Fundamentals of Business Statistics – PT IMBA 2007/8

PART B

Language of Probability

& Random Variables

Chapters 5 & 6

Course Outline

Graphs and tables	Chapter 2	Part A
Summary measures	Chapter 3	Part A
Probability and probability distributions	Chapter 5	Part B
Normal, binomial, Poisson, and exponential distributions	Chapter 6	Part B
Sampling & statistical inference	Chapters 8-11	Part C

Part A: Describe (Summarize) Data in Samples
Part B: Language of Probability & Random Variables
Part C: Sampling & Statistical Inference

"We may at once admit that any inference from the particular to the general must be attended with some degree of uncertainty, but this is not the same as to admit that such inference cannot be absolutely rigorous, for the nature and degree of the uncertainty may itself be capable of rigorous expression,"
R.A. Fisher in *The Design of Experiments*

Overview of Part B – Generics

- **Probability essentials**
  - What is a probability?
  - Where do probabilities come from?
  - How do we calculate complex probabilities from simpler (available, known) ones?
- Random variables
  - What is a random variable?
  - What is a probability distribution?
  - Measures of central location and variability
  - Joint probability distributions

Definition of Probability

- Concept of probability is quite intuitive; however, the rules of probability (cf. later) are not always intuitive or easy to master.
- Mathematically, a probability is a number between 0 and 1 that measures the **likelihood** that some **event** will occur.
  - An event ***A*** with probability **0** cannot occur ($P(\boldsymbol{A}) = 0$).
  - An event with probability **1** is certain to occur ($P(\boldsymbol{A}) = 1$).
  - An event with probability greater than 0 and less than 1 involves **uncertainty**, but the closer its probability is to 1 the more likely it is to occur ($0 < P(\boldsymbol{A}) < 1$).
- Often expressed as % (e.g. probability of Brazil winning its next match against France is 0.6 or 60%)

Definition of Probability (cont.)

- An **event** is defined as a subset of outcomes of an experiment.
- An **experiment** is defined as any process that yields discernable, *possible* outcomes*. It can be 'experimental' or 'non-experimental'.
- The set of all possible outcomes of an experiment is called the universe of discourse or **sample space**.
- An event ***occurs*** whenever one of the defining outcomes occur. (What is the probability associated with the event defined by sample space?)
- In other words, probability is defined as relating measures of likelihood to (subsets of) outcomes of an experiment.
- E.g. flipping a coin, flipping coins, rolling dice, drawing balls from urns, etc.

a (simple)* ***random variable* *is a named quantity that associates a numerical value with each possible outcome of an experiment (cf. infra).*

How Associate Probabilities With Events?

- Two (extreme) possibilities: subjective vs. objective
- **Subjective**:
  - One person's assessment of likelihood consistent with the definition of probability (cf. supra)
  - Based on purely personal information
  - Usually relevant for unique, one-time experiments
  - How sensitive is your assessment to your probabilities?
- *What is the probability for life on Mars?*
- Most situations (experiments) are not completely unique. We usually have some history to guide us.
- We can often come by more **objective** probabilities by using the ***law of large numbers***.

Law of Large Numbers: Frequentist Interpretation of Probability

- The **relative frequency** of an event is the proportion of times the event occurs out of a number of times the experiment is run.
- This relative frequency of an event **in the long run** will get closer and closer to the true probability of the event. (Repetition under identical conditions!)
- **Important result links 'population view' to 'data (sample) view' of the world**
- For any *partitioning* of the sample space into a collectively exhaustive set of *#events* mutually exclusive events (e.g. event per outcome):
  - Repeat experiment *#repetitions* times and count number of times event ***A*** occurs. This is denoted ***#A** occurrences*.
  - *P(**A**) ≈ (#**A** occurrences) / (#repetitions)*
- E.g. flipping a 'fair' coin ('fair' → equally likely events)
- E.g. flipping an 'unfair' coin

cf. counting rules (permutations, combinations, etc.)

This extra piece of information lowers the cost of obtaining probabilities: *P*(*event of interest*) = 1/ (# different events)

Rules of Probability Calculus (A peek at the end – PART 1)

1) Complement : $P(A^c) = 1 - P(A)$

2) Addition : $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

3a) Conditional : $P(A \mid B) = \dfrac{P(A \text{ and } B)}{P(B)}$

3b) Multiplication : $P(A \text{ and } B) = P(B)P(A \mid B) = P(A)P(B \mid A)$

4) Variable Elimination : $P(A) = P(A \text{ and } B) + P(A \text{ and } B^c)$

5) Bayes' Rule : $P(A \mid B) = \dfrac{P(A)P(B \mid A)}{P(B)}$

Rules of Probability Calculus (A peek at the end – PART 2)

$$\text{1) Complement: } p(x^c) = 1 - p(x)$$

$$\text{2) Addition: } p(x \textit{ or } y) = p(x) + p(y) - p(x \textit{ and } y)$$

$$\text{3a) Conditional: } p(x \mid y) = \frac{p(x \textit{ and } y)}{p(y)}$$

$$\text{3b) Multiplication: } p(x \textit{ and } y) = p(y)p(x \mid y) = p(x)p(y \mid x)$$

$$\text{4) Variable Elimination: } p(x) = \sum_y p(x \textit{ and } y)$$

$$\text{5) Bayes' Rule: } p(x \mid y) = \frac{p(x)p(y \mid x)}{p(y)} = \frac{p(x)p(y \mid x)}{\sum_x p(x \textit{ and } y)}$$

