Basic Random Variable Concepts

ECE275A – Lecture Supplement – Fall 2016

Kenneth Kreutz–Delgado
Electrical and Computer Engineering
Jacobs School of Engineering
University of California, San Diego

VERSION F16.LSRV.V1.0b

Copyright © 2006-2016, All Rights Reserved

October 19, 2016

Borel σ–Algebra of the Real Line. As has been repeatedly stressed, probability theory is concerned with the random occurrence of *events*. In particular, we are often concerned with a sample space of *real numbers,* $\Omega = \mathbb{R}$, and corresponding measured real–valued experimental outcomes forming events of the following type:

1. A real number (experimental outcome) takes a specified real value v, $x = v \in \mathbb{R}$, in which case we say that the *singleton* or *point–set event* $\{v\} \subset \mathbb{R}$ occurs.

2. A real number (outcome) takes it value in one of the following open– and closed–*half–spaces*: $(-\infty, w) = \{x \,|\, -\infty < x < w\}$, $(v, \infty) = \{x \,|\, v < x < \infty\}$, $(-\infty, w\,] = \{x \,|\, -\infty < x \le w\}$, $[\,v, \infty) = \{x \,|\, v \le x < \infty\}$. If an outcome takes its value in a specified half–space we say that a *half–space event* occurs.

3. A real number takes its value either in the *closed interval* $[\,v, w\,] = \{x \,|\, v \le x \le w\}$, in the *open interval* $(v, w) = \{x \,|\, v < x < w\}$, in the *half–open interval* $(v, w\,] = \{x \,|\, v < x \le w\}$, or in the *half–open interval* $[\,v, w) = \{x \,|\, v \le x < w\}$. If an outcome takes its value in such an interval, we say that an *interval event* occurs. Note that we can view the occurrence of the point–set event $\{v\}$ as equivalent to the occurrence of the degenerate interval event $[v, v]$. Similarly, we can view the half–space events as interval events of a limiting type, such as $[v, \infty) = \lim_{w \to \infty} [v, w)$, etc.

The smallest σ–Algebra of subsets of $\Omega = \mathbb{R}$ which contains such interval events is called the *Borel algebra*[1] (or Borel field), $\mathcal{B} = \mathcal{B}(\mathbb{R})$.[2] This σ-algebra is closed under countable intersections, unions, and complements of the just–mentioned event–intervals. The Borel algebra $\mathcal{B}$ contains events corresponding to logical boolean questions that can be asked about event–intervals in $\mathbb{R}$.[3] Events (sets) in the Borel algebra $\mathcal{B}$ are called *Borel sets*.

- Given a probability measure, $P_{\mathcal{B}}$, on $\mathcal{B}$, we then have the very important and useful *Borel Probability Space,* $(\mathbb{R}, \mathcal{B}, P_{\mathcal{B}})$, whose events are the Borel sets, and which therefore include the open– and closed–intervals in $\mathbb{R}$.

Measurable Functions and Random Variables. We are interested in real–valued functions, $w = f(\omega) \in \mathbb{R}$, which are single–valued[4] mappings between the outcomes in a sample space Ω and the real numbers $\mathbb{R}$,

$$f : \Omega \to \mathbb{R} .$$

Given a function f, the *Inverse Image*, or *Pre–image*, of a set of points $M \subset \mathbb{R}$ is defined as

$$f^{-1}(M) = \{\omega \,|\, f(\omega) \in M\} \subset \Omega .$$

By definition, a function f is $\mathcal{A}$*–measurable*, or just *measurable*, if and only if,

$$f^{-1}(M) \in \mathcal{A} , \quad \forall M \in \mathcal{B} .$$

Given a probability space $(\Omega, \mathcal{A}, P_{\mathcal{A}})$ and an $\mathcal{A}$–measurable function f, we can legitimately apply the probability measure $P_{\mathcal{A}}$ to the pre–image of any Borel set $M \in \mathcal{B}$ as $f^{-1}(M)$ must be an event in $\mathcal{A}$. If this is the case, we say that the function f is an $\mathcal{A}$–measurable *random variable*.

