Chapter Six

Probability

Approaches to Assigning Probabilities…

There are three ways to assign a probability, $P(O_i)$, to an outcome, O_i, namely:

Classical approach: based on equally likely events.

Relative frequency: assigning probabilities based on experimentation or historical data.

Subjective approach: Assigning probabilities based on the assignor's (subjective) judgment.

Classical Approach…

If an experiment has n possible outcomes, this method would assign a probability of 1/n to each outcome. It is necessary to determine the number of possible outcomes.

Experiment: Rolling a ***die***

Outcomes {1, 2, 3, 4, 5, 6}

Probabilities: Each sample point has a 1/6 chance of occurring.

Classical Approach…

Experiment: Rolling two dice and observing the total

Outcomes: {2, 3, …, 12}

Examples:

P(2) = 1/36

P(6) = 5/36

P(10) = 3/36

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Relative Frequency Approach…

Bits & Bytes Computer Shop tracks the number of desktop computer systems it sells over a month (30 days):

Desktops Sold	# of Days
0	1
1	2
2	10
3	12
4	5

For example,
10 days out of 30
2 desktops were sold.

From this we can construct the probabilities of an event (i.e. the # of desktop sold on a given day)…

Relative Frequency Approach…

Desktops Sold	# of Days	Desktops Sold
0	1	1/30 = .03
1	2	2/30 = .07
2	10	10/30 = .33
3	12	12/30 = .40
4	5	5/30 = .17
		Σ = 1.00

"There is a 40% chance Bits & Bytes will sell 3 desktops on any given day"

Subjective Approach…

"In the subjective approach we define probability as the degree of belief that we hold in the occurrence of an event"

E.g. weather forecasting's "P.O.P."

"Probability of Precipitation" (or P.O.P.) is defined in different ways by different forecasters, but basically it's a subjective probability based on past observations combined with current weather conditions.

POP 60% – based on current conditions, there is a 60% chance of rain (say).

Interpreting Probability…

No matter which method is used to assign probabilities all will be interpreted in the relative frequency approach

For example, a government lottery game where 6 numbers (of 49) are picked. The classical approach would predict the probability for any one number being picked as 1/49=2.04%.

We interpret this to mean that in the long run each number will be picked 2.04% of the time.

Joint, Marginal, Conditional Probability…

We study methods to determine probabilities of events that result from ***combining*** other events in various ways.

There are several types of combinations and relationships between events:

- Complement event
- Intersection of events
- Union of events
- Mutually exclusive events
- Dependent and independent events

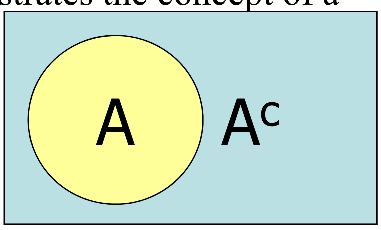
Complement of an Event…

The ***complement of event*** A is defined to be the event consisting of all sample points that are “not in A”.

Complement of A is denoted by A^c

The Venn diagram below illustrates the concept of a complement.

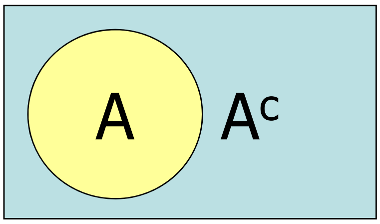
$P(A) + P(A^c) = 1$



Complement of an Event…

For example, the rectangle stores all the possible tosses of 2 dice {(1,1), 1,2),… (6,6)} Let A = tosses totaling 7 {(1,6), (2, 5), (3,4), (4,3), (5,2), (6,1)}

P(Total = 7) + P(Total not equal to 7) = 1

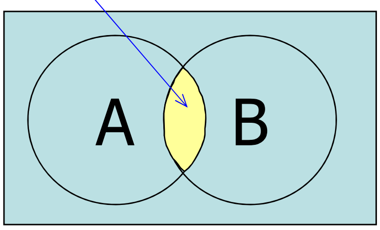


Intersection of Two Events…

The ***intersection of events*** A and B is the set of all sample points that are in both A <u>and</u> B.

