Indian Institute of Technology Bombay
Department of Electrical Engineering

Handout 9 EE 325 Probability and Random Processes
Lecture Notes 6 August 11, 2014

1 Measurable Spaces

Recall the definition of a probability space $(\Omega, \mathcal{F}, P)$. A slightly more general notion is that of a *measurable space*, which comprises of the first two entities, i.e. a sample-space and an associated sigma-field. We can define maps or functions between a pair of measurable spaces, similar to the familiar functions (like $x^2, \sin(x), \log(x)$ etc) which operate on the domain arguments to output values in their range. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces. A function is an association of every element $\omega_1 \in \Omega_1$ to some element of $\omega_2 \in \Omega_2$. We will denote this as

$$f : \Omega_1 \to \Omega_2,$$

where $f(\cdot)$ is the function with domain Ω_1 and range contained in Ω_2.

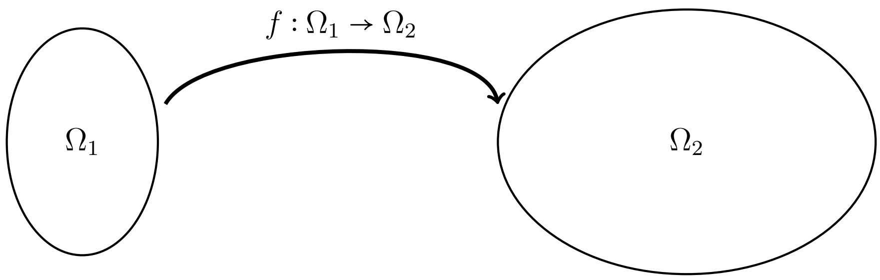

Figure 1: $f : \Omega_1 \to \mathcal{F}_2$

The key notion on functions that we need is known as **measurability** with respect to the domain and range spaces. Unlike the function, which specified between Ω_1 and Ω_2, 'measurability' also depends on the associated sigma-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ that we consider. Measurability is defined in terms of the *inverse map* or *inverse image* of a function.

Definition 1 *The set function $f^{-1}(\cdot)$ given by*

$$f^{-1}(B) = \{\omega \in \Omega_1 : f(\omega) \in B\}$$

is defined as the inverse image of the function $f : \Omega_1 \to \Omega_2$.

Thus the inverse images take sets as inputs and outputs another set. Contrast this with the forward map $f(\cdot)$, whose arguments are elements in Ω_1 and range-space is in Ω_2. Let us now define measurability of a function with respect to the sigma-fields $(\mathcal{F}_1, \mathcal{F}_2)$.

Definition 2 *We call a function $f : \Omega_1 \to \Omega_2$ as $(\Omega_1, \mathcal{F}_1) - (\Omega_2, \mathcal{F}_2)$* measurable*, if $\forall B \in \mathcal{F}_2$, $f^{-1}(B) \in \mathcal{F}_1$.*

In other words, the inverse image of the set $B \in \mathcal{F}_2$ should belong to the sigma-field F_1. The advantage of this definition will get more clear, however, let us be absolutely sure about measurability by doing a few examples.

Example 1 *Let* $\Omega_1 = \{a,b,c,d\}$ *and* $\Omega_2 = \{1,2,3,4,5\}$, *and assume* $\mathcal{F}_i = \mathcal{P}(\Omega_i)$, $i = 1,2$. *Consider a uniform probability assignment over* Ω_1. *For the map* $X(a) = 1, X(b) = 2, X(c) = 3, X(d) = 4$.

- *Find* $X^{-1}(\{3,4\})$, *is it in* $\mathcal{F}_1$?
- *Find* $X^{-1}(\{4,5\})$, *is it in* $\mathcal{F}_1$?
- *Find* $X^{-1}(\{5\})$, *is it in* $\mathcal{F}_1$?
- *Is* X *measurabe w.r.t* $(\mathcal{F}_1, \mathcal{F}_2)$?

From now onward we will shorten the statement 'X is $(\mathcal{F}_1, \mathcal{F}_2)-$ measurable' to simply 'X is measurable', and the sigma-fields should be clear from the context.