Definition of a Random Variable

> Given a sample space, Ω, with an associated σ–algebra, $\mathcal{A}$, of events in Ω, a *Random Variable*, $X : \Omega \to \mathbb{R}$, is an $\mathcal{A}$*–measurable real–valued function* on Ω.

We will always use upper case Roman letters to indicate a random variable to emphasize the fact that a random variable is a *function* and *not* a number. Note that whether or not X is

[1]Named after noted French mathematician Félix Édouard Émile Borel, 1871–1956.

[2]For you math-types, more generally the Borel σ–algebra of a topological space $\mathbb{X}$, $\mathcal{B}(\mathbb{X})$, is the smallest σ–algebra of subsets of $\mathbb{X}$ that contains every open set of $\mathbb{X}$. Such issues are studied in graduate mathematical probability courses, such as Math 280 at UCSD.

[3]I.e., "Did the experiment result in a measurement value between 1 and 5 inclusive, *or* between 11 and 17 noninclusive, *or* strictly greater than 35?"

[4]This statement is actually redundant. By *definition*, functions *must* be single–valued mappings. Non—single–valued mappings are called *relations*. In CSE 20, the properties of both functions and relations are studied in some detail.

an $\mathcal{A}$–measurable random variable depends on the the function X *and* the specific choice of the pair $(\Omega, \mathcal{A})$.[5]

A most important fact is that a random variable, X, on a probability space $(\Omega, \mathcal{A}, P_{\mathcal{A}})$ *induces a probability measure on* $(\mathbb{R}, \mathcal{B})$ via the definition,

$$P_{\mathcal{B}}(M) \triangleq P_{\mathcal{A}}(X^{-1}(M)), \quad \forall M \in \mathcal{B}.$$

Thus, corresponding to the probability space $(\Omega, \mathcal{A}, P_{\mathcal{A}})$ and the random variable X, we have the *induced probability space* $(\mathbb{R}, \mathcal{B}, P_{\mathcal{B}})$. Given a probability space $(\mathbb{R}, \mathcal{B}, P_{\mathcal{B}})$ induced by the random variable X, we refer to $(\Omega, \mathcal{A}, P_{\mathcal{A}})$ as the *base space*, or the *underlying probability space.*

If we are interested solely in real–valued events which are outcomes in the Borel probability space $(\mathbb{R}, \mathcal{B}, P_{\mathcal{B}})$, then often there is no need to work with the underlying space. Typically, in such cases, either a Borel probability space is directly assumed to exist (and we never directly consider the underlying space, except to acknowledge its existence) or the Borel probability space is constructed once from the definition of the induced probability measure on $\mathcal{B}$, after which the underlying space is ignored and we work solely with the Borel probability space thereafter.[6]

It is a fact that if X is a random variable for $(\Omega, \mathcal{A}, P_{\mathcal{A}})$, then there exists a smallest (or coarsest) σ–algebra, $\mathcal{A}'$, of Ω for which X is still measurable. We denote the smallest σ–algebra by $\mathcal{A}' = \sigma(X)$. This smallest σ–algebra, $\sigma(X)$, is given by

$$\sigma(X) = X^{-1}(\mathcal{B}) = \left\{ A \subset \Omega \,\middle|\, A = X^{-1}(M),\ M \in \mathcal{B} \right\} .$$

Necessarily, then, $\sigma(X) \subset \mathcal{A}$ and therefore X is also a random variable for the (possibly[7]) coarser probability space $(\Omega, \sigma(X), P_{\mathcal{A}})$. Measurements of experimental outcomes in the corresponding Borel space $(\mathbb{R}, \mathcal{B}, P_{\mathcal{B}})$ can never allow us to distinguish between the existence of $(\Omega, \mathcal{A}, P_{\mathcal{A}})$ or the coarser underlying space $(\Omega, \sigma(X), P_{\mathcal{A}})$. If all we have is data in the Borel algebra, then no harm incurs by working with the coarser underlying algebra $(\Omega, \sigma(X), P_{\mathcal{A}})$.

