Northwestern University
Department of Electrical Engineering and Computer Science

EECS 454: Modeling and Analysis of Communication Networks — Spring 2008

Probability Review

As discussed in Lecture 1, probability theory is useful for modeling a variety of sources of uncertainty in a communication network. Examples include when a packet arrives at a point in the network, whether a packet is received correctly, the size of a packet, and the destination for a packet. These notes will review some basic facts from probability theory that will be useful during the course.

1 Basic Probability Theory

A probabilistic model can be thought of as describing an experiment with several possible outcomes. Formally, this consists of a **sample space** Ω, which is the set of all possible outcomes, and a **probability law** that assigns a probability $P(A)$ to each *event*, where an event is a subset of the sample space.[1] $P(\cdot)$ can be viewed as a real-valued function whose range is the set of all possible events.

A probability law must satisfy the following three properties:

- *Non-negativity*: $P(A) \geq 0$ for every event A.
- *Additivity*: If A_i's are all disjoint events, $P\left(A_1 \cup A_2 \cup ...\right) = \sum P(A_i)$.
- *Normalization*: The probability of the union of all possible events is 1, i.e., $P(\Omega) = 1$.

For example, a probabilistic model might represent the length of a packet sent over a network. In this case, the sample space will be the set of possible packet lengths, say $\{l_1, l_2, \ldots, l_m\}$ and the $P(l_i)$ would indicate the likelihood a packet has the length l_i.

1.1 Conditional Probability

The **conditional probability** of an event A occurring, given that an event B has occurred, is denoted by $P(A|B)$; this is computed as

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

Under the assumption that $P(B) > 0$ (since we know that B has occurred).

[1] This definition is adequate for discrete sample spaces. For infinite sample spaces, more care is required to be mathematically precise. In particular, a probability $P(A)$ can only be assigned to so-called *measurable events* A; such issues need not concern us in this course.

1.1.1 Total Probability Theorem

Let $A_0, ..., A_n$ be disjoint events that form a partition of the sample space (i.e. each possible outcome is included in one and only one of the events $A_1, ..., A_n$) and assume that $P(A_i) > 0$ for all $i = 1, .., n$. For any event B, we have

$$\begin{aligned} P(B) &= P(A_1 \cap B) + ... + P(A_n \cap B) \\ &= P(A_1)P(B|A_1) + ... + P(A_n)P(B|A_n), \end{aligned}$$

where the first line follows from additivity and the second line is the definition of conditional probability. This is sometimes refereed to as the total probability theorem; often one is interested in calculating one of the conditional probabilities on the right-hand side, and can use this relation to accomplish this.

1.1.2 Independence

Events A and B are defined to be **independent events** if and only if,

$$P(A \cap B) = P(A)P(B).$$

Using the definition of conditional probability, it can be seen that this is equivalent to:

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

Otherwise, the two events are said to be **dependent**. Events being independent means that one occurring does not effect the probability of the other occurring. This can be generalized for any set of n events, i.e. the set of events $A_1, \ldots, A_n$ are independent if and only if,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

2 Random Variables

For a given probability model, a **random variable** is a function $X : \Omega \to \mathbb{R}$, where this notation means that the domain of X is the sample space Ω and its range is the real-line.[2] The probability that a random variable X takes values in a subset A of $\mathbb{R}$ is given by

$$P(X \in A) = P(\omega \in \Omega : X(\omega) \in A),$$

i.e., the probability of those points in the sample space that map into A. Often, we omit references to the underlying sample space and simply work with the random variable. A random variable is said to be **discrete** if it takes on a finite or countable number of values. A **continuous random variable** can take an uncountable number of values, e.g., any value on the real line, or possibly a sub-interval of the real line.

[2] In most engineering texts, random variables are defined to be real-valued, as we have done here. In advanced probability courses often any set is allowed as the range of a random variable. For our purposes, this level of generality is not needed.

2.1 Discrete Random Variables

For a discrete random variable X, the **probability mass function (PMF)** gives the probability that X will take on a particular value in its range. We denote this by p_X, i.e.

$$p_X(x) = P(\{X = x\}),$$

when it is clear from the context, the sub-script X may be omitted. Since P is a probability law, it follows that $p_X(x)$ is non-negative and sums to one, i.e.,

$$\sum_i p_X(x_i) = 1,$$

where $\{x_i\}$ are the values which X takes on (i.e., the set of values so that $X(\omega) = x_i$ for some $\omega \in \Omega$).