Rules of Probability

- Apply to objective as well as subjective probabilities
- The objective is to be able to calculate complex probabilities (often the ones we are really interested in) from simpler (available, known) ones?
- The simplest probability rule involves the ***complement*** of an event.
  - If ***A*** is any event, then the complement of ***A***, denoted by A^c, is the event that ***A*** does not occur.
    - *Characteristic 1*: per definition, if ***A*** occurs, then A^c does not.
    - *Characteristic 2*: per definition, ***A*** U A^c is the sample space.
    - From both characteristics we conclude that either ***A*** occurs or it doesn't; there is no third alternative. Thus, …
  - if the probability of ***A*** is *P*(***A***), then P(A^c) = 1- *P*(***A***).
- **Odds ratio** *in favor of* ***A*** is *P*(***A***)/*P*(A^c) (e.g. for Brazil winning its next match against France this is 3/2 or 3 to 2)

Relationship Among Events

- Two special relationships among events are mutual exclusivity and collective exhaustiveness.
- We say that events are **mutually exclusive** if at most one of them can occur. That is, if one of them occurs, then none of the others can occur.
  - E.g. complements are, per definition, mutually exclusive.
- Events can also be **collectively exhaustive**, which means that they collectively comprise the entire sample space of outcomes, i.e. at least one of these events is certain to occur when running the experiment.
  - E.g. complements are, per definition, collectively exhaustive.
- These relationships are **logically independent**, i.e. the presence or absence of one does not imply the presence or absence of the other.
- In typical applications, the events we study first (= simple events) are chosen to partition the set of all possible outcomes into mutually exclusive events.

Addition Rule

- Let A_1 and A_2 be any ***2*** events. Then the addition rule of probability involves the probability that *at least one* of these events will occur.

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2) - P(A_1 \text{ and } A_2)$$

- The subtraction of the probability of the composite (or joint) event (A_1 *and* A_2) avoids 'double counting'. This **composite or joint probability** is 0 if and only if the 2 events are **mutually exclusive**. Only then is

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2)$$

- These formulas extend to analysis of more than 2 events.
  - E.g. $P(A_1 \text{ or } A_2 \text{ or } A_3) = P(A_1) + P(A_2) + P(A_3)$
    $- P(A_1 \text{ and } A_2) - P(A_1 \text{ and } A_3) - P(A_2 \text{ and } A_3) + P(A_1 \text{ and } A_2 \text{ and } A_3)$
- Note: the complement of "at least one" is "none"

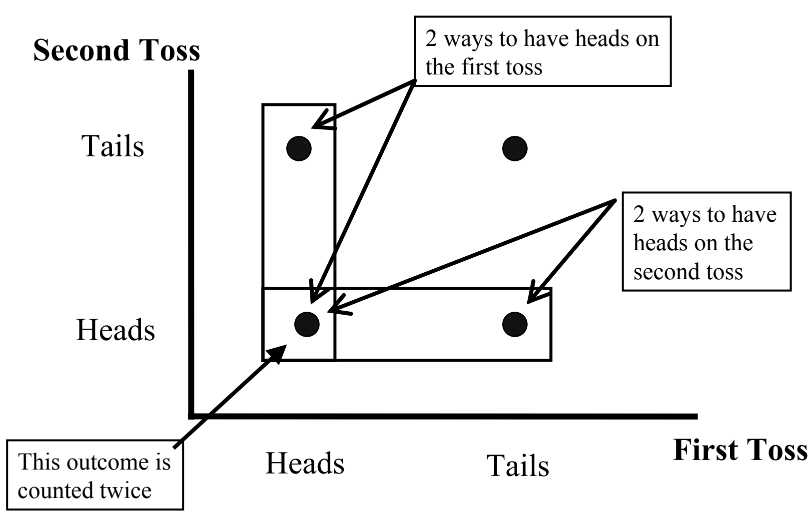
Addition Rule Example

- Suppose we flip a 'fair' coin 2 times (i.e. an experiment composed of two subexperiments). Define event ***A*** as heads on the first toss, and event ***B*** as heads on the second toss. What is *P*(***A*** *or* ***B***)? (= 'at least one')
- Using the **addition rule**:

P(***A*** *or* ***B***) = P(***A***)	+ *P*(***B***)	– *P*(***A*** *and* ***B***)
P(***A*** *or* ***B***) = 1/2	+ 1/2	– 1/4
P(***A*** *or* ***B***) = 3/4		

- What is the probability of no heads?

Addition Rule Example (cont.)



Conditional Probabilities

- Probabilities are ALWAYS assessed relative to the information currently available. As new information becomes available, probabilities often change.
- Let ***A*** and ***B*** be any events with probabilities *P*(***A***) and *P*(***B***). Typically the probability *P*(***A***) is assessed without knowledge of whether ***B*** does or does not occur. However if we are told ***B*** has occurred, the probability of ***A*** *may* change.
  - *What is the probability of a person getting married within one year?*

    vs.

  - *What is this probability of a person getting married within one year after we have observed that he is formally engaged?*

Conditional Probabilities (cont.)

- The new probability of ***A*** is called the **conditional probability** of ***A*** given ***B***. It is denoted *P(**A** | **B**).*
  - Note that there is uncertainty involving the event to the left of the vertical bar in this notation; we do not know whether it will occur or not. However, there is no uncertainty involving the event to the right of the vertical bar; we know that it has occurred or will occur.
- The **conditional probability formula** enables us to calculate *P*(***A*** | ***B***):

$$P(\mathbf{A}|\mathbf{B}) = \frac{P(\mathbf{A} \text{ and } \mathbf{B})}{P(\mathbf{B})}$$

Note: assuming P(B) is not 0.