Example 2 *Let* Ω_1 *correspond to the tossing of a die and* $\Omega_2 = a,b,c,d$. *Consider the map* $X(1) = a, X(2) = X(3) = b, X(4) = X(5) = X(6) = c$. *For* $\mathcal{F}_i = \mathcal{P}(\Omega_i), i = 1,2$, *is* X *measurable?*

Example 3 *For the same* $\Omega_i, i = 1,2$ *and map* $X(\cdot)$ *in the last example, let us take* $\mathcal{F}_1 = \{\varnothing, \{1\}, \{2,3,4,5,6\}, \Omega_1\}$, $\mathcal{F}_2 = \mathcal{P}(\Omega_2)$. *Is* X *measurable.*

Thus, we have to take an appropriate sigma-field. Special care has to be taken about the sigma-field $\mathcal{F}_1$ in particular.

Example 4 *Let us keep* Ω_i *and* $\mathcal{F}_i$ *from the previous example and change the map to a new one, say* Y. *Just to make you think a bit all that I will tell you is some inverse images of* Y. *In particular,*

$$Y^{-1}(\{a,b\}) = \{1\}; Y^{-1}(\{b,c\}) = \{2,3,4,5,6\}.$$

Show that Y *is indeed a measurable function.*

Solution: The first thing that you should notice is that $\{a,b\}$ and $\{b,c\}$ generate the whole $\mathcal{P}(\Omega_2)$, i.e.

$$\sigma(\{a,b\},\{b,c\}) = \mathcal{P}(\Omega_2).$$

We have already shown this in the class, as any field which containts $\{a,b\}$, $\{b,c\}$ has to contain $\{a\},\{b\},\{c\}$ and $\{d\}$ also, which implies that this field is no smaller than $\mathcal{P}(\Omega_2)$. While we can check for each subset that its inverse is in $\mathcal{F}_1$, it turns out that checking $Y^{-1}(\{a,b\})$ and $Y^{-1}(\{b,c\})$ is sufficient for this example, and it is easy to see that both in $\mathcal{F}_1$. The sufficiency follows from the next result.

Proposition 1 *If* $X^{-1}(A_1) \in \mathcal{F}_1$ *and* $X^{-1}(A_2) \in \mathcal{F}_1$ *then*

$$X^{-1}(A_1 \bigcup A_2) \in \mathcal{F}_1.$$

Proof: By definition,

$$\begin{aligned} X^{-1}(A_1 \bigcup A_2) &= \{\omega_1 \in \Omega_1 : X(\omega_1) \in A_1 \text{ or } X(\omega_1) \in A_2)\} \\ &= \{\omega_1 \in \Omega_1 : X(\omega_1) \in A_1\} \bigcup \{\omega_1 \in \Omega_1 : X(\omega_1) \in A_2\} \\ &= X^{-1}(A_1) \bigcup X^{-1}(A_2). \end{aligned}$$

Clearly the RHS is in $\mathcal{F}_1$, as the latter is a field. ■

The above argument also works for countable unions. Let us now look at complements.

Proposition 2 *If* $X^{-1}(A) \in \mathcal{F}_1$, *then*

$$X^{-1}(A^c) \in \mathcal{F}_1.$$

Proof:

$$\begin{aligned} X^{-1}(A^c) &= \{\omega_1 \in \Omega_1 : X(\omega_1) \in A^c\} \\ &= \{\omega_1 \in \Omega_1 : X(\omega_1) \in A\}^c \\ &\left(X^{-1}(A)\right)^c \end{aligned}$$

Since $\mathcal{F}_1$ is a field, the result follows. ■

Putting this all together, we have the following simpler check for measurability.

Theorem 1 *Let* C *be a collection of sets such that* $X^{-1}(A) \in \mathcal{F}_1$ *for every* $A \in C$. *Then the map* X *is* $(\mathcal{F}_1, \sigma(C))-$ *measurable.*

The advantage of measurability becomes obvious if we know how to associate a measure to every event in $\mathcal{F}_1$. For example, this is the case when the first space is a probability space $(\Omega, \mathcal{F}, P)$. Measurability guarantees that we can associate a appropriate probability to every event in $\mathcal{F}_2$, by simply assigning $P(X^{-1}(B))$, $\forall B \in \mathcal{F}_2$. All that we did is to take the probability of the inverse image. If we can argue that this is a valid probability association, then $(\Omega_2, \mathcal{F}_2)$ will also become a probability space.

