p) such that $x \geq 0$. The expectation of x is defined*

$$E(x) = \sup \{E(z) | z = \mathcal{L}(x)\}$$

where $\mathcal{L}(x)$ is the set of all simple random variables on X such that $z \leq x$

This notion is clearly similar to that of an integral, where the weights put on any rectangle is not the length of that rectangle, but its probability. This is actually how the Lesbegue interval is defined

$$\int_X x dp = E(x)$$

Note that we can say something about the relative expectations of random variables by knowing about their almost sure properties:

Remark 4 *Let x and y be two random variables on (X, Σ, p)*

1. *If $x \geq_{a.s.} y$ then $E(x) \geq E(y)$*
2. *If $x =_{a.s.} y$ then $E(x) = E(y)$*

It is also true, but not easy to show, that the linearity properties of simple random variables extend to arbitrary positive random variables.

One final order of business is to extend these results to arbitrary random variables. To do this, we essentially use a trick. Let x be some arbitrary random variable, and define the following random variables

$$\begin{aligned} x_+ &= \max(x, 0) \\ x_- &= \max(-x, 0) \end{aligned}$$

These are now two positive random variables, so the following is well defined

$$E(x) = E(x_+) - E(x_-)$$

6 Weak Convergence

The aim of this final section is to discuss how to put a metric structure on probability measures. This is going to be very important when it comes to decision making under uncertainty. Why? Well, effectively choosing a lottery (which is the bread and butter of expected utility theory) is equivalent to choosing a random variable on some mother space, whcih (as we have seen) is equivalent to choosing amongst probability measures. Thus, if we want to (say) have a model in which people choose the random variable with the highest expected utility, we better make sure that this concept is well defined. We know that, to guarantee this, we need a continuous function on a compact set. But in order to define continuity and compactness, we need metrics. This is what we now do.

For simplicity, we will think about metricizing probability measures on the borel sets on a metric space X, which we will denote as $\Delta(X)$. If X is a metric space, then at least we know how we would like to metricize to degenerate probability measures: one that assigns all its probability to $t \in X$ and the other to $s \in X$: we would simply use the distance between s and t. To go beyond this, we are going to reverse the order of events that we learned in the first year. There, you began with the notion of a metric, used this to define a topology, and thence convergence and continuity. Here we are going to begin by thinking about what functions we would like to be continuous, then use this to define convergence, and use this to generate a metric and then a topology.

Thus, the starting point that we are going to take is that we would like the expectation of all continuous and bounded functions[2] to be contininuous with respect to our probability measures. Think of this the following way: Let X be the real line and let u be some continuous and bounded untility function. We are going to be interested in the expectations of this utilty function with respect to p's defined on the borel sets of X. In particular, we are going to want to find the p in some subset $p \in P \subset \Delta(X)$ that maximizes utility. For this we are going to need the expectation of utility to be continuous, and P to be compact. The route we are going to take is do define a metric on $\Delta(X)$ such that the expectations operator is continuous on continuous and bounded functions

First, let x be a random variable, and note that the mapping $L_x : \Delta(X) \to \mathbb{R}$ defined by

$$L_x(p) = E_p(x)$$

defines a mapping from the space of all Borel probability measures to the real numbers. The discussion above suggests that we would like to define convergence in $\Delta(X)$ is such a way that $p_n \to p$ implies $L_x(p_n) \to L_x(p)$, if x is a continuous and bounded function

Definition 10 *Let X be a metric space and $\{p_m\}$ be a sequence in $\Delta(X)$. for any $p \in \Delta(X)$, we say that $\{p_m\}$ converges weakly to p is*

$$\int_X \theta dp_m \to \int_X \theta dp$$

[2] Note that by 'continuous random variable' we mean a random variable whose distribution function is continuous. This is distinct from a random variable defined by a function that happens to be continuous. Let X be a finite metric space with the discrete metric. Any real map on X is a contnuous function, but the resulting random varaible will not be continuous.

for every continuous, bounded function on θ. Equivalently, we have

$$E_{p_m}(x) \to E_p(x)$$

for every random variable on X that is continuous and bounded.

Note that, because we have not started with a metric, we do not know that any sequence p_m has a unique weak limit, but this in fact the case. A corrolary of this is that any two borel probabilty measures $p, q \in \Delta(X)$ are distinct if and only if

$$\int_X \theta dp \neq \int_X \theta dq$$

for some continuous bounded θ.

Here are some examples

Example 2 *Consider the probability space $(\{0,1\}, 2^{\{0,1\}} p_m)$ defined as*

$$p_m\{0\} = 1 - \frac{1}{m}$$

Then, for any real function θ on $(0,1)$ we have

$$\int_X \theta dp_m = \theta(0)\left(1 - \frac{1}{m}\right) + \theta(1)\frac{1}{m} \to \theta(0)$$

Thus, it must be the case that $p_m \to p$ where p puts probability 1 on $\{0\}$

Example 3 *Define a Dirac measure on a metric space in the following way: for some $w \in X$,*

$$\begin{aligned} \delta_w(S) &= 1 \text{ if } w \in S \\ &= 0 \text{ otherwise} \end{aligned}$$

for any borel set S. Intuitively, we would like a sequence of dirac measures δ_{w_m} to converge to δ_w if and only if $w_m \to w$. Is this the case? Let θ be a continuous bounded map, and note that

$$\int_X \theta dp_m = \theta(w_m) \to \theta(w) = \int_X \theta dp$$

by the convergence properties of continuous functions. Note also that if $w_m \nrightarrow w$, then δ_{w_m} does not converge weakly to δ_w. To see this, define $\theta(u)$ as $min\{1, d(u, w)\}$. This is a continuou sand bounded function, so if δ_{w_m} converges weakly to δ_w, it must be the case that

$$min\{1, d(w_m, w)\} = \int_X \theta dp_m \rightarrow \int_X \theta dp = 0$$

which in turn implies that it must be the case that $d(w_m, w) \rightarrow 0$

It is useful to have some equivalence results for weak convergence

Theorem 3 (Portmanteau) *For and metric space X and $p_1, p_2.... \in \Delta(X)$, the following are equivalent*

1. $p_m \rightarrow^w p$
2. $\limsup p_m(C) \leq p(C)$ *for every closed set C*
3. $\liminf p_m(O) \geq p(O)$ *for every open set O*
4. $\lim p_m(S) = p(S)$ *for every $S \in B(X)$ with $p(\delta_X S) = 0$ (i.e. the measure of the boundary of S is zero - these are sometimes called continuity sets)*