The intersection is denoted: **A and B**

The ***joint probability*** of A and B is the probability of the intersection of A and B, i.e. P(A and B)



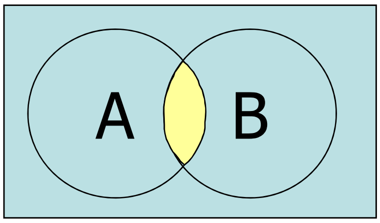
Intersection of Two Events…

For example, let A = tosses where first toss is 1 {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)}

and B = tosses where the second toss is 5 {(1,5), (2,5), (3,5), (4,5), (5,5), (6,5)}

The intersection is {(1,5)}

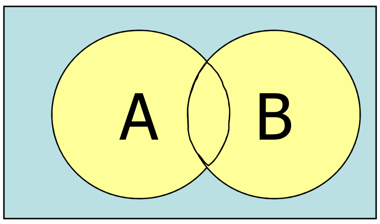
The ***joint probability*** of A and B is the probability of the intersection of A and B, i.e. P(A and B) = 1/36



Union of Two Events…

The ***union of two events*** A and B, is the event containing all sample points that are in A or B or both:

Union of A and B is denoted: **A or B**



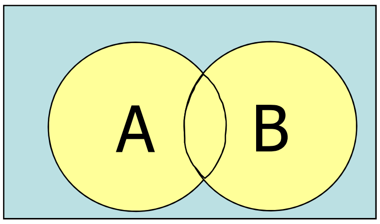
Union of Two Events…

For example, let A = tosses where first toss is 1 {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)}

and B is the tosses that the second toss is 5 {(1,5), (2,5), (3,5), (4,5), (5,5), (6,5)}

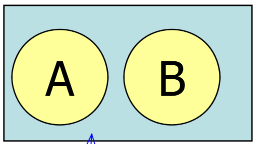
Union of A and B is {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)}

(2,5), (3,5), (4,5), (5,5), (6,5)}



Mutually Exclusive Events…

When two events are ***mutually exclusive*** (that is the two events cannot occur together), their joint probability is 0, hence:

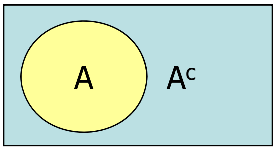


Mutually exclusive; no points in common…

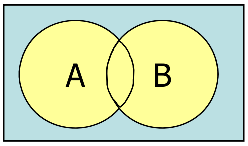
For example A = tosses totaling 7 and B = tosses totaling 11

Basic Relationships of Probability…

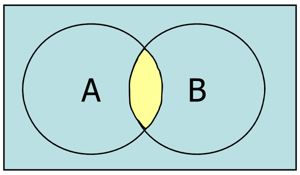
Complement of Event



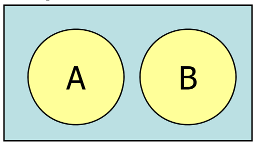
Union of Events



Intersection of Events



Mutually Exclusive Events



Example 6.1…

Why are some mutual fund managers more successful than others? One possible factor is where the manager earned his or her MBA. The following table compares mutual fund performance against the ranking of the school where the fund manager earned their MBA:

	Mutual fund outperforms the market	Mutual fund doesn't outperform the market
Top 20 MBA program	**.11**	**.29**
Not top 20 MBA program	**.06**	**.54**

E.g. This is the probability that a mutual fund outperforms **AND** the manager was in a top-20 MBA program; it's a **joint probability**.

Example 6.1…

Alternatively, we could introduce shorthand notation to represent the events:

A_1 = Fund manager graduated from a top-20 MBA program
A_2 = Fund manager did not graduate from a top-20 MBA program
B_1 = Fund outperforms the market
B_2 = Fund does not outperform the market

	B_1	B_2
A_1	**.11**	**.29**
A_2	**.06**	**.54**

E.g. P(A_2 and B_1) = .06
= the probability a fund outperforms the market
and the manager isn't from a top-20 school.

Marginal Probabilities…

Marginal probabilities are computed by adding across rows and down columns; that is they are calculated in the ***margins*** of the table:

$P(A_2) = .06 + .54$
"what's the probability a fund manager isn't from a top school?"