Theorem 2 *Consider a measurable map* $X : (\Omega_1, \mathcal{F}_1, P) \to (\Omega_2, \mathcal{F}_2)$, *let us define* Q *on* $(\Omega_2, \mathcal{F}_2)$ *as*

$$Q(B) = P(X^{-1}(B)).$$

Then $(\Omega_2, \mathcal{F}_2, Q)$ *is a probability space. The measure* Q *is known as the* ***induced probability measure****.*

Proof: Observe that we need to verify the axioms for the measure Q. First,

$$Q(B) = P(X^{-1}(B)) \geq 0.$$

Notice that $\{\omega \in \Omega_1 : X(\omega) \in \Omega_2\} = \Omega_1$. Then

$$Q(\Omega_2) = P(X^{-1}(\Omega_2)) = P(\Omega_1) = 1.$$

Finally, for disjoint sets B_i, $i \geq 1$,

$$\begin{aligned} Q(\bigcup B_i) &= P(X^{-1}(\bigcup B_i)) \\ &= P(\bigcup X^{-1}(B_i)) \\ &= \sum_{i \geq 1} P(X^{-1}(B_i)) \\ &= \sum_{i \geq 1} Q(B_i). \end{aligned}$$

Thus the sigma-additivity property is also satisfied the measure Q, indeed guaranteeing that it is a proper probability measure.

■

The term induced probability measure is used for Q to emphasize that Q is induced by the function X by its operation on $(\Omega_1, \mathcal{F}_1, P)$. Just like functions are extremely useful in analysis and elsewhere, measurable functions are very useful in probability theory. That they automatically induce a probability measure on the range-space is a boon, which allows scientists/engineers/students to use measurable functions like a microscope, seamlessly zooming in to the details of the ambient probability space.

2 Random Variables

When the range space of a measurable function is $\mathbb{R}$ (or a subset), and the corresponding sigma-field employed is $\mathcal{B}(\mathbb{R})$, the measurable function is also known as a **random variable**.

Definition 3 *A* ***random variable*** X *is a measurable function from* $(\Omega_1, \mathcal{F}_1, P)$ *to* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Some examples are in order.

Example 5 *Show that the given function X in the previous question is a random variable and compute the probability of the event $\{2,4,5\}$.*

Solution: Let us find the probability, assuming that X is a random variable.

$$Q(\{2,4,5\}) = P(X^{-1}(\{2,4,5\})) = P(\{b,d\}) = \frac{1}{2}.$$

One can verify that for all elements of $\mathcal{F}_2$, X will satisfy the measurability condition. Verifying this is in fact a painful job. Fortunately, there is a more straightforward way to check measurability, which we state without proof.

Theorem 3 *Let C be a class of subsets of Ω_2 and $\mathcal{F}_2 = \sigma(C)$. Then $X : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ is measurable iff $X^{-1}(C) \subset \mathcal{F}_1$.*

We have used the bi-directional iff statement above, and the condition states that the inverse image of each element in the class C should be in $\mathcal{F}_1$, i.e. $\forall A \in C, X^{-1}(A) \in \mathcal{F}_1$.

Example 6 *Consider tossing a fair die, i.e. $\Omega_1 = \{1,2,3,4,5,6\}$. Consider the identity map $X : (\Omega_1, \mathcal{P}(\Omega_1)) \to (\Omega_1, \mathcal{P}(\Omega_1))$, in other words, $X(i) = i$. Certainly all measurability properties are satisfied, Notice that the assertion of a fair die specifies a uniform distribution on Ω_1. Thus X is a random variable.*