*Note: in 'frequentist' terms this is just the relative frequency of **A** within the <u>reduced sample space</u> that includes only those outcomes that include the occurrence of **B** (e.g. coin tosses).*

Conditional Probabilities (cont.)

- In the conditional probability rule the numerator is the probability that both ***A*** and ***B*** occur. It must be known in order to determine $P(\mathbf{A} \mid \mathbf{B})$.
- However, in some applications $P(\mathbf{A} \mid \mathbf{B})$ and $P(\mathbf{B})$ are known. We are really interested in the probability of the composite event (***A** and **B***). In these cases we can multiply both side of the conditional probability formula by $P(\mathbf{B})$ to obtain the **multiplication rule (a.k.a. *de-composition or chain rule*)**.

$$P(\mathbf{A} \text{ and } \mathbf{B}) = P(\mathbf{B})\, P(\mathbf{A} \mid \mathbf{B})$$

- Or generalized to more than 2 events

$$P(\mathbf{A_1} \text{ and } \mathbf{A_2} \text{ and } \mathbf{A_3}) = P(\mathbf{A_1})\, P(\mathbf{A_2} \mid \mathbf{A_1})\, P(\mathbf{A_3} \mid \mathbf{A_2} \text{ and } \mathbf{A_1})$$

Note: decompose in function of what conditional probabilities are available!
e.g. check formula for $P(\mathbf{B} \mid \mathbf{A})$

Independent Events

- A concept that is closely tied to conditional probability is **probabilistic independence**.
- There are situations when $P(\mathbf{A})$, $P(\mathbf{A}|\mathbf{B})$ are not all that different. There are situations where these probabilities are all equal. In this case we can say that events ***A*** and ***B*** are probabilistically independent.
- **This does not mean they are mutually exclusive!** It means that knowledge of one event is of no value when assessing the probability of the other.
- The main advantage is that the multiplication rule simplifies to

$$P(\mathbf{A} \text{ and } \mathbf{B}) = P(\mathbf{A})P(\mathbf{B})$$

- **In order to determine whether events are probabilistically independent or not we usually cannot use mathematical arguments. Often, our only way out is to use empirical data (cf. law of large numbers) to decide whether independence is reasonable.**

Example 5.1

Assessing Uncertainty at the Bendrix Company

Probability Essentials

Bendrix Company's Situation

- The Bendrix Company supplies contractors with materials for the construction of houses.
- Bendrix currently has a contract with one of its customers to fill an order by the end of July.
- There is **uncertainty** about whether this deadline can be met, **due to** uncertainty about whether Bendrix will receive the materials it needs from one of its suppliers by the middle of July.
- It is currently July 1.
- How can the uncertainty in this situation be assessed? (What are we interested in?)

Assessing Bendrix

- **Objective:** we are interested in the likelihood that Bendrix will meet its end-of-July deadline, given the information the company has at the beginning of July.
- We discern the following events July 1.
  - Let ***A*** be the event that Bendrix meets its end-of-July deadline.
  - Let ***B*** be the event that Bendrix receives the materials form its supplier by the middle of July.
- Suppose that the probabilities that we are best able to assess on July 1 are
  - $P(B)$ and
  - $P(A|B)$

Assessing Bendrix (cont.)

- They estimate
  - a 2 in 3 chance of getting the materials on time; thus $P(B)=2/3$.
  - that if they receive the materials on time, then the chances of meeting the deadline are 3 out of 4; thus $P(A|B)=3/4$.
- We can now use the decomposition (multiplication) rule to obtain:

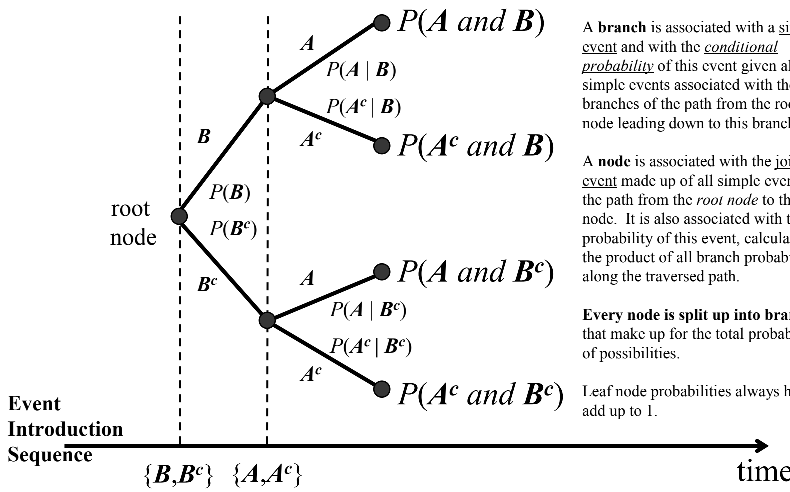
$$P(A \text{ and } B) = P(B)P(A \mid B) = (2/3)(3/4) = 0.5$$

- There is a 50-50 chance that Bendrix will get its materials on time *and* meet its deadline.

Assessing Bendrix (cont.)

- This uncertain situation is *decomposed graphically* in the form of a **probability tree**.
- Note that *in this case* the decomposition (or chaining of events) is governed by time.
  - Bendrix *initially* faces the uncertainty of whether event ***B*** or its complement ($\boldsymbol{B^c}$) will occur.
  - Regardless of whether event ***B*** takes place, Bendrix must *next* confront the uncertainty regarding event ***A***.
- Often, time specifies the natural *event introduction sequence* for the probability tree.