	B_1	B_2	$P(A_i)$
A_1	**.11**	**.29**	**.40**
A_2	**.06**	**.54**	**.60**
$P(B_j)$	**.17**	**.83**	**1.00**

$P(B_1) = .11 + .06$
"what's the probability a fund outperforms the market?"

BOTH margins must add to 1
(useful error check)

Conditional Probability…

Conditional probability is used to determine how two events are related; that is, we can determine the probability of one event ***given*** the occurrence of another related event.

Conditional probabilities are written as **P(A | B)** and read as "the probability of A *given* B" and is calculated as:

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

Conditional Probability…

Again, the probability of an event ***given*** that another event has occurred is called a conditional probability…

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

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

Note how "A given B" and "B given A" are related…

Conditional Probability…

Example 6.2 What's the probability that a fund will outperform the market ***given*** that the manager graduated from a top-20 MBA program?

Recall:

A_1 = Fund manager graduated from a top-20 MBA program

A_2 = Fund manager did not graduate from a top-20 MBA program

B_1 = Fund outperforms the market

B_2 = Fund does not outperform the market

Thus, we want to know "what is $\mathbf{P(B_1 \mid A_1)}$?"

Conditional Probability…

We want to calculate $\mathbf{P(B_1 \mid A_1)}$

	B_1	B_2	$P(A_i)$
A_1	**.11**	**.29**	**.40**
A_2	**.06**	**.54**	**.60**
$P(B_j)$	**.17**	**.83**	**1.00**

$$P(B_1 \mid A_1) = \frac{P(A_1 \text{ and } B_1)}{P(A_1)} = \frac{.11}{.40} = .275$$

Thus, there is a 27.5% chance that that a fund will outperform the market given that the manager graduated from a top-20 MBA program.

Independence…

One of the objectives of calculating conditional probability is to determine whether two events are related.

In particular, we would like to know whether they are ***independent***, that is, if the probability of one event is ***not affected*** by the occurrence of the other event.

Two events A and B are said to be ***independent*** if

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

or

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

Independence…

For example, we saw that

$P(B_1 \mid A_1) = .275$

The marginal probability for B_1 is: $P(B_1) = 0.17$

Since $P(B_1|A_1) \neq P(B_1)$, B_1 and A_1 are ***not independent*** events.

Stated another way, they are ***dependent***. That is, the probability of one event (B_1) ***is affected*** by the occurrence of the other event (A_1).

Union…

We stated earlier that the union of two events is denoted as: **A or B**. We can use this concept to answer questions like:

Determine the probability that a fund outperforms the market ***or*** the manager graduated from a top-20 MBA program.

Union…

Determine the probability that a fund outperforms (B_1) ***or*** the manager graduated from a top-20 MBA program (A_1).

A_1 or B_1 occurs whenever:

A_1 and B_1 occurs, A_1 and B_2 occurs, or A_2 and B_1 occurs…

	B_1	B_2	$P(A_i)$
A_1	**.11**	**.29**	**.40**
A_2	**.06**	**.54**	**.60**
$P(B_j)$	**.17**	**.83**	**1.00**

$P(A_1 \text{ or } B_1) = .11 + .06 + .29 = \mathbf{.46}$

Union…

Determine the probability that a fund outperforms (B_1) ***or*** the manager graduated from a top-20 MBA program (A_1).

	B_1	B_2	$P(A_i)$
A_1	.11	.29	.40
A_2	.06	.54	.60
$P(B_j)$	.17	.83	1.00

$P(A_1 \text{ or } B_1) = .11 + .06 + .29 = \mathbf{.46}$

Alternatively…

Take 100% and subtract off "when doesn't A_1 or B_1 occur"?

i.e. at A_2 and B_2

	B_1	B_2	$P(A_i)$
A_1	.11	.29	.40
A_2	.06	.54	.60
$P(B_j)$	.17	.83	1.00

$P(A_1 \text{ or } B_1) = 1 - P(A_2 \text{ and } B_2) = 1 - .54 = \mathbf{.46}$

Probability Rules and Trees…

We introduce three rules that enable us to calculate the probability of more complex events from the probability of simpler events…

The Complement Rule,

The Multiplication Rule, and

The Addition Rule

Complement Rule…

As we saw earlier with the complement event, the ***complement rule*** gives us the probability of an event NOT occurring. That is:

$P(A^C) = 1 – P(A)$

For example, in the simple roll of a die, the probability of the number "1" being rolled is 1/6. The probability that some number other than "1" will be rolled is 1 – 1/6 = 5/6.

