Northwestern University
Department of Electrical and Computer Engineering

ECE 333: Introduction to Communication Networks Fall 2002

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 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[1] A (a set of possible outcomes). The

The following properties must hold:

- *Non-negativity*: $P(A) \geq 0$ for every event A.
- *Additivity*: If A_i's are all disjoint events, $P(A_1 \cup A_2 \cup ...) = \sum P(A_i)$.
- *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 of course that $P(B) > 0$ (since we know that B has occurred.)

[1] This definition is adequate for discrete sample spaces. For infinite sample spaces, slightly more care is required to be mathematically precise; such concerns are beyond the scope of this course.

1.1.1 Total Probability Theorem

Let $A_0, ..., A_n$ be disjoint events that form a partition of the sample space (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 $i = 1, .., n$, then for any event B, we have

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

1.1.2 Independence

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

$$\begin{aligned} P(A \cap B) &= P(A)P(B) \\ P(A|B) &= P(A) \end{aligned}$$

Otherwise, the two events are said to be dependent. 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,

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

2 Random Variables

For a given probability model, a random variable is a function whose domain is the sample space Ω. A discrete random variable takes on a finite or countable number of values. A continuous random variable can take any value on the real line (or possibly a subinterval of the real line).

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.

$$P_X(x) = P(\{X = x\}).$$

2.1.1 Expectation

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

$$E[X] = \sum x P_X(x).$$

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

$$E[g(X)] = \sum g(x) P_X(x).$$

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

$$\sum_{n=0}^{\infty} P_n = 1.$$

This can be shown as follows:

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

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

The expected value of a Poisson random variable is

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

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 can be 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 such that

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

and

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

For a continuous random variable X 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

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

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

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

So that,

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

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

$$f_X(x) = \sum_{i=0}^{n} P(A_i) f_{X|A}(x).$$

It follows that,

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

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 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.

4 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 and 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 Bernoulli random variables. This is a discrete-time stochastic process called a Bernoulli process.

5 Useful Results

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

5.1 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:

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

5.2 Exponentials

The Taylor series expansion of e^x is:

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

and

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