Assessing Bendrix (cont.)



A **branch** is associated with a <u>simple event</u> and with the <u>*conditional probability*</u> of this event given all simple events associated with the branches of the path from the root node leading down to this branch.

A **node** is associated with the <u>joint event</u> made up of all simple events on the path from the *root node* to this node. It is also associated with the probability of this event, calculated as the product of all branch probabilities along the traversed path.

Every node is split up into branches that make up for the total probability of possibilities.

Leaf node probabilities always have to add up to 1.

e.g. $P(\boldsymbol{B})\ P(\boldsymbol{A} \mid \boldsymbol{B}) = P(\boldsymbol{A} \textit{ and } \boldsymbol{B})$

Assessing Bendrix (cont.)

- Other probabilities of interest exist in this example.
- The event that the materials from the supplier do not arrive on time. We know that $P(\boldsymbol{B}^c) = 1 - P(\boldsymbol{B}) = 1/3$ from the rule of complements.
- Bendrix estimates that the chances of meeting the deadline are 1 out of 5 if the materials do not arrive on time, that is, $P(\boldsymbol{A} \mid \boldsymbol{B}^c) = 1/5$. The multiplication rule gives

$$P(\boldsymbol{A} \text{ and } \boldsymbol{B}^c) = P(\boldsymbol{A} \mid \boldsymbol{B}^c)P(\boldsymbol{B}^c) = (1/5)(1/3) = 1/15$$

Assessing Bendrix (cont.)

- The bottom line for Bendrix is whether it will meet its end-of-July deadline.
- **After the middle of July** the probability is either 3/4 (= $P(\boldsymbol{A} \mid \boldsymbol{B})$) or 1/5 (= $P(\boldsymbol{A} \mid \boldsymbol{B}^c)$) because by this time they will know whether the materials have arrived on time.
- But since it is July 1 the probability to be assessed is $P(\boldsymbol{A})$; there is still uncertainty about whether $\boldsymbol{B}$ or $\boldsymbol{B}^c$ will occur.
- We can calculate $P(\boldsymbol{A})$ from the probabilities we already know. Using the *addition rule for mutually exclusive events* we obtain

$$P(\boldsymbol{A}) = P((\boldsymbol{A} \text{ and } \boldsymbol{B}) \text{ or } (\boldsymbol{A} \text{ and } \boldsymbol{B}^c))$$

$$P(\boldsymbol{A}) = P(\boldsymbol{A} \text{ and } \boldsymbol{B}) + P(\boldsymbol{A} \text{ and } \boldsymbol{B}^c) = (1/2)+(1/15) = 17/30$$

- Note:
  - This ‘trick’ is sometimes referred to as ‘variable or event elimination’.
  - The sum of the probabilities associated with a set of nodes linked to a single starting node always equals the probability associated with the starting node.

Conditional Probability Revisited

- Whenever new information comes in, a probability of interest may change.
- This was the rationale behind the introduction of the concept of conditional probability, and corresponding multiplication rule:

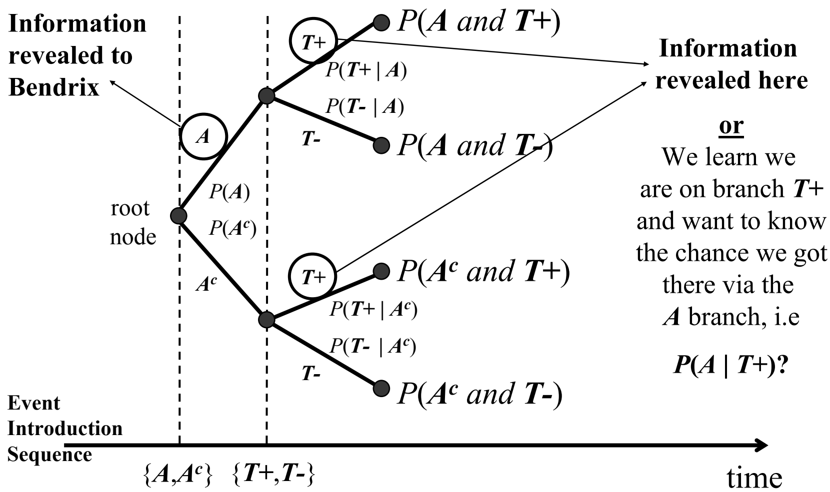
$$P(\boldsymbol{A} \mid \boldsymbol{B}) = \frac{P(\boldsymbol{A}\ and\ \boldsymbol{B})}{P(\boldsymbol{B})} \quad \& \quad P(\boldsymbol{A}\ and\ \boldsymbol{B}) = P(\boldsymbol{B})P(\boldsymbol{A} \mid \boldsymbol{B})$$

  where ***B*** is an event that occurs before ***A*** in time (or ***B*** is at the origin or cause of effect ***A***).
- Suppose we are unable to observe ***B.*** Then we are not interested in *P*(***A*** | ***B***). But we may be interested in *P*(***B*** | ***A***) !

Example: AIDS Test

- You are tested for AIDS and obtain a positive test result. What is the probability you actually have AIDS given this new information?
- Let ***A*** be the event that you have AIDS and let ***T+*** be the event that you test positive.
- You want to know *P*(***A*** | ***T+***).
- *There is a definite <u>difference</u> with the assessment of the uncertainty in the Bendrix example, as is clear from the following probability tree.*

Example: AIDS Test (cont.)