Multiplication Rule…

The ***multiplication rule*** is used to calculate the ***joint probability*** of two events. It is based on the formula for conditional probability defined earlier:

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

If we multiply both sides of the equation by P(B) we have:

P(A and B) = P(A | B)•P(B)

Likewise, P(A and B) = P(B | A) • P(A)

If A and B are independent events, then P(A and B) = P(A)•P(B)

Example 6.5…

A graduate statistics course has seven male and three female students. The professor wants to select two students at random to help her conduct a research project. What is the probability that the two students chosen are female?

Let A represent the event that the first student is female

P(A) = 3/10 = .30

What about the second student?

Example 6.5…

A graduate statistics course has seven male and three female students. The professor wants to select two students at random to help her conduct a research project. What is the probability that the two students chosen are female?

Let B represent the event that the second student is female

P(B | A) = 2/9 = .22

That is, the probability of choosing a female student ***given*** that the first student chosen is 2 (females) / 9 (remaining students) = 2/9

Example 6.5…

A graduate statistics course has seven male and three female students. The professor wants to select two students at random to help her conduct a research project. **What is the probability that the two students chosen are female?**

Thus, we want to answer the question: what is **P(A and B)** ?

P(A and B) = P(A)•P(B|A) = (3/10)(2/9) = 6/90 = .067

"There is a 6.7% chance that the professor will choose two female students from her grad class of 10."

Example 6.6

Refer to Example 6.5. The professor who teaches the course is suffering from the flu and will be unavailable for two classes. The professor's replacement will teach the next two classes. His style is to select one student at random and pick on him or her to answer questions during that class. What is the probability that the two students chosen are female?

Let A represent the event that the first student is female

P(A) = 3/10 = .30

What about the second student?

Example 6.6

Let B represent the event that the second student is female

P(B | A) = 3/10 = .30

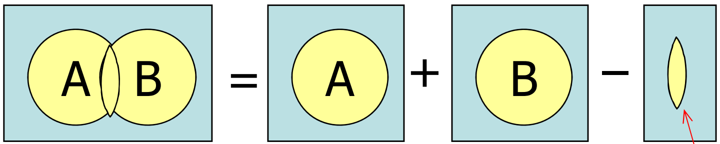
That is, the probability of choosing a female student ***given*** that the first student chosen is unchanged since the student selected in the first class can be chosen in the second class.

P(A and B) = P(A)•P(B|A) = (3/10)(3/10) = 9/100 = .090

Addition Rule…

Recall: the ***addition rule*** is used to compute the probability of event A ***or*** B ***or*** both A and B occurring; i.e. the union of A and B.

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



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

If A and B are mutually exclusive, then this term goes to zero

Example 6.7…

In a large city, two newspapers are published, the Sun and the Post. The circulation departments report that 22% of the city's households have a subscription to the Sun and 35% subscribe to the Post. A survey reveals that 6% of all households subscribe to both newspapers. What proportion of the city's households subscribe to either newspaper?

That is, what is the probability of selecting a household at random that subscribes to the Sun or the Post or both?

i.e. what is **P(Sun or Post)** ?

Example 6.7…

In a large city, two newspapers are published, the Sun and the Post. The circulation departments report that 22% of the city's households have a subscription to the Sun and 35% subscribe to the Post. A survey reveals that 6% of all households subscribe to both newspapers. What proportion of the city's households subscribe to either newspaper?

