ECE 6010
Lecture 1 – Introduction; Review of Random Variables

Readings from G&S: Chapter 1. Section 2.1, Section 2.3, Section 2.4, Section 3.1, Section 3.2, Section 3.5, Section 4.1, Section 4.2, Section 4.4, Section 4.5

Why study probability?

1. Communication systems. Noise. Information.
2. Control systems. Noise in observations. Noise in interference.
3. Computer systems. Random loads. Networks. Random packet arrival times.

Probability can become a powerful engineering tool. One way of viewing it is as "quantified common sense." Great success will come to those whose tool is sharp!

Set theory

Probability is intrinsically tied to set theory. We will review some set theory concepts.

We will use ^{c} to denote complementation of a set with respect to its universe.

$\cup$ – union: $A \cup B$ is the set of elements that are in A **or** B.

$\cap$ – intersection: $A \cap B$ is the set of elements that are in A **and** B. We will also denote this as AB.

$a \in A$: a is an element of the set A.

$A \subset B$: A is a subset of B.

$A = B$: $A \subset B$ and $B \subset A$.

Note that

$$A \cup A^c = \Omega$$

(where Ω is the universe).

Notation for some special sets:

$\mathbb{R}$ – set of all real numbers

$\mathbb{Z}$ – set of all integers

$\mathbb{Z}^+$ – set of all positive integers

$\mathbb{N}$ – set of all natural numbers, 0,1,2,…,

$\mathbb{R}^n$ – set of all n tuples of real numbers

$\mathbb{C}$ – set of complex numbers

Definition 1 A **field** (or algebra) of sets is a collection of sets that is closed under complementation and finite union. □

That is, if $\mathcal{F}$ is a field and $A \in \mathcal{F}$, then A^c must also be in $\mathcal{F}$ (closed under complementation). If A and B are in $\mathcal{F}$ (which we will write as $A, B \in \mathcal{F}$) then $A \cup B \in \mathcal{F}$.

Note: the properties of a field imply that $\mathcal{F}$ is also closed under finite intersection. (DeMorgan's law: $AB = (A^c \cup B^c)^c$)

Definition 2 A σ-field (or σ-algebra) of sets is a field that is also closed under *countable* unions (and intersections). □

What do we mean by countable?

- A set with a finite number in it is countable.
- A set whose elements can be matched one-for-one with $\mathbb{Z}$ is countable (even if it has an infinite number of elements!)

Are there non-countable sets?

Note: For any collection F of sets, there is a σ-field containing F, denoted by $\sigma(F)$. This is called the σ-field generated by F.

Definition of probability

We now formally define what we mean by a probability space. A probability space has three components.

The first is the **sample space**, which is the collection of all possible outcomes of some experiment. The outcome space is frequently denoted by Ω.

Example 1 Suppose the experiment involves throwing a die.

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

□

We deal with **subsets** of Ω. For example, we might have an event which is "all even throws of the die" or "all outcomes ≤ 4". We denote the collection of subsets of interest as $\mathcal{F}$. The elements of $\mathcal{F}$ (that is, the subsets of Ω) are called **events**. $\mathcal{F}$ is called the **event class**. *We will restrict $\mathcal{F}$ to be a σ-field.*

Example 2 Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, and let $\mathcal{F} = \{\{1, 2, 3, 4\}, \{5, 6\}, \{2, 4, 6\}, \{1, 3, 5\}, ...\}$ (What do we need to finish this off?) □

Example 3 $\Omega = \{1, 2, 3\}$. We could take $\mathcal{F}$ as the set of all subsets of Ω:

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$$

This is frequently denoted as $\mathcal{F} = 2^\Omega$, and is called the **power set** of Ω. □

Example 4 $\Omega = \mathbb{R}$. $\mathcal{F}$ is restricted to something smaller than all subsets of Ω. (So that probabilities can be applied consistently.) $\mathcal{F}$ could be the smallest σ-field which contains all intervals of $\mathbb{R}$. This is called the **Borel field**, $\mathcal{B}$. □

The tuple $(\Omega, \mathcal{F})$ is called a **pre-probability space** (because we haven't assigned probabilities yet). This brings us to the third element of a probability space: Given a pre-probability space $(\Omega, \mathcal{F})$, a **probability distribution** or a **measure** on $(\Omega, \mathcal{F})$ is a mapping P from $\mathcal{F}$ to $\mathbb{R}$ (which we will write: $P : \mathcal{F} \to \mathbb{R}$) with the properties:

- $P(\Omega) = 1$ (this is a normalization that is always applied for probabilities, but there are other measures which don't use this)
- $P(A) \geq 0 \forall A \in \mathcal{F}$. (Measures are nonnegative)
- If $A_1, A_2, \ldots \in \mathcal{F}$ such that $A_i A_j = \emptyset$ for all $i \neq j$ (that is, the sets are "disjoint" or "mutually exclusive") then

  $$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

  This is called the σ-additive, or additive, property.

