An abridged review of the basic theory of probability.

For: Students of Math 605, Stochastic Models in Biology, at the University of Wisconsin at Madison in the Fall semester of 2013.

By: David F. Anderson

Review of Proability

Probability theory is used to model experiments (defined loosely) whose outcome can not be predicted with certainty beforehand.

For any such experiment, there is a triple $(\Omega, \mathcal{F}, P)$, called a *probability space*, where

- Ω is the *sample space*,
- $\mathcal{F}$ is a collection of *events*,
- P is a *probability measure*.

We will consider each in turn.

The sample space Ω

The *sample space* of an experiment is the set of all possible outcomes.

Elements of Ω are called *sample points* and are often denoted by ω. Subsets of Ω are referred to as *events*.

Example

Consider the experiment of rolling a six-sided die. Then the natural sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. □

Example

Consider the experiment of tossing a coin three times. Let us write 1 for heads and 0 for tails. Then the sample space consists of all sequences of length three consisting only of zeros and ones. Each of the following representations is valid:

$$\begin{aligned}
\Omega &= \{0,1\}^3 \\
&= \{0,1\} \times \{0,1\} \times \{0,1\} \\
&= \{(x_1, x_2, x_3) : x_i \in \{0,1\} \text{ for } i = 1,2,3\} \\
&= \{(0,0,0),(0,0,1),(0,1,0),(1,0,0),(0,1,1),(1,0,1),(1,1,0),(1,1,1
\end{aligned}$$

The sample space Ω

Example

Consider the experiment of counting the number of mRNA molecules transcribed by a given gene in some interval of time. Here it is most natural to let

$$\Omega = \{0, 1, 2, \dots\}.$$

□

Example

Consider the experiment of waiting for a bacteria to divide. In this case, it is natural to take as our sample space all values greater than or equal to zero. That is,

$$\Omega = \{t \; : \; t \geq 0\},$$

where the units of t are specified as hours, for example. □

The sample space Ω

Terminology: A set that is finite or countably infinite is called *discrete*.

Most, though not all, of the sample spaces encountered in this course will be discrete.

The collection of events $\mathcal{F}$

- Events are simply subsets of the state space Ω.
- They are often denoted by A, B, C, etc., and they are usually the objects we wish to know the probability of.
- They can be described in words, or using mathematical notation.

Example 1, continued. Let A be the event that a 2 or a 4 is rolled. That is, $A = \{2, 4\}$. □

Example 2, continued. Let A be the event that the final two tosses of the coin are tails. Thus,

$$A = \{(1, 0, 0), (0, 0, 0)\}.$$

□

Example 3, continued. Let A be the event that no more than 10 mRNA molecules have appeared. Thus,

$$A = \{0, 1, 2, \ldots, 10\}.$$

□

The collection of events $\mathcal{F}$

For discrete sample spaces, $\mathcal{F}$ will contain all subsets of Ω, and will play very little role. This is the case for nearly all of the models in this course.

When the state space is more complicated, $\mathcal{F}$ is assumed to be a σ-algebra. That is, it satisfies the following three axioms:

1. $\Omega \in \mathcal{F}$.
2. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$, where A^c is the complement of A.
3. If $A_1, A_2, \dots, \in \mathcal{F}$, then

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

The probability measure P

Definition

The real valued function P, with domain $\mathcal{F}$, is a *probability measure* if it satisfies the following three axioms

1. $P(\Omega) = 1$.
2. If $A \in \mathcal{F}$, then $P(A) \geq 0$.
3. If for a sequence of events $A_1, A_2, \ldots,$, we have that $A_i \cap A_j = \emptyset$ for all $i \neq j$ (i.e. the sets are *mutually exclusive*) then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

The probability measure P

The following is a listing of some of the basic properties of any probability measure.

Lemma

Let $P(\cdot)$ be a probability measure. Then

1. *If $A_1, \ldots, A_n$ is a finite sequence of mutually exclusive events, then*
$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$
2. $P(A^c) = 1 - P(A)$.
3. $P(\emptyset) = 0$.
4. *If $A \subset B$, then $P(A) \leq P(B)$.*
5. $P(A \cup B) = P(A) + P(B) - P(AB)$.