Example: AIDS Test (cont.)

- $P(\boldsymbol{A} \mid \boldsymbol{T+}) = ?$

  $P(\boldsymbol{A} \mid \boldsymbol{T+}) = P(\boldsymbol{A} \text{ and } \boldsymbol{T+}) \,/\, P(\boldsymbol{T+})$

  ? ? ?

  $P(\boldsymbol{A} \mid \boldsymbol{T+}) = [P(\boldsymbol{A})P(\boldsymbol{T+} \mid \boldsymbol{A})] \,/\, P(\boldsymbol{T+})$ (1)

  ? ! ! ?

- $P(\boldsymbol{T+}) = P(\boldsymbol{A} \text{ and } \boldsymbol{T+}) + P(\boldsymbol{A}^c \text{ and } \boldsymbol{T+})$

  ? ? ?

  $P(\boldsymbol{T+}) = P(\boldsymbol{A})P(\boldsymbol{T+} \mid \boldsymbol{A}) + P(\boldsymbol{A}^c)P(\boldsymbol{T+} \mid \boldsymbol{A}^c)$ (2)

  ! ! ! ! !

- Substitute (2) in (1)

$$P(A \mid T+) = \frac{P(A)P(T+ \mid A)}{P(A)P(T+ \mid A) + P(A^c)P(T+ \mid A^c)}$$

Example: AIDS Test (cont.)

- Data from the U.S. gives the following.
  - $P(\boldsymbol{A}) = 0.006$, thus $P(\boldsymbol{A}^c) = 0.994$
  - $P(\boldsymbol{T+} \mid \boldsymbol{A}) = 0.999$, thus $P(\boldsymbol{T-} \mid \boldsymbol{A}) = 0.001$
  - $P(\boldsymbol{T+} \mid \boldsymbol{A}^c) = 0.01$, thus $P(\boldsymbol{T-} \mid \boldsymbol{A}^c) = 0.99$
- Inserting the values gives
  - $P(\boldsymbol{A} \mid \boldsymbol{T+}) = (0.999 \times 0.006) / (0.999 \times 0.006 + 0.01 \times 0.994)$
  - $P(\boldsymbol{A} \mid \boldsymbol{T+}) = 0.38$
- The probability YOU do not have aids is 0.62!

Bayes' Rule

$$P(\boldsymbol{A} \mid \boldsymbol{T}+) = \frac{P(\boldsymbol{A})P(\boldsymbol{T}+ \mid \boldsymbol{A})}{P(\boldsymbol{A})P(\boldsymbol{T}+ \mid \boldsymbol{A}) + P(\boldsymbol{A}^c)P(\boldsymbol{T}+ \mid \boldsymbol{A}^c)}$$

or

$$P(\boldsymbol{A} \mid \boldsymbol{T}+) = P(\boldsymbol{A}) \left(\frac{P(\boldsymbol{T}+ \mid \boldsymbol{A})}{P(\boldsymbol{A})P(\boldsymbol{T}+ \mid \boldsymbol{A}) + P(\boldsymbol{A}^c)P(\boldsymbol{T}+ \mid \boldsymbol{A}^c)} \right)$$

- Interpretation:
  - $P(\boldsymbol{A})$ is your **prior probability** of having AIDS, i.e. what you assume if you believe you are just like the rest of the population
  - $P(\boldsymbol{A} \mid \boldsymbol{T+})$ is your **posterior probability**, i.e. updated probability to account for new information (evidence)
- In the end, it's 'just' a conditional probability!

Rules of Probability Calculus (The end – PART 1)

1) Complement: $P(A^c) = 1 - P(A)$

2) Addition: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

3a) Conditional: $P(A \mid B) = \dfrac{P(A \text{ and } B)}{P(B)}$

3b) Multiplication: $P(A \text{ and } B) = P(B)P(A \mid B) = P(A)P(B \mid A)$

4) Variable Elimination: $P(A) = P(A \text{ and } B) + P(A \text{ and } B^c)$

5) Bayes' Rule: $P(A \mid B) = \dfrac{P(A)P(B \mid A)}{P(B)}$

Overview of Part B – Generics

- Probability essentials (~ ‘event’ view)
  - What is a probability?
  - Where do probabilities come from?
  - How do we calculate complex probabilities from simpler (available, known) ones?
- **Random variables**
  - What is a random variable?
  - What is a probability distribution?
  - Measures of central location and variability
  - Joint probability distributions

Random Variables

- A **random variable** is a named quantity that associates a numerical value* with each possible outcome of an experiment.
  - A *discrete* random variable has only a finite number of possible values.
  - A *continuous* random variable has a continuum of possible values.
- Mathematically, there is an important difference between discrete and continuous random variables. A proper treatment of continuous variables (esp. their distributions) requires calculus.
- Though we shall use both discrete and continuous random variables later on, we shall restrict our introduction of the concepts of random variables and their probability distributions to the case of discrete random variables.