2.1.1 Expectation

The **expected value** of a discrete random variable X is defined by

$$E[X] = \sum_i x_i p_X(x_i).$$

Let $g(X)$ be a real-valued function of X, the expected value of $g(X)$ is calculated by

$$E[g(X)] = \sum_i g(x_i) p_X(x_i).$$

This is also called the mean or average of X.

When $g(X) = (X - E(X))^2$, the expected value of $g(X)$ is called the variance of X and denoted by σ_X^2, i.e.

$$\sigma_X^2 = E(X - E(X))^2.$$

Next we discuss a few common discrete random variables:

2.1.2 Bernoulli Random variable with parameter p

X is a Bernoulli random variable with parameter p if it can take on values 0 and 1 with

$$p_X(1) = p$$

$$p_X(0) = 1 - p$$

Bernoulli random variables provide a simple model for an experiment that can either result in a success (1) or a failure (0).

2.1.3 Binomial Random Variable with parameters p and n

This is the number S of successes out of n independent Bernoulli random variables

The PMF is given by

$$p_S(k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k},$$

for $k = 0, 1, .., n$. The expected number of successes is given by

$$E[S] = np.$$

For example, if packets arrive correctly at a node in a network with probability p (independently); then the number of correct arrivals out of n is a Binomial random variable.

2.1.4 Geometric Random Variable with parameter p

Given a sequence of independent Bernoulli random variables, let T be the number observed up to and including the first success. Then T will have a geometric distribution; its PMF is given by

$$p_T(t) = (1-p)^{t-1} p$$

for $t = 1, 2, ...$; the expectation is

$$E[T] = \frac{1}{p}.$$

2.2 Poisson random variable with parameter μ

A discrete random variable, N is said to have Poisson distribution with parameter μ if

$$p_N(n) = \frac{(\mu)^n}{n!} e^{-\mu}, \quad n = 0, 1, 2, \ldots$$

We verify that this is a valid PMF, i.e. that

$$\sum_{n=0}^{\infty} p_N(n) = 1.$$

This can be shown as follows:

$$\begin{aligned}
\sum_{n=0}^{\infty} p_N(n) &= \sum_{n=0}^{\infty} \frac{(\mu)^n}{n!} e^{-\mu} \\
&= e^{-\mu} \sum_{n=0}^{\infty} \frac{(\mu)^n}{n!} \\
&= e^{-\mu} e^{\mu} \\
&= 1
\end{aligned}$$

Here we have used that $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ...$

The expected value of a Poisson random variable is

$$\begin{aligned} E[N] &= \sum_{n=0}^{\infty} n \frac{(\mu)^n}{n!} e^{-\mu} \\ &= (\mu) e^{-\mu} \sum_{n=1}^{\infty} \frac{(\mu)^{n-1}}{(n-1)!} \\ &= (\mu) e^{-\mu} e^{\mu} \\ &= \mu \end{aligned}$$

As we will see, Poisson random variables are often used for modeling traffic arrivals in a network. For example in the telephone network, the number of calls that arrive in an interval of T seconds is well modeled by a Poisson random variable with parameter λT, where λ is the call arrival rate.

2.3 Continuous Random Variables

For a continuous random variable X, a **probability density function (PDF)** f_X is a non-negative function[3] such that for all $a < b$,

$$P(a < X \leq b) = \int_a^b f_X(x)dx.$$

and

$$\int_{-\infty}^{+\infty} f_X(x)dx = P(-\infty < X < +\infty) = 1.$$

For a continuous random variable X (with a PDF) and for any value a,

$$P(X = a) = \int_a^a f(x)dx = 0.$$

This implies that

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b).$$

2.3.1 Expectation

The expectation of a continuous random variable is defined as

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)dx.$$

Again, for a real-valued function $g(X)$ we have

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x)dx.$$

[3] Not all continuous random variables have a probability density, but this will be the case for those considered here.

2.3.2 Exponential Random Variable

As an example of a continuous random variable, we consider an exponential random variable. These random variables are also used for modeling traffic in a network, for example to model the time between packet arrivals. An exponential random variable has a PDF of the form

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

where $\lambda > 0$. Verify that for an exponential random variable,

$$E[X] = \frac{1}{\lambda}.$$

2.4 Cumulative Distribution Functions

The **cumulative distribution function (CDF)** of a random variable X is the probability $P(X \leq x)$, denoted by $F_X(x)$.