Conditional probability and independence

Definition

For two events $A, B \subset \Omega$, the *conditional probability of A given B* is

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

provided that $P(B) > 0$.

Independence

We intuitively think of A being independent from B if

$$P(A|B) = P(A), \quad \text{and} \quad P(B|A) = P(B).$$

More generally, we have the following definition.

Definition

The events $A, B \in \mathcal{F}$ are called *independent* if

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

It is straightforward to check that the definition of independence implies both

$$P(A|B) = P(A), \quad \text{and} \quad P(B|A) = P(B).$$

Random variables

Recall that for some probability space $(\Omega, \mathcal{F}, P)$, a random variable X is a real-valued function defined on the sample space Ω.

That is,

$$X : \Omega \to \mathbb{R}$$

and $X(\omega) \in \mathbb{R}$ for each $\omega \in \Omega$.

The values that a random variable can take (i.e. its *range*) is called its *state space*.

- If the state space of X is finite or countably infinite, then X is said to be a *discrete random variable*.
- Otherwise X is said to be a *continuous random variable*.

Random variables

Example

Suppose we roll two die and take $\Omega = \{(i, j) \mid i, j \in \{1, \ldots, 6\}\}$. We let

$$X(i, j) = i + j$$

be the discrete random variable giving the sum of the rolls. The range is $\{2, \ldots, 12\}$. $\Box$

Random variables

We can now ask for probabilities associated with the random variable X. For example, we may be interested in

$$P\{\omega : 1 \leq X(\omega) \leq 6\}. \tag{1}$$

Note that we are still asking for a probability of a subset of Ω, that is, an element of $\mathcal{F}$.

We often do not care about the probability space itself and will simply write

$$P\{1 \leq X \leq 6\},$$

where it is understood that the equation above is shorthand for (1).

Random variables

A stochastic process $\{X_t\}$, $t \in I$, is a collection of random variables defined on a common probability space, and where I is some index set.

As in the case of a random variable, it is technically correct to write things like $X_t(\omega)$, where $\omega \in \Omega$, the common probability space, and

$$P\{\omega : X_{t_1}(\omega) \in A_1, X_{t_2}(\omega) \in A_2\}.$$

However, we will most often instead simply write X_t and

$$P\{X_{t_1} \in A_1, X_{t_2} \in A_2\}$$

and largely ignore the probability space upon which the process is defined.

Transformations of random variables

Theorem

Let U be uniformly distributed on the interval $(0, 1)$ and let F be an invertible distribution function. Then $X = F^{-1}(U)$ has distribution function F.

Proof.

Letting $X = F^{-1}(U)$ where U is uniform$(0, 1)$, we have

$$\begin{aligned} P\{X \le t\} &= P\{F^{-1}(U) \le t\} \\ &= P\{U \le F(t)\} \\ &= F(t). \end{aligned}$$

□

Transformations of random variables

Example

Suppose we want an $\exp(\lambda)$ random variable.

Has distribution function $F : \mathbb{R}_{\geq 0} \to [0, 1)$

$$F(t) = 1 - e^{-\lambda t}, \quad t \geq 0.$$

Therefore, $F^{-1} : [0, 1) \to \mathbb{R}_{\geq 0}$ is given by

$$F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u), \quad 0 \leq u \leq 1.$$

If U is uniform$(0, 1)$, then so is $1 - U$. Thus, to simulate a realization of $X \sim \text{Exp}(\lambda)$, you first simulate U from uniform$(0, 1)$, and then set

$$x = -\frac{1}{\lambda} \ln(U) = \ln(1/U)/\lambda.$$

□

Transformations of random variables

Theorem

Let U be uniformly distributed on the interval $(0,1)$. Suppose that $p_k \geq 0$ for each $k \in \{0,1,\ldots,\}$, and that $\sum_k p_k = 1$. Let

$$X = \min\left\{k \;\middle|\; \sum_{i=0}^{k} p_i \geq U\right\}.$$

Then,

$$P\{X = k\} = p_k.$$