*or a tuple of values for a joint experiment (cf. infra)

Probability Distribution of a Discrete Random Variable

- The **probability distribution** of a discrete random variable X specifies all its possible values and their associated probabilities.
- All probability distributions should satisfy the following **property.**

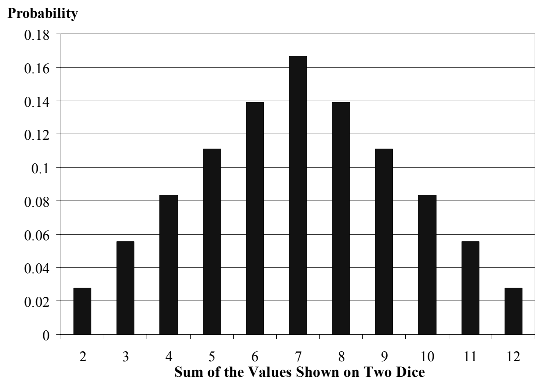
$$\sum_{i=1}^{k} P(X = v_i) = \sum_{i=1}^{k} p(v_i) = 1, \quad p(v_i) \geq 0$$

where $P(X = v_i)$ (or shorthand: $p(v_i)$) stands for the probability that X has value v_i. We assume that there are k possible values.

- *The notation p(x) is conventionally used to refer to the mapping of any value of X to its appropriate probability. It is referred to as probability distribution function (or probability density function).*

Probability Distribution of a Discrete Random Variable (cont.)

- A bar chart can be used to represent a probability distribution (function) of a discrete random variable.



The height of the bar at each *possible* value x of X is $P(X = x)$

"exactly"

Probability Distribution of a Discrete Random Variable (cont.)

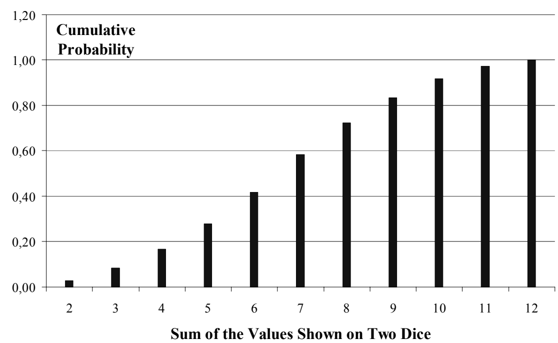
- The **cumulative probability distribution** (function) of a discrete random variable X specifies for all its values the probability that X has a value less than or equal to that value.

$$P(X \leq v_i) = \sum_{k:v_k \leq v_i} p(v_k)$$

- It specifies a monotonically increasing function from 0 to 1 over the ordered range of X-values from lowest to highest.

Probability Distribution of a Discrete Random Variable (cont.)

- A bar chart can be used to represent the cumulative probability distribution (function) of a discrete random variable.



The height of the bar at each *possible* value x of X is $P(X \leq x)$

"at most"

Probability Distribution of a Discrete Random Variable (cont.)

- A probability distribution can essentially be described or **summarized** in the same way as the data sample distributions we covered in part A of this course.
  - Measures of central location (e.g. **mean**, median, mode)
  - Measures of variability (e.g. percentiles, IQR, **variance, standard deviation**)
- There is a conceptual difference, though. But the link between both distribution concepts is clear from the law of large numbers.

Expected Value (Mean)

- The **expected value** or **mean** of a random variable X, denoted as $E(X)$, is a probability weighted sum of the possible values of X.

$$E(X) = \mu_X = \sum_{i=1}^{k} v_i p(v_i)$$

where we assume that X has k possible values v_i.

- $E(X)$ is a measure of central location for the probability distribution of X.
- Notice that we describe the mean by making use of the probability distribution of X. Compare this to the definition of the mean of a data sample distribution in Part A of this course.

Expected Value Example

- The probability distribution for the value shown on one die (X) is

v_i	$p(v_i)$	$v_i p(v_i)$
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6

$$E(X) = \sum_{i=1}^{6} v_i p(v_i) = \frac{21}{6} = 3.5$$

- Note:
  - The expected value need not be one of the possible outcomes from the experiment.
  - What is the median, the mode for the above example?

Properties of *E(X)*

- $E(X)$ is a 'linear operator':
  - $E(c) = c$ *(where 'c' is a constant value)*
  - $E(cX) = cE(X)$
  - $E(c+X) = E(c)+E(X) = c+E(X)$
- These three properties are summarized as
  - $E(a+bX) = E(a)+E(bX) = a+bE(X)$ *(where 'a' and 'b' are constants)*
- Hence, if we define a new random variable Z:

$$Z = a+bX \quad \text{then} \quad E(Z) = a+bE(X)$$

Variance of a Random Variable

- The **variance** of a random variable X, denoted as $\sigma^2(X)$, is the expected value of the squared deviations of the possible values from $E(X)$; i.*e. $\sigma^2(X)$ is the probability weighted sum of the squared deviations of the possible values from the mean*.

$$\sigma^2(X) = \sigma_X^2 = E((X - E(X))^2) = \sum_{i=1}^{k}(v_i - E(X))^2 p(v_i)$$

- The **standard deviation** of a random variable X, denoted as $\sigma(X)$, is a more natural measure of variability for the probability distribution of X.

$$\sigma(X) = \sigma_X = \sqrt{\sigma_X^2}$$

Properties of $\sigma^2(X)$

- Let $Z = a+bX$ then

$$\sigma^2(Z) = \sigma_Z^2 = E((Z - E(Z))^2)$$
$$\sigma^2(Z) = E((a + bX - E(a + bX))^2)$$
$$\sigma^2(Z) = E((a + bX - E(a) - bE(X))^2)$$
$$\sigma^2(Z) = E((bX - bE(X))^2)$$
$$\sigma^2(Z) = E(b^2(X - E(X))^2)$$
$$\sigma^2(Z) = b^2E((X - E(X))^2)$$
$$\boxed{\sigma^2(Z) = b^2\sigma^2(X)}$$

Thus, the variance of Z is proportional to the variance of X.