These three properties are called the **axioms of probability**.

The triple $(\Omega, \mathcal{F}, P)$ is called a **probability space**.

- Ω tells what individual outcomes are possible
- $\mathcal{F}$ tells what sets of outcomes — events — are possible.
- P tells what the probabilities of these events are.

Some properties of probabilities (which follow from the axioms):

1. $P(A^c) = 1 - P(A)$.
2. $P(\emptyset) = 0$
3. $A \subset B \Rightarrow P(A) \le P(B)$.
4. $P(A \cup B) = P(A) + P(B) - P(AB)$
5. If $A_1, A_2, \ldots, \in \mathcal{F}$ then $P(\cup_{i=1}^\infty A_i) \le \sum_{i=1}^\infty P(A_i)$

There is also the **continuity of probability** property. Suppose $A_1 \subset A_2 \subset A_3 \cdots$ (this is called an increasing sequence). Define

$$\lim_{n\to\infty} A_n = \cup_{i=1}^\infty A_i \equiv A.$$

We write this as $A_n \uparrow A$ (A_n converges up to A). Similarly, if

$$A_1 \supset A_2 \supset A_3 \cdots$$

(a decreasing sequence), define

$$\lim_{n\to\infty} A_n = \cap_{i=1}^\infty A_i \equiv A_n \downarrow A.$$

If $A_n \uparrow A$, then $P(A_n) \uparrow P(A)$. If $A_n \downarrow A$, then $P(A_n) \downarrow P(A)$.
Example 5 Let $(\Omega, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$.

1. Take $A_n = (-\infty, 1/n]$, n=1,2,.... Then $A_n \downarrow (-\infty, 0]$. So

$$P((-\infty, 0]) = \lim_{n\to\infty} P((-\infty, 1/n]) = \lim_{x\to 0^+} P((-\infty, x)).$$

2. Take $A_n = (-\infty, -1/n]$, $n = 1, 2, \ldots,$. Then $A_n \uparrow (-\infty, 0)$.

$$P((-\infty, 0)) = \lim_{n\to\infty} P((-\infty, -1/n]) = \lim_{x\to 0^-} P((-\infty, x]).$$

□

We will introduce more properties later.

Some examples of probability spaces

1. $\Omega = \{w_1, w_2, \ldots, w_n\}$ (a discrete set of outcomes). $\mathcal{F} = 2^\Omega$. Let $p_1, p_2, \ldots, p_n$ be a set of nonnegative numbers satisfying $\sum_{i=1}^n p_i = 1$. Define the function $P : \mathcal{F} \to \mathbb{R}$ by

$$P(A) = \sum_{w_i \in A} p_i.$$

   Then $(\Omega, \mathcal{F}, P)$ is a probability space.

2. Uniform distribution: Let $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}_{[0,1]}$ =smallest σ-field containing all intervals in [0,1]. We can take (without proving that this actually works — but it does)

$$P([a, b]) = |b - a|$$

   for $0 \le a \le b \le 1$, and

$$P(\text{unions of disjoint intervals}) = \text{sum of probabilities of individual intervals.}$$

Conditional probability and independence

Conditional probability is perhaps the most important probability concept from an engineering point of view. It allows us to describe mathematically how our information changes when we are "given" a measurement.

We define conditional probability as follows: Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $A, B \in \mathcal{F}$ with $P(B) > 0$. Define the conditional probability of A given B as

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Essentially what we are saying is that the sample space is restricted from Ω down to B. Dividing by $P(B)$ provides the correct normalization for this probability measure.

Some properties of conditional probability (consequences of the axioms of probability) are as follows:

1. $P(A|B) \geq 0$.
2. $P(\Omega|B) = 1$
3. For $A_1, A_2, \ldots \in \mathcal{F}$ with $A_i A_j = \emptyset$ for $i \neq j$,
$$P(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B)$$
4. $AB = \emptyset \Rightarrow P(A|B) = 0$.
5. $P(B|B) = 1$
6. $A \subset B \Rightarrow P(A|B) \geq P(A)$
7. $B \subset A \Rightarrow P(A|B) = 1$.