P(Sun or Post) = P(Sun) + P(Post) – P(Sun and Post)

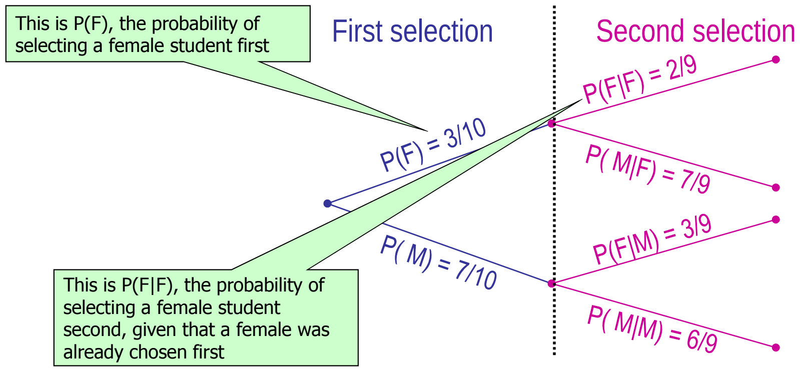
= .22 + .35 – .06 = .51

"There is a 51% probability that a randomly selected household subscribes to one or the other or both papers"

Probability Trees

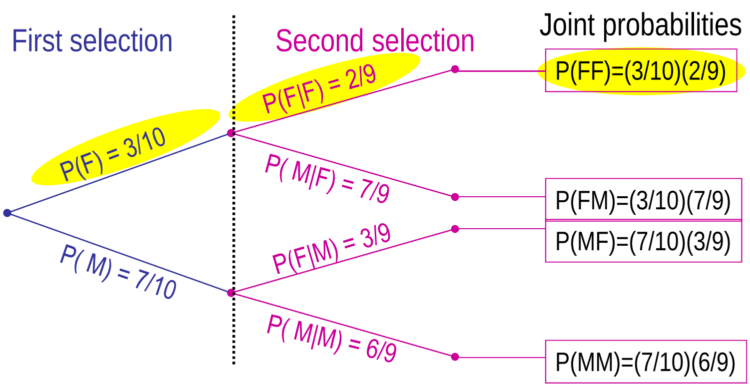
An effective and simpler method of applying the probability rules is the probability tree, wherein the events in an experiment are represented by lines. The resulting figure resembles a tree, hence the name. We will illustrate the probability tree with several examples, including two that we addressed using the probability rules alone.

Example 6.5



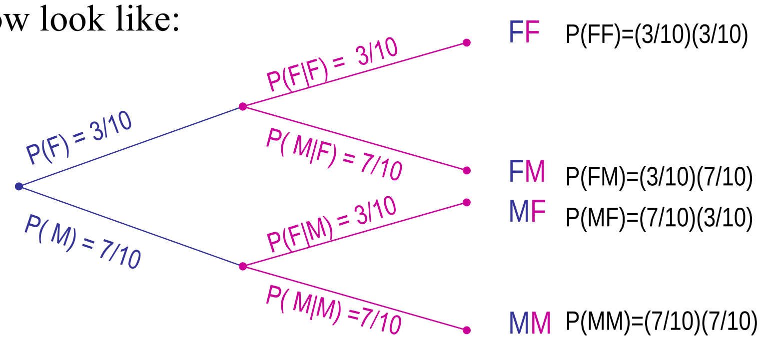
Probability Trees…

At the ends of the "branches", we calculate ***joint probabilities*** as the **product** of the individual probabilities on the preceding branches.



