**School of Oriental and African Studies** **September 2013**
Department of Economics

# Preliminary Statistics
# Lecture 2: Probability Theory (Outline)

Gujarati D. (2003) *Basic Econometrics*, Appendix A.2 and A.3
Barrow M. (2009) *Statistics for Economics, Accounting and Business Studies* Ch. 2

## 1 Introduction

We need to analyse cases where we do not know what is going to happen: where there are risks, randomness, chances, hazards, gambles, etc. Probabilities provide a way of doing this. Probabilities are also useful for statistical inference, since we will need to make some probabilistic statements to infer about the population (random variable) when using a sample (just a subset of the values of the random variable).

## 2 Approaches of Probability

There are two main approaches to Probability Theory, the frequentist view and the subjective view.

- Relative frequency: Empirical definition of probability

$$P(A) = \frac{m}{n} = \frac{\text{Number of Outcomes favourable to } A}{\text{Total Number of Observations (trials)}}$$

- Subjective probability: Bayesian definition of probability

  Probability as a degree of belief

## 3 Terminology

An experiment (or trial) is a process leading to a unique outcome (numerical value) with uncertainty. An outcome is the result of an experiment. For example,

| Experiment | Outcome |
|---|---|
| Tossing a coin | Head (h) or Tail (T) |
| Tossing two coins | HH, HT, TH, TT |
| Throwing a die | 1, 2, 3, 4, 5, 6 |

All possible outcomes of an experiment define the sample space or the population. Each member (outcome) of the sample space is called a sample point. The collection of the possible outcomes of an experiment is called an event. An event is a subset of the sample space.

- Events are *mutually exclusive* if the occurrence of one event prevents the occurrence of another event at the same time.

- Two events are said to be *equally likely* if we are confident that one event is as likely to occur as the other event.

- Events are said to be *collectively exhaustive* if they exhaust all possible outcomes of an experiment.
- Two events are said to be *independent* if the occurrence of the one does not depend on the other.

## 4 Probability Axioms and Rules

Probabilities are numbers which represent the chance of an event happening. Denote the probability of event $A$ happening as $P(A)$ and the probability of event $B$ happening as $P(B)$.

- The probability of an event always lies between 0 and 1, $0 \leq P(A) \leq 1$. $P(A)$ is often known as the marginal probability.
- The probability of event A not happening is $1-P(A)$. This is called the complement of A, sometimes denoted as $\overline{A}$. It holds that $P(A) + P(\overline{A}) = 1$.
- The probability of two events, $A$, $B$ both happening, $A$ **AND** $B$, is denoted as $P(A \cap B)$, the *intersection*, and is often known as the joint probability. It is the probability of observing both events happening together.
- For two mutually exclusive events, A and B, their intersection is equal to zero, by definition. So, $P(A \cap B) = 0$.
- Denote the probability of event $A$ **OR** event $B$ happening as $P(A \cup B)$ (the union). It is the probability of one event or the other happening.
- The union of two events, $A$ and $B$ is calculated as

  $$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

  Notice that if the events are mutually exclusive, $P(A \cap B) = 0$, so $P(A \cup B) = P(A) + P(B)$.
- Two events are said to be independent if the probability of the one occurring does not depend on the other one happening. For two independent events, $A$ and $B$, the joint probability, i.e. the probability of both happening, is equal to the product of the two marginal probabilities:

  $$P(A \cap B) = P(A) \times P(B).$$
- The probability of $A$ happening given that event $B$ has happened is called a conditional probability and is given by:

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

  Conditional probabilities play a very important role in decision making. They ask how the information that $B$ happened changes your estimate of the probability of $A$ happening. If A and B are independent, $P(A \mid B) = P(A)$. By definition, knowing

that $B$ happened does not change the probability of $A$ happening. Similarly, the probability of B happening given that A happens is:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \tag{2}$$

Multiply both sides of (1) by $P(B)$ and both sides of (2) by $P(A)$, and rearrange to give

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

The joint probability is the product of the conditional probability and the marginal probability in each case. Using the two right hand side relations gives Bayes Theorem:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

This formula is widely used to update probabilities of an event $A$, in the light of new information, $B$. In this context $P(A)$ is called the prior probability of $A$, $P(B \mid A)$ is called the likelihood, and $P(A \mid B)$ is called the posterior probability.