Definition 3 $A_1, A_2, \ldots, A_n \in \mathcal{F}$ is a **partition** of Ω if $A_i A_j = \emptyset$ for $i \neq j$, and $\cup_{i=1}^{n} A_i = \Omega$, and $P(A_i) > 0$. □

Example 6 Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, and $A_1 = \{1\}$, $A_2 = \{2, 5, 6\}$, $A_3 = \{3, 4\}$. □

The Law of Total Probability: If $A_1, \ldots, A_n$ is a partition of Ω and $A \in \mathcal{F}$, then

$$P(A) = \sum_{i=1}^{n} P(A|A_i)P(A_i)$$

(Draw picture).

Bayes Formula is a simple formula for "turning around" the conditioning. Because conditioning is so important in engineering, Bayes formula turns out to be a tremendously important tool (even though it is very simple). We will see applications of this throughout the semester.

Suppose $A, B \in \mathcal{F}$, $P(A) \neq 0$ and $P(B) \neq 0$. Then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Why?

Definition 4 The events A and B are **independent** if

$$P(AB) = P(A)P(B).$$

□

Is independent the same as disjoint?

Note: For $P(B) > 0$, if A and B are independent then $P(A|B) = P(A)$. (Since they are independent, B can provide no information about A, so the probability remains unchanged. If $P(B) = 0$, then B is independent of A for any other event $A \in \mathcal{F}$. (Why?).

Definition 5 $A_1, \ldots, A_n \in \mathcal{F}$ are **independent** if for each $k \in \{2, \ldots, n\}$ and each subset $\{i_1, \ldots, i_k\}$ of $\{1, \ldots, n\}$,

$$P(\cap_{j=1}^k A_{i_j}) = \prod_{j=1}^k P(A_{i_j}).$$

□

Example 7 Take $n = 3$. Independent if:

$$P(A_1A_2) = P(A_1)P(A_2) \qquad P(A_1A_3) = P(A_1)P(A_3) \qquad P(A_2A_3) = P(A_2)P(A_3)$$

and

$$P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3).$$

□

The next idea is important in a lot of practical problem of engineering interest.

Definition 6 A_1 and A_2 are **conditionally independent** given $B \in \mathcal{F}$ if

$$P(A_1A_2|B) = P(A_1|B)P(A_2|B)$$

□

(draw picture to illustrate the idea).

Random variables

Up to this point, the outcomes in Ω could be anything: they could be elephants, computers, or mitochondria, since Ω is simple expressed in terms of sets. But we frequently deal with *numbers*, and want to describe events associated with sets of numbers. This leads to the idea of a random variable.

Definition 7 Given a probability space $(\Omega, \mathcal{F}, P)$, a **random variable** is a function X mapping Ω to $\mathbb{R}$. (That is, $X : \Omega \to \mathbb{R}$), such that for each $a \in \mathbb{R}$,

$$\{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F}.$$

□

A function $X : \Omega \to \mathbb{R}$ such that $\{w \in \Omega : X(\omega) \leq a\} \in \mathcal{F}$, that is, such that the events involved are in $\mathcal{F}$, is said to be **measurable** with respect to $\mathcal{F}$. That is, $\mathcal{F}$ is divided into sufficiently small pieces that the events in it can describe all of the sets associated with X.

Example 8 Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} = \{\{1, 2, 3\}, \{4, 5, 6\}, \emptyset, \Omega\}$. Define

$$X(\omega) = \begin{cases} 0 & \omega \text{ odd} \\ 1 & \omega \text{ otherwise} \end{cases}$$

Then for X to be a random variable, we must have

$$\{\omega \in \Omega | X(\omega) \leq a\}$$

to be an event in $\mathcal{F}$.

$$\{\omega \in \Omega | X(\omega) \leq a\} = \begin{cases} \emptyset & a < 0 \\ \{1,3,5\} & 0 \leq a < 1 \\ \Omega & a \geq 1 \end{cases}$$

So X is not a random variable — it not measurable with respect to $\mathcal{F}$. The field $\mathcal{F}$ is too "coarse" to measure if ω is odd.

On the other hand, let us now define

$$X(\omega) = \begin{cases} 0 & \omega \leq 3 \\ 1 & \omega > 3 \end{cases}$$

Is this a random variable? □

We observe that a random variable cannot generate partitions of the underlying sample space which are not events in the σ-field $\mathcal{F}$.