Example 5.2

Market Return Scenarios for the National Economy

Distribution of a Single Random Variable

Market Returns Example

- The Excel file specifies market return outcomes, variable values and corresponding probabilities estimated by an investor.

	A	B	C	D
1	**Mean, variance, and standard deviation of a random variable**			
2				
3	Economic outcome	Probability	Market	Sq dev from mean
4	Rapid Expansion	0.12	0.23	0.005929
5	Moderate Expansion	0.40	0.18	0.000729
6	No Growth	0.25	0.15	0.000009
7	Moderate Contraction	0.15	0.09	0.003969
8	Serious Contraction	0.08	0.03	0.015129
9				
10	Mean return	0.153	**Range names**	
11	Variance of return	0.002811	Mean: B10	
12	Stdev of return	0.053	Probs: B4:B8	
13			Returns: C4:C8	
14			SqDevs: D4:D8	
15			Var: B11	
16				

Mean return: **=SUMPRODUCT(Returns,Probs)**
Squared Deviations: **=(C4-Mean)^2**
Variance: **=SUMPRODUCT(SqDevs,Probs)**
Standard Deviation: **=SQRT(Var)**

Market
turn Scenar

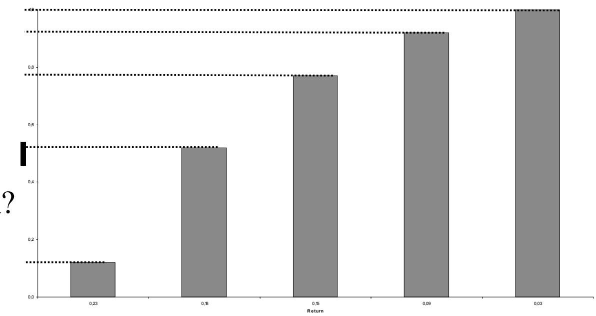
Market Returns Example (cont.)

- We see that the mean return is 15.3% and the standard deviation is 5.3%. *What do these mean?*
- First, the mean or expected return does **not imply** that the most likely return is 15.3%, **nor** is this the value that the investor "expects" to occur. The value 15.3% is not even a possible market return.
- *We can understand these measures better in terms of long-run averages.*
  - If we can see the coming year repeated many times, using the same probability distribution, then the average of these times would be close to 15.3% and their standard deviation would be 5.3%.
  - You can simulate this with Excel using *Rand()*. Simulation can also be used for sensitivity analysis (cf. subjective probabilities).

Market
turn Scenar

Market Returns Example (cont.)

How to generate
a sample from
a given distribution?



Step 1: Draw (correct) cumulative probability chart

Step 2: Cut Y-axes up into line segments proportional to the probabilities of probability distribution

Step 3: Generate a random number between 0 and 1

Step 4: Map random number into the matching return value on the X-axes

See book p. 204

Market
turn Scenari

Rules of Probability Calculus Revisited

$$1)\ \text{Complement}: p(x^c) = 1 - p(x)$$

$$2)\ \text{Addition}: p(x\ or\ y) = p(x) + p(y) - p(x\ and\ y)$$

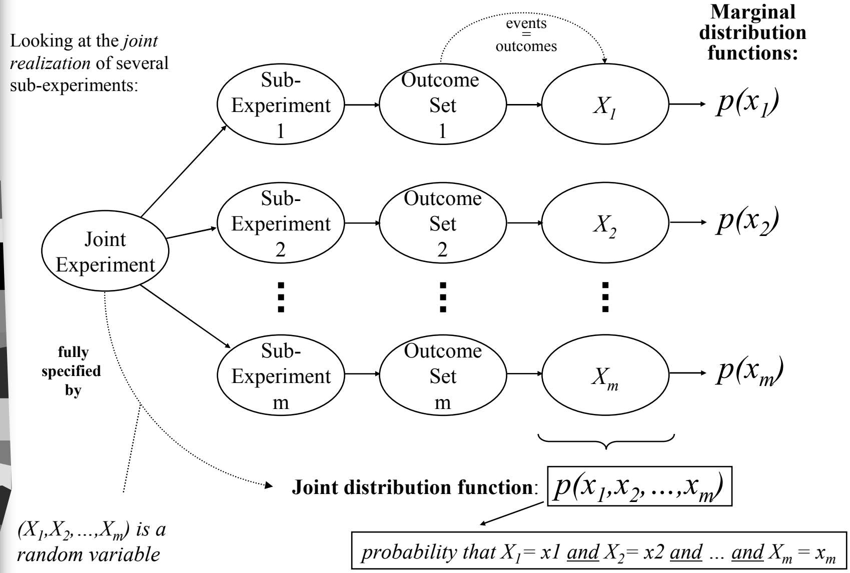
$$3a)\ \text{Conditional}: p(x \mid y) = \frac{p(x\ and\ y)}{p(y)}$$

$$3b)\ \text{Multiplication}: p(x\ and\ y) = p(y)p(x \mid y) = p(x)p(y \mid x)$$

$$4)\ \text{Variable Elimination}: p(x) = \sum_y p(x\ and\ y)$$

$$5)\ \text{Bayes' Rule}: p(x \mid y) = \frac{p(x)p(y \mid x)}{p(y)} = \frac{p(x)p(y \mid x)}{\sum_x p(x\ and\ y)}$$

Joint Probability Distributions



Joint Probability Distributions (cont.)