If X is a discrete random variable, then we get

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} p_X(k)$$

And similarly if X is a continuous random variable we get

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$$

In this case we can therefore define f_X in terms of F_X:

$$f_X(x) = \frac{dF_X(x)}{dx},$$

i.e., the PDF of a continuous random variable is the derivative of its CDF.

Example: The CDF of an exponential random variable is

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t)dt \\ &= \int_{0}^{x} \lambda e^{-\lambda t}\, dt \\ &= 1 - e^{-\lambda x}, \quad x \geq 0. \end{aligned}$$

2.5 Conditional PDF

The **conditional PDF** $f_{X|A}$ of a continuous random variable X given an event A with $P(A) > 0$, is defined as

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(A)} & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

So that,

$$P(X \in B|A) = \int_B f_{X|A}(x)dx.$$

An analogous definition holds for discrete random variables. The conditional expectation is defined by

$$E[X|A] = \int_{-\infty}^{+\infty} x f_{X|A}(x)dx.$$

In this case, the total probability theorem can be restated as

$$f_X(x) = \sum_{i=1}^{n} P(A_i) f_{X|A_i}(x),$$

where $\{A_1, \ldots, A_n\}$ are disjoint events that cover the sample space. It follows that,

$$E[X] = \sum_{i=1}^{n} P(A_i) E[X|A_i],$$

this is often a useful way to calculate an expectation.

Example: Let X be an exponential random variable, and A the event that $X > t$. Then $P(A) = e^{-\lambda t}$ and

$$f_{X|A}(x) = \begin{cases} \lambda e^{-\lambda(x-t)} & x \geq t \\ 0 & otherwise \end{cases}$$

From this it follows that

$$P\{X > r + t \mid X > t\} = P\{X > r\}, \qquad r, t \geq 0.$$

This is an important property of an exponential random variable called the memoryless property.

3 Stochastic Processes

A **stochastic process** is a sequence of random variables indexed by time. Stochastic processes are used, for example to model arrivals of packets in a network. In this case $\{A(t), t \geq 0\}$ could denote the total number of packets to arrive at a node up to time t. For each time t, the quantity $A(t)$ is a random variable. Stochastic processes can also be defined in discrete time. For example, let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) Bernoulli random variables. This is a discrete-time stochastic process called a Bernoulli process.

A given outcome of each random variable comprising a stochastic process is referred to as a **realization** or a **sample path** of the process.

4 Law of large numbers

Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed (i.i.d.) random variables, each with expected value $\bar{X}$. The **strong law of large numbers** states that with probability one,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i = \bar{X}.$$

Basically what this says is that if we observe a long enough sequence of outcomes of these random variables and take the arithmetic average of these outcomes, this will converge to the expected value of each outcome.

When each X_i is a Bernoulli random variable, the law of large numbers can be interpreted as stating that the long-run fraction of successes will be equal to EX_i.

We can also view the sequence $X_1, X_2, \ldots$ as a discrete time stochastic process (note that this is a very special process in that each random variable is i.i.d., which in general is not required for a stochastic process). In this case, we can view the quantity $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X_i$ as the *time-average* of a given realization of the stochastic process. On the other hand, the quantity $\bar{X}$ is a *statistical average*, i.e. it is calculated by averaging a particular value of X over all possible realizations. The strong law of large numbers states that we can replace the time average with the statistical average for a process made up of i.i.d. random variables. This is useful often useful because for a given model statistical averages are easier to calculate, but in practice time-averages are what are important (i.e. you only see one realization). In many cases where stochastic processes are not made up of i.i.d. random variables, we can still replace time-averages with statistical averages (such processes are referred to as *ergodic*). In most cases, the processes we deal with here will be ergodic.

Appendix: Useful Results

The following are some results are useful for manipulating many of the equations that may arise when dealing with probabilistic models.

4.1 Geometric Series

For $x \neq 1$,

$$\sum_{k=0}^{n} x^k = \frac{x^{n+1} - 1}{x - 1},$$

and when $|x| < 1$,

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1 - x}.$$

Differentiating both sides of the previous equation yields another useful expression:

$$\sum_{k=0}^{\infty} k x^k = \frac{x}{(1 - x)^2}$$

4.2 Exponentials

The Taylor series expansion of e^x is:

$$e^x = 1 + x + \frac{x^2}{2!} + ... = \sum_{i=0}^{\infty} \frac{x^i}{i!},$$

and

$$\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x.$$