Another way of saying that X is measurable: For any $B \in \mathcal{B}$,

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

Recalling the idea of Borel sets $\mathcal{B}$ associated with the real line, we see that a random variable is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$:

$$X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}).$$

We will abbreviate the term random variable as **r.v.**.

We will use a notational shorthand for random variables.

Definition 8 Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$. For each $B \in \mathcal{B}$ we define

$$P(X \in B) = P(\{\omega \in \Omega : X(\omega) \in B\}).$$

□

By this definition, we can identify a new probability space. Using $(\mathbb{R}, \mathcal{B})$ as the pre-probability space, we use the measure

$$P_X(B) = P(X \in B).$$

So we get the probability space $(\mathbb{R}, \mathcal{B}, P_X)$.

As a matter of practicality, if the sample space is $\mathbb{R}$, with the Borel field, *most* mappings to $\mathbb{R}$ will be random variables.

To summarize:

$$(\Omega, \mathcal{F}, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}, P_X)$$

where

$$P_X(B) = P(\{\omega \in \Omega | X(\omega) \in B\})$$

for $B \in \mathcal{B}$.

Distribution functions

The **cumulative distribution function** (cdf) of an r.v. X is defined for each $a \in \mathbb{R}$ as

$$\begin{aligned} F_X(a) &= P(X \leq a) \\ &= P(\{\omega \in \Omega | X(\omega) \leq a\}) \\ &= P_X((-\infty, a]) \end{aligned}$$

Properties of cdf:

1. F_X is non-decreasing: If $a < b$ then $F_X(a) \leq F_X(b)$.
2. $\lim_{a\to\infty} F_X(a) = 1$
3. $\lim_{a\to-\infty} F_X(a) = 0$.
4. F_X is right-continuous: $\lim_{b\to a^+} F_X(b) = F_X(a)$.

Draw "typical" picture.

These four properties completely characterize the family of cdfs on the real line. Any function which satisfies these has a corresponding probability distribution.

5. For $b > a$: $P(a < X \leq b) = F_X(b) - F_X(a)$.
6. $P(X = a_0) = F_X(a_0) - \lim_{a\to a_0^-} F_x(a)$. Thus, if F_X is continuous at a_0, $P(X = a_0) = 0$.

From these properties, we can assign probabilities to all intervals from knowledge of the cdf. Thus we can extend this to all Borel sets.

Thus F_X determines a unique probability distribution on $(\mathbb{R}, \mathcal{B})$, so F_X and P_X are uniquely related.

Pure types of r.v.s

1. Discrete r.v.s — an r.v. whose possible values can be enumerated.
2. Continuous r.v.s — an r.v. whose distribution function can be written as the (regular) integral of another function
3. Singular, but not discrete — Any other r.v.

Discrete r.v.s

A random variable that can take on at most a *countable* number of possible values is said to be a discrete r.v.:

$$X : \Omega \to \{x_1, x_2, \ldots\}$$

Definition 9 For a discrete r.v. X, we define the probability mass function (pmf) (or discrete density function) by

$$p_X(a) = P(X = a) \qquad a \in \mathbb{R}$$

where $p_X(a) = 0$ if $a \neq x_i$ for any r.v. outcome x_i. □

Properties of pmfs:

1. Nonnegativity:

$$p_X(a) = \begin{cases} \geq 0 & a \in \{x_1, x_2, \ldots\} \\ 0 & \text{else} \end{cases}$$

2. Total probability:
$$\sum_{i=1}^{\infty} p_X(x_i) = 1.$$
(These two properties completely characterize the class of all pdfs on a given set $\{x_1, x_2, \ldots\}$.)