Example 7 *Consider rolling two fair dice independently.*

$$\Omega_1 = \{(i,j) : 1 \le i,j \le 6\},$$

Consider the map $X(i,j) = i + j$ with domain Ω_1. The range of X is the set $\{2,3,\cdots,12\}$. Assuming $\mathcal{F}_1 = \mathcal{P}(\Omega_1)$, we can find

$$X^{-1}(5) = \{(2,3),(3,2),(4,1),(1,4)\} \in \mathcal{P}(\Omega_1).$$
$$X^{-1}(2,3) = \{(1,1),(2,1),(1,2)\} \in \mathcal{P}(\Omega_1).$$

In fact for each element in $\{2,\cdots,12\}$ one can verify that the inverse map is in $\mathcal{F}_1$, guaranteeing that X is indeed a random variable. Using the induced probability measure Q on Ω_2,

$$Q(\{5\}) = P(X^{-1}\{5\}) = P\{(2,3),(3,2),(4,1),(1,4)\} = \frac{4}{36} = \frac{1}{9}.$$

■

As we mentioned in the class, random variables are like credit-cards. Once you set up a bank account in a good place (say Switzerland), i.e. constructed a suitable probability space $(\Omega_2, \mathcal{F}_1, P)$, then you can swipe the card in whichever shop you like and buy goods. No worries about currency conversions etc, it will all be calculated and accounted in terms of the original measure $P(\cdot)$. Some people may not like associating random-variables to one currency. They ask, why just reals? why not other objects?. No problems, the terminology variables are conventionally associated with reals and complex-values, hence the usage. To exemplify, let us do an experiment on random chords, as in the Bertrand's paradox. I should warn that we may be jumping the gun a bit, we have not completed the theory that is being built, but please bear with me, and ask questions if there is any confusion.

2.1 Choosing a Chord

This example also helps us to understand random objects which are not just numbers or variables. Here we will revisit our *paradoxical* problem of random chords of a unit circle centered at origin. Recall that we motivated the modern axiomatic approach based on this example. We will consider two different situations in this section, one is the classical version of Bertrand. We will also mention another problem where we wish to find the probability of a chord cutting both the positive $y-$axis and the negative $x-$axis. This will further exemplify the concepts.

Let us throw more light to the Betrand's problem by using what we have learned. In particular, choosing a chord is about choosing two end-points of the chord, and this can be done in several ways. Notice that each end-point is specified by one real-value, as the circle constrains the other coordinate. Thus, two end-points will correspond to two real values, and we will propose a method to choose those. starting at $(\cos(2\pi x), \sin(2\pi x))$ and ending at $(\cos(2\pi y), \sin(2\pi y))$. Clearly both points lie in the unit circle.

Instead of assigning probability to the chords, let us take recourse our *credit utility* of random variables or random chords. Our strategy is as follows: since there are two parameters involved for each chord, let us construct a probability space on $\Omega_1 = [0,1]^2$, i.e. the unit square. We have already showed that taking rectangular intervals and assigning the area of the interval as the probability measure will do. This procedure associates itself with the sigma-fied generated by open rectangles $[0,1]^2$, let us call this as $\mathcal{F}_1$. The probability space $(\Omega_1, \mathcal{F}_1, P)$ is thus ready.

We will create a second space Ω_2 which will contain the objects of interest, i.e. chords of the unit circle. Without much ado define Ω_2 as the collection of all chords of the unit circle centered at origin. Let us now define a suitable map from Ω_1 to Ω_2. Since the outputs are not variables or vectors, rather chords, let us denote the map by $RC : \Omega_1 \to \Omega_2$. Consider the map

$$RC(\omega_x, \omega_y) = \left[(\cos(2\pi\omega_x), \sin(2\pi\omega_x)), (\cos(2\pi\omega_y), \sin(2\pi\omega_y))\right].$$

Clearly each point in Ω_1 is mapped to some chord of the unit circle. The figure below illustrates this for the point $p = (\frac{1}{4}, \frac{5}{8})$ on the unit square, where p_1 is the start of the chord and p_2 the endpoint. Also shown is a map for another point q. Clearly, with the map above, every point in Ω_1 will be mapped to a directed chord (possibly de-generate).