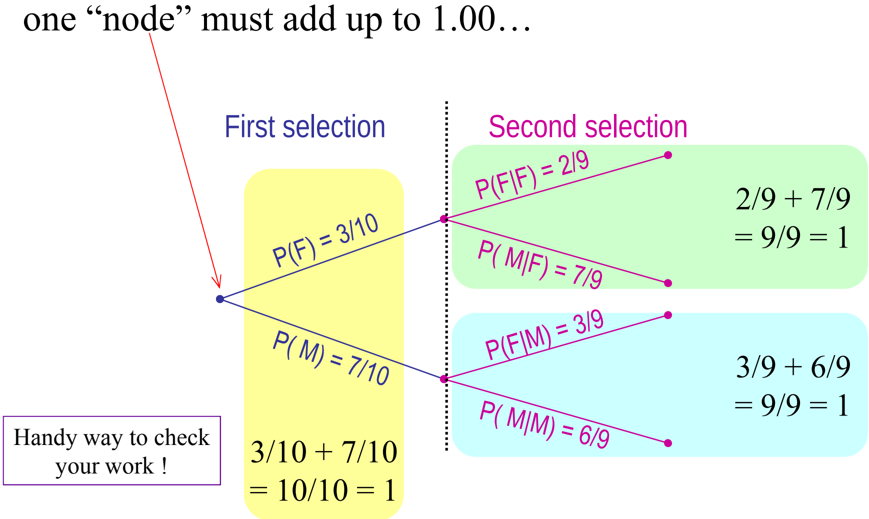
Example 6.6

Suppose we have our grad class of 10 students again, but make the student sampling ***independent***, that is "with replacement" – a student could be picked first and picked again in the second round. Our tree and joint probabilities now look like:



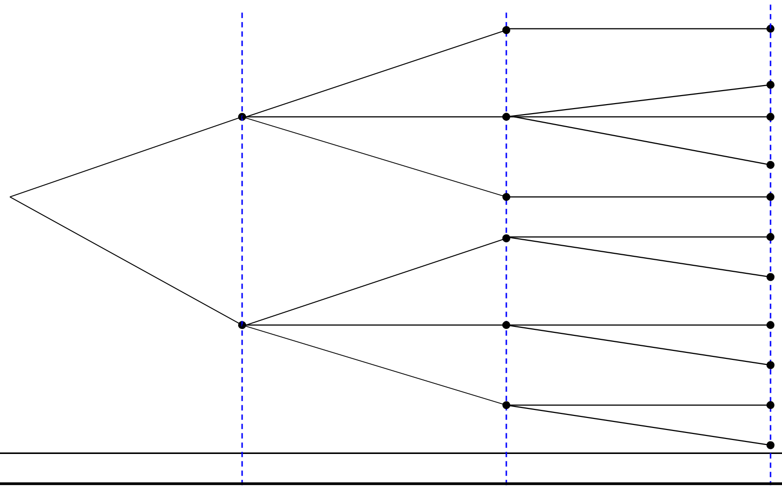
Probability Trees…

The probabilities associated with any set of branches from one "node" must add up to 1.00…



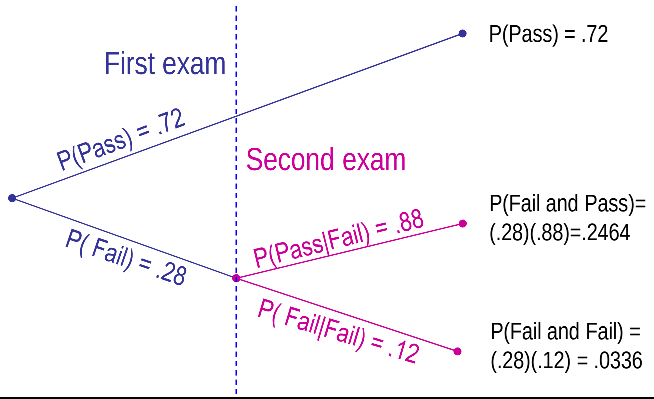
Probability Trees…

Note: there is no requirement that the branches splits be binary, nor that the tree only goes two levels deep, or that there be the same number of splits at each sub node…



Example 6.8

Law school grads must pass a bar exam. Suppose pass rate for first-time test takers is 72%. They can re-write if they fail and 88% pass their second attempt. What is the probability that a randomly grad passes the bar?



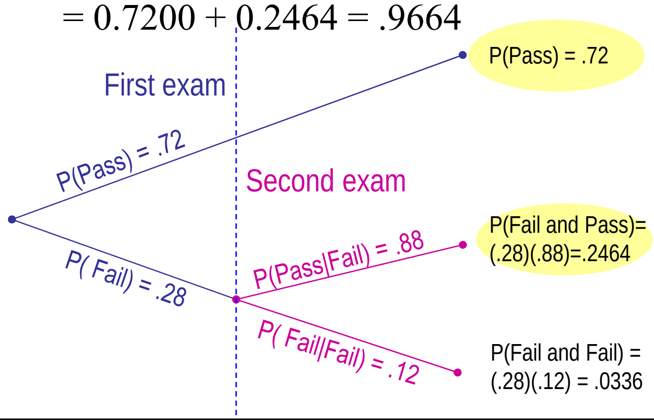
Example 6.8

What is the probability that a randomly grad passes the bar?