Measurable Space $(\Omega, \mathcal{A})$. If $\mathcal{A}$ is a σ–algebra of outcomes in Ω, we call the pair $(\Omega, \mathcal{A})$ a *measurable space*. A random variable X, then, is a measurable function from the measurable space $(\Omega, \mathcal{A})$ to the measurable Borel space $(\mathbb{R}, \mathcal{B})$.[8] Note that a probability space, $(\Omega, \mathcal{A}, P)$, is a measurable space $(\Omega, \mathcal{A})$ plus a probability measure P.

---

[5] That is, an $\mathcal{A}$–measurable random variable X might not be measurable with respect to a *different* σ–algebra of Ω. However, if $\mathcal{A}'$ is a σ–algebra of Ω such that $\mathcal{A} \subset \mathcal{A}'$, then X is also $\mathcal{A}'$ measurable and is then a random variable for the "finer" system $(\Omega, \mathcal{A}')$. For example, if $\mathcal{A}'$ is the power set of Ω (the "finest" σ–algebra of Ω) and $\mathcal{A}$ the trivial σ–algebra of Ω (the coarsest σ–algebra), then an $\mathcal{A}$–measurable X is also $\mathcal{A}'$ measurable.

[6] However, when dealing with jointly random variables, random vectors, or random processes (as discussed subsequently) one usually postulates the existence of an underlying probability space.

[7] It might be the case that $\sigma(X) = \mathcal{A}$.

[8] We also say that X is $\mathcal{A}$–measurable to emphasize the fact that measurability is the requirement that $X^{-1}(M) \in \mathcal{A}$ for all $M \in \mathcal{B}$. Thus X can be measurable with respect to a σ–algebra $\mathcal{A}$ while failing to be measurable with respect to a different σ–algebra, $\mathcal{A}'$, of Ω.

Distribution Function. Let the Borel set M_x denote the closed half–space,

$$M_x = (-\infty, x\,] = \{\xi \,|\, -\infty < \xi \le x\} \ , \quad x \in \mathbb{R}\,.$$

If X is a random variable for the probability space $(\Omega, \mathcal{A}, P_{\mathcal{A}})$, its (Cumulative) *Distribution Function* F_X is defined by

$$F_X(x) \triangleq P_{\mathcal{B}}(M_x) = P_{\mathcal{A}}(X^{-1}(M_x)) = P_{\mathcal{A}}(\{\omega \,|\, X(\omega) \in M_x\})\,, \quad \forall x \in \mathbb{R}\,.$$

Note that the $\mathcal{A}$–measurability of the random variable X and the existence of the probability measure $P_{\mathcal{A}}$ on $\mathcal{A}$–events are both crucial to this definition. A distribution function has several important properties which are discussed in Section 3.2 of Leon-Garcia.

It can be shown that knowing the probability distribution function, F_X, for a random variable X is entirely equivalent to knowing its X–induced probability measure, $P_{\mathcal{B}}$, on the algebra of Borel subsets of $\mathbb{R}$. They contain exactly the same information about the probability of events which are Borel sets in $\mathbb{R}$. Therefore, if we know the distribution function for the random variable X and if we are interested solely in real–valued events which are outcomes in the Borel probability space $(\mathbb{R}, \mathcal{B}, P_{\mathcal{B}})$, then there is no need to work with the underlying probability space or the induced probability measure $P_{\mathcal{B}}$. The distribution function, F_X, alone contains all the information we need to compute the probability of Borel events.

Random Variable defined in Basic Probability Courses. In many basic courses on probability a random variable is not defined as a function. In such courses a distinction is made between a random variable X and its realization value x, but this distinction can be unclear. We give two different (but closely related) interpretations for X as presented in such basic probability courses that clearly shown the relationship between the random variable X and the realization variable x.