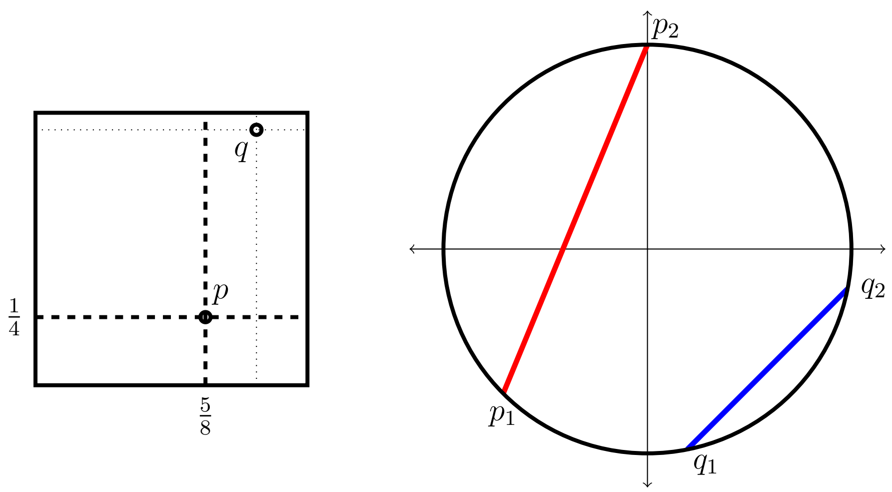

Figure 2: Unit Square and Circle (not to scale), the point p corresponds to the red chord, and point q corresponds to the blue one

The next question is about $\mathcal{F}_2$. For simplicity, let us take A as the set of all chords with length greater than $\sqrt{3}$. The associated sigma-field then is

$$\mathcal{F}_2 = \{\Omega_1, A, A^c, \varnothing\}$$

It now makes sense to ask, 'what is the probability of the chord having length greater than $\sqrt{3}$?. In other words, what is $P(A)$?. More specifically, we are being asked to evaluate the induced probability.

Alternately, the uniform distribution on the square $[0,1]^2$ corresponds to taking two coordinates, where each coordinate is chosen uniformly from $[0,1]$. We also need to say that the values are taken independently, a notion we are yet to define(please come next class). Observe that an $x-$coordinate value 0 is mapped to the point $(1,0)$ on the circle and the $x-$coordinate $\frac{1}{4}$ is mapped to $(0,1)$. Tracing the values of x from 0 to 1 will traverse the circular perimeter once in the anti-clockwise direction. The length of the chord will be greater than $\sqrt{3}$ whenever the chosen points x and y obeys

$$\frac{1}{3} \leq y - x \leq \frac{2}{3}.$$

This is marked in the following figure. Clearly the marked area is $\frac{1}{3}$, same as the probability

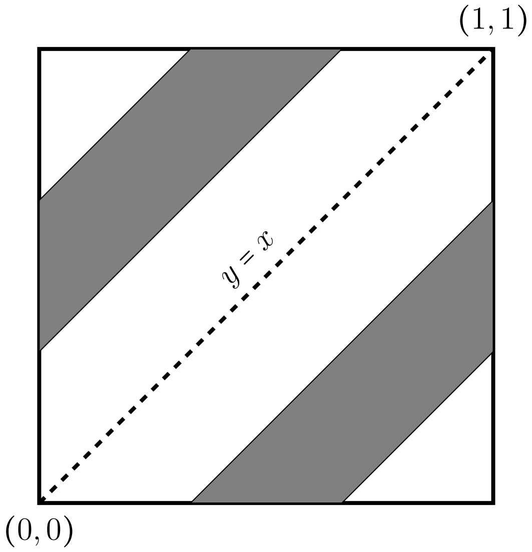

Figure 3: Region (shaded) Corresponding Chord-length at least $\sqrt{3}$

that we have computed earlier. More importantly, there is no ambiguity in this experiment, as the notion of uniform randomness was defined on the unit-square, which is our Ω. The random chords were generated using the above method to give an induced measure on the space of chords. If we choose some other mapping, we will get a different random chord experiment, and the probability computations may give a different answer. This is okay, as the induced measure by different random variables or random chords need not be the same.

Exercise 1 *Let us consider the experiment of a random chord cutting both the positive $y-$ axis and negative $x-$axis. Using the same map as above, can you find this probability. Write down all the spaces, fields and measure that you use.*