Random variables

Sayan Mukherjee

The key ideas covered in this section will be measurability and observability. Measureability is a analytic or measure theoretic property while observability is a empirical property. The two are directly related.

Definition

In the following definition think of $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, and $\mathbf{P} =$ No.

Definition 0.0.1 *Given a measurable space* $(\Omega, \mathcal{F})$*. A real valued random variable is a function* $X : \Omega \to \mathbb{R}$ *that is* $\mathcal{F} \setminus \mathcal{B}$ *if*

$$\{\omega : X(\omega) \in B\} \in \mathcal{F}, \ \forall \ B \in \mathcal{B},$$

or more commonly

$$X^{-1}(B) \equiv \{\omega : X(\omega) \in \mathcal{B}\},$$

is a measureable set in Ω

Given two measurable spaces $(\Omega, \mathcal{B})$ and $(\Omega', \mathcal{B}')$ a map or function $X : \Omega \to \Omega'$ is measureable in $\mathcal{B} \setminus \mathcal{B}'$ if $X^{-1}(\mathcal{B}' \subset \mathcal{B}$. A random variable is the spcial case wehere $(\Omega', \mathcal{B}')$ is $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

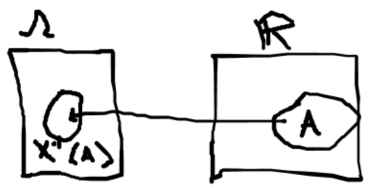

Figure 1: A random variable, the map $X^{-1}(A)$.

Properties of set inverses:

1. $X^{-1}(\emptyset) = \emptyset, \quad X^{-1}(\Omega') = \Omega$
2. $X^{-1}(A'^c) = (X^{-1}(A'))^c$
   $X^{-1}(\Omega' \setminus A') = \Omega \setminus X^{-1}(A')$
3. $X^{-1}(\cup A'_t) = \cup X^{-1}(A'_t)$
   $X^{-1}(\cap A'_t) = \cap X^{-1}(A'_t)$

Definition 0.0.2 *Given the triple* $(\Omega, \mathcal{F}, \mathbf{P})$ *the distribution of a random variable* X *is*

$$\mu_X \equiv \mathbf{P} \circ X^{-1}(A') = \mathbf{P}(X^{-1}(A')).$$

Note that $\mathbf{P} \circ X^{-1}$ *is a function on* $\mathcal{B}(\mathbb{R})$*.*

To examine that μ_X is a measureon $\mathcal{B}(\mathbb{R})$ the following need to be verified

1. $\mathbf{P} \circ X^{-1}(\mathbb{R}) = \mathbf{P}(\Omega) = 1$

2. $\mathbf{P} \circ X^{-1}(A') \geq 0, \forall A' \in \mathcal{B}(\mathbb{R})$
3. If $\{A'_n, n \geq 1\}$ are disjoint

$$\begin{aligned}\mathbf{P} \circ X^{-1}(\cup_n A'_n) &= \mathbf{P}(\cup_n X^{-1}(A'_n)) \\ &= \sum \mathbf{P}(X^{-1}(A'_n)) \\ &= \sum \mathbf{P} \circ X^{-1}(A'_n).\end{aligned}$$

The idea of the composition follows

$$\begin{aligned}X_1 : (\Omega_1, \mathcal{B}_1) &\rightarrow (\Omega_2, \mathcal{B}_2) \\ X_2 : (\Omega_2, \mathcal{B}_2) &\rightarrow (\Omega_3, \mathcal{B}_3).\end{aligned}$$

$X_2 \circ X_1 : \Omega_1 \rightarrow \Omega_3$ and $X_2 \circ X_1(\omega_1) = X_2(X_1(\omega_1))$.

ADD COMPOSITION PICTURE.

Constructing Lebesgue measure

Set $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, and $\mathbf{P} = \lambda$ be the Lebesgue measure. Our objective will be to construct random variables with the Lebesgue measure on Ω. We will show how we can use a dyadic representation of real numbers coupled with iid Bernoulli variables to construct the Lebesgue measure.