1. X is a *random variable* as defined above. Let the underlying probability space $(\Omega, \mathcal{A}, P_{\mathcal{A}})$ be given by $\Omega = \mathbb{R}$ (so that $\omega = x \in \mathbb{R}$), $\mathcal{A} = \mathcal{B}$, and $P_{\mathcal{A}} = P_{\mathcal{B}}$. Then the random variable $X : \Omega \to \mathbb{R}$ is taken to be the *identity map* $X = I$, $\omega = x \mapsto x$. With this interpretation we have $X(\omega) = X(x) = I(x) = x$.

2. Let $X = \Omega$ stand for a *sample space* of real numbers (e.g., $\Omega = X = \mathbb{R}$). Thus a sample space realization is given by $\omega = x \in X$, which (by an abuse of notation) we denote by $X = x$.[9] Tacitly, the σ-algebra of events, $\mathcal{X}$, is comprised of the Borel sets in X. Thus we are dealing with a probability space $(X, \mathcal{X}, P_{\mathcal{X}})$ where in lieu of the probability measure $P_{\mathcal{X}}$ we work with a distribution function. In this interpretation, the "random variable" X is shorthand for the measurable space $(X, \mathcal{X})$.

[9] This is an abuse of notation because X is a set while x is a member of that set. Note that in the standard interpretation of X as a function over a probability space Ω (which we used in the first interpretation) there is no problem in stating that $X(\omega) = x$ (i.e., that x is the value of the function X operating on the realization $\omega \in \Omega$.). There is potential ambiguity in writing $X = x$ since X is a function and x is a value in the codomain of x, but there is no problem if it is understood that we mean $X(\omega) = x$.

Jointly Random Variables. We say that random variables X and Y are *jointly random* if they are both measurable mappings to the Borel algebra $(\mathbb{R}, \mathcal{B})$ from the *same* measurable space $(\Omega, \mathcal{A})$. Note that if X and Y are jointly random, then subsets of Ω like

$$X^{-1}(M) \cap Y^{-1}(N)\,, \quad M, N \in \mathcal{B}\,,$$

are events in $\mathcal{A}$.

Stochastic Process, Random Process, Random Vector. A *random process*, or *stochastic process*, is an indexed collection of jointly random variables $\{X_t(\cdot),\ t \in T\}$ where the *index set*, T, is some subset of the real numbers.

If the index set is the set of natural numbers, $\mathbb{N} = \{1, 2, \cdots\}$, then we call a collection of jointly random variables $\{X_k(\cdot),\ k \in \mathbb{N}\}$ a *random sequence*, or a *discrete–time random process*

A finite collection of n jointly random variables, $\{X_k(\cdot),\ 1 \le k \le n\}$, is called an n–dimensional *random vector* and is denote by

$$\mathbf{X} = (X_1, \cdots, X_n)^T \ .$$

Note that the n–dimensional random variable $\mathbf{X}$ takes its values in $\mathbb{R}^n$,

$$\mathbf{X}(\omega) = \mathbf{x} \in \mathbb{R}^n\,.$$

Almost Sure Equality of Random Variables. Two jointly random variables X and Y are said to be *equal almost surely*, or in *equal with probability 1*, designated as $X = Y$ *a.s.* iff,

$$P(\{\omega \,|\, X(\omega) \neq Y(\omega)\}\,) = 0\,.$$

It can be shown that $X = Y$ a.s. iff the events $X^{-1}(M)$ and $Y^{-1}(M)$ are equal almost surely for each Borel set $M \in \mathcal{B}$.[10]

Independence of Two Random Variables. Two jointly random variables on a probability space $(\Omega, \mathcal{A}, P)$, are said to be *independent*, iff the events $X^{-1}(M)$ and $Y^{-1}(N)$ are independent for all Borel sets M, N. I.e., iff,

$$P(X^{-1}(M) \cap Y^{-1}(N)) = P(X^{-1}(M)) \cdot P(Y^{-1}(N)) \quad \forall M, N \in \mathcal{B}\,.$$

Note that *independence of two random variables depends on the probability measure* P. Two $\mathcal{A}$–measurable random variables which are independent on the probability space $(\Omega, \mathcal{A}, P)$ might be dependent on a probability space, $(\Omega, \mathcal{A}, P')$, which has a different probability measure $P' \neq P$. Thus, independence of random variables is a property which has to be determined for each possible probability measure, P, used on a measurable space $(\Omega, \mathcal{A})$.