3. Relation to cdf:
$$p_X(a) = F_X(a) - \lim_{b \to a^-} F_X(b)$$
(To the pdf and the cdf contain the same information. Note that for a discrete r.v., the cdf is piecewise constant. Draw picture)

4. $F_X(a) = \sum_{\{x_i | x_i \leq a\}} p_X(x_i)$

Example 9 Bernoulli with parameter π, $(0 \leq \pi \leq 1)$.
$X \in \{0, 1\}$,
$$p_X(a) = \begin{cases} 1-\pi & a = 0 \\ \pi & a = 1 \\ 0 & \text{otherwise.} \end{cases}$$
$$F_X(a) = \begin{cases} 0 & a < 0 \\ 1-\pi & 0 \leq a \leq 1 \\ 1 & a \geq 1 \end{cases}$$
The Bernoulli is often used to model bits, bit errors, random coin flips, etc. □

Example 10 Binomial (n, π), $(0 \leq \pi \leq 1, n \in \mathbb{Z}^+)$.
$$X \in \{0, 1, \ldots, n\}.$$
$$p_X(a) = \begin{cases} \frac{n!}{(n-a)!a!}\pi^a(1-\pi)^{n-a} & a \in \{0, 1, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}$$
The binomial (n, π) can be viewed as the sum of n repeated Bernoulli trials. We can ask such questions as: what is the probability of k bits out of n being flipped, if the probability of an individual bit being flipped is π?

How do we show the total probability property? Use the **binomial theorem**:
$$\boxed{(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}}$$
(Plot the probability function.) □

Example 11 Poisson λ:
$$X : \Omega \to \mathbb{N}$$
$$p_X(a) = \begin{cases} \frac{\lambda^a e^{-\lambda}}{a!} & a \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$
The Poisson distribution is often used to model the number of rare events occurring in a length of time. (For example, what is the number of photon emmissions from a substance over a period of time. What is the number of cars passing on a road.) Later in the course we will clarify the assumptions and derive this expression.

How do we find $\sum_{k=0}^{\infty} P(X = k)$? □

Continuous r.v.s

A r.v. X is said to be continuous is there is a function $f_X : \mathbb{R} \to \mathbb{R}$ such that

$$F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx$$

for all $a \in \mathbb{R}$. In this case, $F_X(a)$ is an absolutely continuous function.[1]

Properties:

1. $f_X(x) \geq 0$
2. $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

These two properties completely characterize the f_Xs.

f_X called the **probability density function** (pdf) of X.

3. $P(X \in B) = \int_B f_X(x)\,dx = P_X(B)$, $B \in \mathcal{B}$.
4. $P(X = a) = 0$ for all $a \in \mathbb{R}$.
5. $P(X \in [x_0, x_0 + \Delta x] \approx f(x_0)\Delta x$.

$$P(X \in dx) = P_X(dx) = dP = f(x)dx.$$

Example 12 Uniform on (α, β) (with $\beta > \alpha$):

$$f_X(x) = \begin{cases} \frac{1}{\beta-\alpha} & \alpha < x < \beta \text{ (or } \alpha \leq x \leq \beta) \\ 0 & \text{otherwise.} \end{cases}$$

Plot pdf and cdf.

Uses: "random" number. Phase distribution. □

Example 13 Exponential(λ). $(\lambda > 0)$.

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Plot pdf and cdf.

Uses: Waiting time. (We'll see later what we mean by this.) □

Example 14 Gaussian (normal) $N(\mu, \sigma^2)$:

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-(x-\mu)^2/2\sigma^2)}$$

Plot.

Uses: Noise. "Sums of random variables."

This is the most important distribution we will deal with!

The cdf: Let $Z \sim N(0, 1)$. Define:

$$\Phi(x) = F_Z(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$$

We also use

$$Q(x) = 1 - \Phi(x) = P(Z > x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-z^2/2}\,dz$$

□

[1] A function F is said to be absolutely continuous if for every $\epsilon > 0$ there exists a δ such that for each finite collection of nonoverlapping intervals $[a_i, b_i]$, $i = 1, \ldots, k$, $\sum_{i=1}^{k} |F(b_i) - F(a_i)| < \epsilon$ if $\sum_{i=1}^{k} b_i - a_i < \delta$ [1, p. 433]. Being absolutely continuous is a stronger property than being continuous.

Properties of Gaussian Random Variables

Let $X \sim N(\mu, \sigma^2)$.

1. If $Y = \alpha X + \beta$ then $Y \sim N(\alpha\mu + \beta, \alpha^2\sigma^2)$
2. $(X - \mu)/\sigma \sim N(0, 1)$
3. $F_X(a) = \Phi((a - \mu)/\sigma)$.
4. $\Phi(-x) = 1 - \Phi(x)$.
5. A Gaussian r.v. is completely characterized by its first two moments (i.e., the mean and variance).

We will also see, and make use of, many other properties of Gaussian random processes, such as the fact that an uncorrelated Gaussian r.p. is also independent.

References

[1] P. Billingsley, *Probability and Measure*. New York: Wiley, 1986.