First note that every $x \in [0,1]$ can be written as the following sum which we can also think of as a binary representation of x

$$x = \frac{\delta_1}{2} + \frac{\delta_2}{2^2} + ..., \quad \delta_i = \{0,1\}, \tag{1}$$

or x is written as $.\delta_1\delta_2\delta_3\cdots$, this representation is not unique becuase

$$\frac{1}{2} = \frac{1}{2} + \frac{0}{2^2} + \frac{0}{2^3} + \cdots = \frac{0}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots.$$

We will use the second representation, the infinite string of ones. The reason for this will become apparent soon and is based on the convention that the intervals we have used to construct the Borel sets take the form $(a,b]$, for example the distribution function in this case is on $(0,a]$. We need to construct a series of random variables $\{\xi_i\}, i \geq 1$ such that the following property of the Lebesgue measure holds

$$\begin{aligned}\mathbf{P}(x : \xi_1 = \delta_1, ..., \xi_n = \delta_n) &= \mathbf{P}\left(x : \frac{\delta_1}{2} + \frac{\delta_2}{2^2} + \cdots + \frac{\delta_n}{2^n} \leq x < \frac{\delta_1}{2} + \frac{\delta_2}{2^2} + \cdots + \frac{\delta_n}{2^n} + \frac{1}{2^n+1}\right) \\ &= \mathbf{P}\left(x : x \in \left[\frac{\delta_1}{2} + \frac{\delta_2}{2^2} + \cdots + \frac{\delta_n}{2^n}, \frac{\delta_1}{2} + \frac{\delta_2}{2^2} + \cdots + \frac{\delta_n}{2^n} + \frac{1}{2^{n+1}}\right]\right) = \frac{1}{2^{n+1}}.\end{aligned}$$

There are 2^{n+1} intervals so $\sum_{i=1}^{2^{n+1}} \frac{1}{2^{n+1}} = 1$.

Define the following Bernoulli sequence $\xi_n \overset{iid}{\sim} \text{Be}(p = 1/2)$, for any finite $n \geq 1$ this random variable measurable with measure

$$\mathbf{P}(\xi_1 = \delta_1, ..., \xi_n = \delta_n) = p^{\sum \delta_i}(1-p)^{n - \sum \delta_i}.$$

We can see that the probability of the sequence $\mathbf{P}(\xi_1 = \delta_1, ..., \xi_n = \delta_n, \xi_{n+1} = 0) = \frac{1}{2^{n+1}}$ which is equivalent to $P(x : \xi_1 = \delta_1, ..., \xi_n = \delta_n)$. This verifies for intervals (a semi-algebra) the Bernoulli construction for the Lebesgue measure holds.Define $a_i < b_i \in \{0, 1, ..., 2^n\}$ and the collection $\mathcal{F}_n\{\bigcup_i \left(\frac{a_i}{2^n}, \frac{b_i}{2^n}\right]\}$. Again for a fixed n the Bernoulli construction will hold and $\mathcal{F}_n$ is an algebra but not a sigma algebra, $\mathcal{F} = \sigma(\bigcup_n \mathcal{F}_n) = \mathcal{B}([0,1])$ is the Borel sigma algebra. By extension of measure this construction will give rise to Lebesgue measure on $\mathcal{F}$, one can also show uniqueness by the extension arguments in the previous notes. In summary differnt distributions $\mathbf{P}$ on the Borel set can be placed using various joint distributions on the Bernoulli random variables $(\xi_n), n \geq 1$. For any distribution function $F(x)$ (a non-decreasing right-continuous function $F(x)$ satisfying $F(0) = 0$ and $F(1) = 1$) the measure $\mathbf{P}_F((a,b]) \equiv F(b) - F(a)$ determines a probability distribution on every $\mathcal{F}_n$ that extends uniquely to $\mathcal{F}$ and implies a joint distribution on the sequence $\{\delta_n\}$.

Measurability and observability

We develop in this section the relation between measurability and observability. This relation is often developed in the theory of Martingales and in filtering approaches such as the Kalman-Bucy filter.

We will develop the idea using the example of a symmetric random walk which is a sequence of random variables Y_n on $(\Omega, \mathcal{F}, \mathbf{P})$ with

$$Y_n(\omega) = 2S_n - n, \quad S_n = \sum_{i=1^n} \delta_i \sim \text{Bin}(n, p = .5),$$

the walk starts at the origin and moves to the left or right with probability $.5$.

The sigma algebra generated by the first n S_i's are the finite algebra $\mathcal{F}_n$ of all unions of half-open intervals with endpoints at $j2^{-n}$ with $j = \{0, 1, ..., 2^n\}$. A random variable Z on $(\Omega, \mathcal{F}, \mathbf{P})$ will be $\mathcal{F}_n$ measurable iff Z has the form $Z = \phi_n(\delta_1, ..., \delta_n$ on the first n δ_i's. Thus measurability in the$\mathcal{F}_n$ sense means that we can determine the value of the r.v. by observing the first n values of δ_i or S_i. Z on Ω is $\mathcal{F}$ measurable iff

$$Z = \lim_{n \to \infty} \phi_n(\delta_1, ..., \delta_n),$$

where "=" above and the limit above will be studied in the next sections, what does convergence of a r.v. mean.