[10] Recall the definition of almost sure equality of events given the lecture supplement on probability concepts.

It can be shown that two random variables X and Y are independent iff the joint distribution function of X and Y is equal to the product of the marginal distributions,

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y) \,.$$

Equivalently, they are independent iff the joint probability density function, or probability mass function, is equal to the product of the marginals,

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \,.$$

Equivalence, Equality in Law, Equality in Distribution. Two random variables X and Y are said to be *equivalent*, or *equal in law*, or *equal in distribution*, iff they have the same probability distribution function,

$$F_X(x) = F_Y(x) \,, \quad \forall x \in \mathbb{R} \,.$$

Equivalently, X and Y are equal in law iff

$$f_X(x) = f_Y(x) \,, \quad \forall x \in \mathbb{R} \,.$$

Equality in law is a very weak (indeed the weakest) form of stochastic equivalence. If two *continuous* random variables X and Y are equal in law *and independent*[11] then it is the case that $X \neq Y$ a.s. That is the *equivalent* random variables X and Y are *almost surely not equal*, $P(\{\omega \,|X(\omega) = Y(\omega)\}) = 0$, which is the exact opposite of almost sure equality! *Equality almost surely does implies equality in law (equivalence), but (as seen here) the converse is not true.* Equivalence of random variables has a very specific technical meaning that must *not* be confused with the ordinary English language meaning of equivalence.

Independent Collection of Random Variables. An indexed set of random variables $\{X_t \,,\ t \in T\}$ is an *independent collection of random variables* iff given any finite subset of the collection, the set of pre–images obtained by letting the random variables in this subset pull–back all possible Borel sets forms an independent collection of events.

It can be shown that an indexed set of random variables is an independent collection iff for any subset of random variables, $X_{\alpha_1}, \cdots, X_{\alpha_n}$, drawn from this collection we have

$$f_{X_{\alpha_1},\cdots,X_{\alpha_n}}(x_1,\cdots x_n) = f_{X_{\alpha_1}}(x_1)\cdots f_{X_{\alpha_n}}(x_n) \,,$$

or equivalently,

$$F_{X_{\alpha_1},\cdots,X_{\alpha_n}}(x_1,\cdots x_n) = F_{X_{\alpha_1}}(x_1)\cdots F_{X_{\alpha_n}}(x_n) \,,$$

Note, in particular, that if $\mathbf{X}$ is an n–dimensional random vector with independent components, we must have that the joint pdf and cdf are both equal to the product of their marginals,

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1)\cdots f_{X_n}(x_n) \quad \text{and} \quad F_{\mathbf{X}}(\mathbf{x}) = F_{X_1}(x_1)\cdots F_{X_n}(x_n) \,.$$

[11] As discussed subsequently, we say that they are *independent and identically distributed* or *iid.*

IID Sequence or IID Vector of Random Variables. Given a denumerable collection of random variables (e.g., a countable random sequence or a finite–dimensional random vector) we say that the collection is *independent and identically distributed (iid)* iff the collection is independent and all random variables in the collection have exactly the same distribution function.[12]

If the random variables in an iid collection are all *continuous* random variables, then it can be shown that *almost surely no two of them are equal.* That is, almost surely they all take different values. This generalizes the two iid random variable case discussed above.

A collection of iid samples is known as a *statistical sample* or, more simply, as a *sample.*

[12]I.e., all random variables in the collection are equal in law to each other and to a single random variable having the distribution function in question.