Recitation 2, Probability Review

Paolo Bertolotti

1 Elementary Probability

Probability allows us to study the likelihood of an event's occurrence. Consider rolling a fair die. We do not know what number will be rolled with certainty but we know there is a chance of rolling a 4. Before assigning likelihoods to values, we define the basic underlying framework.

- **Experiment:** an action where the result is uncertain
- **Sample space:** the set of all possible outcomes of an experiment
- **Event:** a subset of the sample space

In our die example, the experiment is rolling a die once. The sample space is {1, 2, 3, 4, 5, 6}. Rolling a 4 would be an example of an event.

Given a sample space S, the **probability** P is a function from the space of events in S to the interval [0, 1]. It satisfies the following properties:

1. Countable additivity: For any sequence A_i of events in S such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. In words, the probability of any disjoint event occurring in some sequence is equal to the sum of their individual probabilities.
2. Normalization: P(S) = 1

Question: If $A \in S$ with probability $P(A)$, what is the probability of the event A_c, A's complement? By their definition, A and A_c satisfy $A \cup A_c = S$ and $A \cap A_c = \emptyset$. Therefore, we have $1 = P(S) = P(A \cup A_c) = P(A) + P(A_c)$ and $P(A_c) = 1 - P(A)$. For example, the probability of not rolling a 4 is equal to 1 minus the probability of rolling a 4.

Given events A and B where $P(B) > 0$, the **conditional probability** of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1}$$

Example: Consider a fair die, where fair means every outcome in {1, 2, 3, 4, 5, 6} happens with equal probability 1/6. Define the event $A = \{4or5\}$. What is $P(A)$? Simple, $P(A) = 1/6 + 1/6 = 1/3$. Now, what happens if we have additional information regarding the toss.

- Case 1: If B_1 = "the outcome is an even number", then $P(A|B_1) = 1/3$
- Case 2: If B_2 = "the outcome is larger than 3", then $P(A|B_2) = 2/3$

- Case 3: If B_3 = "the outcome is less or equal to 3", then $P(A|B_3) = 0$

Two events are **independent** iff $P(A \cap B) = P(A)P(B)$. Independence implies $P(A) = P(A|B)$ assuming $P(B) > 0$, which means that B's occurrence provides no information about A. Independence simplifies many calculations. Two important theorems using the concept of conditional expectation are:

Theorem 1 (Law of Total Probability) If A is an event and B_i is a sequence of n events that partitions the sample space (meaning they are all disjoint and their union equals the sample space), then

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i). \tag{2}$$

Theorem 2 (Bayes' Theorem) For events A and B with $P(B) > 0$,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A_c)P(A_c)} \tag{3}$$

where the second equality follows from the Law of Total Probability.

2 Random Variables

A **random variable** (rv) is a function $X : S \to \mathbb{R}$ that assigns a real number to each event in the sample space. For example, consider an experiment where we toss a coin ten times. The sample space is the collection of all possible combinations of H and T of size 10. A possible event is $s = \{HHHTHHHTTT\} \in S$. We can define the random variable Y that counts the number of heads. In our example, $Y(s) = 6$.

A random variable is called **discrete** if it takes on at most a countable set of values. For every discrete random variable X we define the **probability mass function** (pmf) of X by

$$p_X(x) = P(\{s \in S : X(s) = x\}) \tag{4}$$

We usually omit the argument of the rv X and simply write

$$p_X(x) = P(X = x) \tag{5}$$

Assume the random variable X takes values in the set $a_1, a_2, ..., a_n$ and the random variable Y takes values in the set $b_1, b_2, ..., b_m$. We say that X and Y are **independent random variables** if

$$P(\{X = a_i\} \cap \{Y = b_j\}) = P(X = a_i)P(Y = b_j) \tag{6}$$

for every $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$.

All random variables have a **cumulative distribution function** F, which is defined as

$$F_X(x) = P(X \leq x) \tag{7}$$

Continuous random variables, which can take a uncountably infinite set of values, do not have a pmf. Instead, they have a **probability density function** (pdf) f, defined as

$$f_X(x) = \frac{d}{dx}F(x) \tag{8}$$

where

$$P(x \in A) = \int_A f_X(x)dx \tag{9}$$

2.1 Expectation and Variance

The **expected value** of a random variable X, also called its mean, is the probability-weighted average value for the variable. It is defined as

$$E[X_{continuous}] = \int x f_X(x)dx \tag{10}$$

$$E[X_{discrete}] = \sum_i x_i\, p_X(x_i) \tag{11}$$

For independent random variables X and Y, $E[XY] = E[X]E[Y]$.

The **variance** of a random variable X measures how dispersed it is around its mean. In a sense, it captures how variable it is.

$$Var(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2 \tag{12}$$

The **covariance** of two random variables X and Y measures how much they co-vary or co-move. I.e., if X goes up, does Y go up or down. It is defined as

$$cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] \tag{13}$$

Theorem 3 (Variance of Sums) For any sequence X_i of random variables

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + \sum_{i \neq j} cov(X_i, X_j) \tag{14}$$

Remark: As a special case, when X_i and X_j are independent for each $i \neq j$, the second term above vanishes and

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) \tag{15}$$

2.2 Common Distributions

A quick reminder of some useful random variables and their distributions.

- **Bernoulli:** Models a biased coin toss with bias parameter p. Random variable X takes value 1 with probability p and value 0 with probability $1 - p$

- **Binomial:** Models the sum of outcomes of n biased coin tosses with bias parameter p. Random variable X takes the values in $\{0, ..., n\}$ with pmf $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$
- **Geometric:** Models the number of tosses until a heads appears in biased coin tosses with bias parameter p (giving the probability of heads). I.e. the number of trials until the first success. Random variable X takes the values in $\{1, ...\}$ with pmf $p_X(x) = (1-p)^{x-1} p$
- **Poisson:** One form of continuum limit for the binomial distribution when p becomes small and n becomes large. Governed by an intensity / rate parameter λ. Random variable X takes values in $\{0, ...\}$ with pmf $p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

[1]

[1] The skeleton for these notes was provided by the recitation notes for MIT 6.268 Network Science and Models

1.022 Introduction to Network Models
Fall 2018