- **Probability distribution table**:
- Very similar to the way we described *relationships* between variables in Part A of the course via **pivot tables**, i.e. each cell could be expressed by a *count of observations in the cell as a % of the total number of observations.*
- There is a **conceptual difference**.
- Still, there is a clear relationship via the **law of large numbers**. This is the trick to interpretation of the probability distribution table.

Look at this example:

	A	B	C	D	E	F
1	**Probability distribution of demands for substitute products**					
2						
3			Demand for product 1			
4			100	200	300	400
5		50	0.015	0.040	0.050	0.035
6	Demand	100	0.030	0.080	0.075	0.025
7	for	150	0.050	0.100	0.100	0.020
8	product 2	200	0.045	0.100	0.050	0.010
9		250	0.060	0.080	0.025	0.010

How does the demand of both products move together?

→Specified by p(d1,d2)

→Represented in the form of a table

substitutes.xls

Joint Probability Distributions (cont.)

- A joint probability distribution contains an enormous amount of information. It allows us to compute all marginal and conditional probability distributions.
- **Marginal distributions** are easily obtained via the rule of *variable elimination* (cf. supra).

Probability distribution of demands for substitute products

		Demand for product 1				
		100	200	300	400	
	50	0.015	0.040	0.050	0.035	0.14
Demand	100	0.030	0.080	0.075	0.025	0.21
for	150	0.050	0.100	0.100	0.020	0.27
product 2	200	0.045	0.100	0.050	0.010	0.21
	250	0.060	0.080	0.025	0.010	0.18
		0.20	0.40	0.30	0.10	

e.g.
***P*(D_2=50)** =
P(D_2=50 and D_1=100) +
P(D_2=50 and D_1=200) +
P(D_2=50 and D_1=300) +
P(D_2=50 and D_1=400) =
0.015+0.040+0.050+0.035=
0.14

- This comes down to row and column sums in the joint probability table. But these tell us nothing about the relationship between the variables.

substitutes.xls

Joint Probability Distributions (cont.)

- **Conditional probabilities** do tell us about the relationship between the variables.
- The best way to study relationships among variables is often to look at conditional probability distributions.

e.g. take another look at the joint behavior of D_2 and D_1

Conditional distribution of demand for product 1, given demand for product 2

		Demand for product 1				
		100	200	300	400	
	50	0,11	0,29	0,36	0,25	1
Demand	100	0,14	0,38	0,36	0,12	1
for	150	0,19	0,37	0,37	0,07	1
product 2	200	0,22	0,49	0,24	0,05	1
	250	0,34	0,46	0,14	0,06	1

$p(d_1|d_2)=p(d_1,d_2)/p(d_2)$

$p(d_2|d_1)=p(d_1,d_2)/p(d_1)$

Conditional distribution of demand for product 2, given demand for product 1

		Demand for product 1			
		100	200	300	400
	50	0,08	0,10	0,17	0,35
Demand	100	0,15	0,20	0,25	0,25
for	150	0,25	0,25	0,33	0,20
product 2	200	0,23	0,25	0,17	0,10
	250	0,30	0,20	0,08	0,10
		1	1	1	1

or plotted:

substitutes.xls

Joint Probability Distributions (cont.)

- Sometimes we are not able to **specify** the joint probability distribution directly.
- An alternative way to come by the joint probability distribution is to **calculate** it after specifying all prior and conditional probability distributions.

  e.g. the joint behavior of D_2 and D_1 can be calculated via

  $$p(d_1,d_2) = p(d_1|d_2)p(d_2) = p(d_2|d_1)p(d_1)$$

  In case we know (or assume) that D_2 and D_1 are **independent random variables**, i.e. _all_ conditional probabilities are equal to marginal ones, then this formula simplifies to

  $$p(d_1,d_2) = p(d_1)p(d_2)$$

substitutes.xls

Covariance and Correlation

- The covariance and correlation of **two** random variables X and Y, denoted $Cov(X,Y)$ or σ_{XY}, respectively $Corr(X,Y)$ or ρ_{XY}, are defined as

  $$Cov(X,Y) = \sigma(X,Y) = \sigma_{XY} = \sum_{i=1}^{k}(x_i - E(X))(y_i - E(Y))p(x_i, y_i)$$

  $$Corr(X,Y) = \rho(X,Y) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

  where k is the number of cells in the joint probability table of X and Y.
- The interpretation of the covariance and correlation based on a joint probability distribution is essentially the same as for known data (cf. Part A of this course), though the concepts are different.
- **You can check that the covariance (correlation) of two independent random variables is always 0.**

Properties of Weighted Sums of Random Variables

- Let $Z = a_1X_1+a_2X_2+\ldots+a_mX_m$ then

$$E(Z) = a_1E(X_1) + a_2E(X_2) + \ldots + a_mE(X_m)$$

$$E(Z) = \sum_{i=1}^{m} a_i E(X_i)$$

$$\sigma^2(Z) = \sum_{i=1}^{m} a_i^2 \sigma^2(X_i) + \sum_{i<j} 2a_i a_j \sigma(X_i, Y_j)$$

- In case all variables X_i are independent, $\sigma^2(Z)$ simplifies to

$$\sigma^2(Z) = \sum_{i=1}^{m} a_i^2 \sigma^2(X_i)$$

Overview of Part B – Generics Summary

- Probability essentials (~ 'event' view)
  - What is a probability?
  - Where do probabilities come from?
  - How do we calculate complex probabilities from simpler (available, known) ones?
- Random variables (~ 'variable' view)
  - What is a random variable?
  - What is a probability distribution? (+ recasting of calculus)
  - Measures of central location and variability
  - Joint probability distributions (~ relationships between variables)