- We can use the joint probability to redefine the marginal probability of an event $A$ happening with reference to event $B$:

  $$P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

  i.e. the probability of event $A$ happening is irrespective of whether event $B$ happens or not.

# 5 Random Variables

We now turn to consider how we apply probability to variables where we are uncertain what values they will take. These are called *random variables*. Usually, we denote random variables with capital letters, $X, Y, Z...$, and their realisations (particular values) with small letters $x_i, y_j, z_k$.

## 5.1 Discrete Random Variables

A discrete random variable, $X$ can take a number of distinct possible values, say $x_1, x_2, .., x_N$ with probabilities $p_1, p_2, ..., p_N$. The observed values are called the realisations of the random variable. For instance, $X$ the total obtained from throwing two dice is a discrete random variable. It can take the values 2 to 12. After you throw the dice, you observe the outcome, the realisation, a particular number, $x_i$. Associated with the random variable is a probability distribution, $p_i = f(x_i)$, which gives the probability of obtaining each of the possible outcomes the random variable can take. Formally, the probability distribution is

$$\begin{gathered} f(x_i) = P(X = x_i) = \text{the probability that the} \\ \text{random variable} \quad X \text{ takes the value } x_i, \text{ e.g. the value 3.} \end{gathered}$$

The cumulative probability distribution gives the probability of getting a value less than or equal to $x_j$. That is,

$$F(x_j) = \sum_{i=1}^{j} f(x_i) = P(X \leq x_j)$$

### 5.2 Continuous Random Variables

Whereas a discrete random variable can only take specified values, continuous random variables (e.g. inflation) can take an infinite number of values. Corresponding to the probabilities $f(x_i)$ for discrete random variables there is a probability density function, *PDF*, also denoted $f(x_i)$ for continuous random variables. Since there are an infinite number of points on the real line, the probability of any one of those points is zero, although the *PDF* will be defined for it. However, we can always calculate the probability of falling into a particular interval. Hence, if we want to calculate the probability that $X$ lies between $a$ and $b$, then

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

Similarly, the cumulative density function, *CDF*, $F(x_j) = \Pr(X \leq x_j)$ gives the probability that the continuous random variable will take a value less than or equal to a specified value $x_j$.

$$F(x_j) = P(X \leq x_j) = \int_{-\infty}^{x_j} f(x_i)dx_i$$

### 5.3 Marginal, Joint and Conditional Distributions

Suppose we have two random variables $X$ and $Y$ with individual probability density functions $f(x_i)$ and $f(y_i)$. $f(x_i)$ is also known as the marginal $PDF$ of $X$, the probability distribution that $X$ assumes a given value, for all values of $X$, regardless of the values taken by $Y$.

The joint probability of $X$ and $Y$, $f(x_i, y_i)$, indicates the probability of both $X$ taking a particular value, $x_i$, and $Y$ taking a particular value, $y_i$, and corresponds to $P(A \cap B)$ (intersection) in terms of events.

We can now define the marginal probability distribution for the random variable $X$ with reference to the joint distribution:

$$f(x_i) = \sum_{j=1}^{\infty} f(x_i, y_j) = f(x_i, y_1) + f(x_i, y_2) + ...$$

The conditional $PDF$ of $X$, gives the probability that $X$ takes on the value of $x_i$ given that $Y$ has assumed the value $y_i$.

$$f(x_i|y_i) = \frac{f(x_i, y_i)}{f(y_i)}$$

### 5.4 Statistical Independence

If two random variables, $X, Y$, are independent, then for all values of $X$ and $Y$, their joint probability is just the product of the individual probability distributions,

$$f(x_i, y_i) = f(x_i)f(y_i).$$