"There is almost a 97% chance they will pass the bar"

P(Pass) = P(Pass 1^{st}) + P(Fail 1^{st} and Pass 2^{nd}) =

= 0.7200 + 0.2464 = .9664



Bayes' Law…

Bayes' Law is named for Thomas Bayes, an eighteenth century mathematician.

In its most basic form, if we know P(B | A),

we can apply Bayes' Law to determine P(A | B)

$$P(B\,|\,A) \Rightarrow P(A\,|\,B)$$

for example …

Example 6.9 – Pay $500 for MBA prep??

The Graduate Management Admission Test (GMAT) is a requirement for all applicants of MBA programs. There are a variety of preparatory courses designed to help improve GMAT scores, which range from 200 to 800. Suppose that a survey of MBA students reveals that among GMAT scorers above 650, 52% took a preparatory course, whereas among GMAT scorers of less than 650 only 23% took a preparatory course. An applicant to an MBA program has determined that he needs a score of more than 650 to get into a certain MBA program, but he feels that his probability of getting that high a score is quite low--10%. He is considering taking a preparatory course that cost $500. He is willing to do so only if his probability of achieving 650 or more doubles. What should he do?

Example 6.9 – Convert to Statistical Notation

Let A = GMAT score of 650 or more,

hence A^C = GMAT score less than 650

Our student has determined the probability of getting greater than 650 (without any prep course) as 10%, that is:

$P(A) = .10$

It follows that $P(A^C) = 1 – .10 = .90$

Example 6.9 – Convert to Statistical Notation

Let B represent the event "take the prep course"

and thus, B^C is "do not take the prep course"

From our survey information, we're told that ***among GMAT scorers above 650, 52% took a preparatory course***, that is:

$P(B \mid A) = .52$

(Probability of finding a student who took the prep course ***given that*** they scored above 650…)

But our student wants to know **P(A | B)**, that is, ***what is the probability of getting more than 650 given that a prep course is taken?***

If this probability is > 20%, he will spend $500 on the prep course.

Example 6.9 – Convert to Statistical Notation

Among GMAT scorers of less than 650 only 23% took a preparatory course. That is:

$P(B \mid A^C) = .23$

(Probability of finding a student who took the prep course ***given that*** he or she scored less than 650…)

Example 6.9 – Convert to Statistical Notation

Conditional probabilities are

$P(B \mid A) = .52$

and

$P(B \mid A^C) = .23$

Again using the complement rule we find the following conditional probabilities.

$P(B^C \mid A) = 1 - .52 = .48$

and

$P(B^C \mid A^C) = 1 - .23 = .77$

Example 6.9 – Continued…

We are trying to determine P(A | B), perhaps the definition of conditional probability from earlier will assist us…

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

We don't know P(A and B) and we don't know P(B). Hmm.

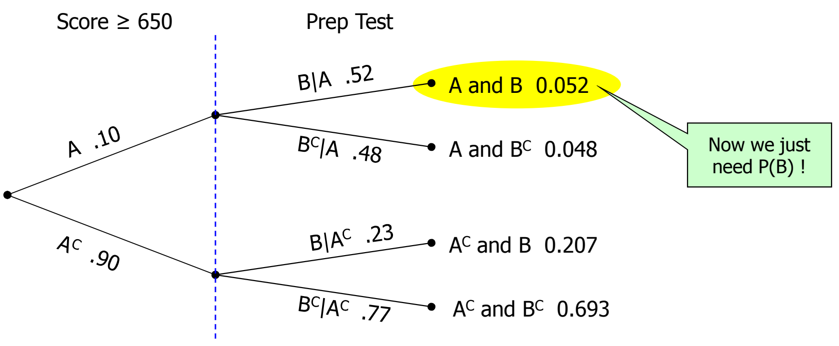
Perhaps if we construct a probability tree…

Example 6.9 – Continued…

In order to go from

$P(B \mid A) = 0.52$ to $P(A \mid B) = ??$

we need to apply Bayes' Law. **Graphically:**

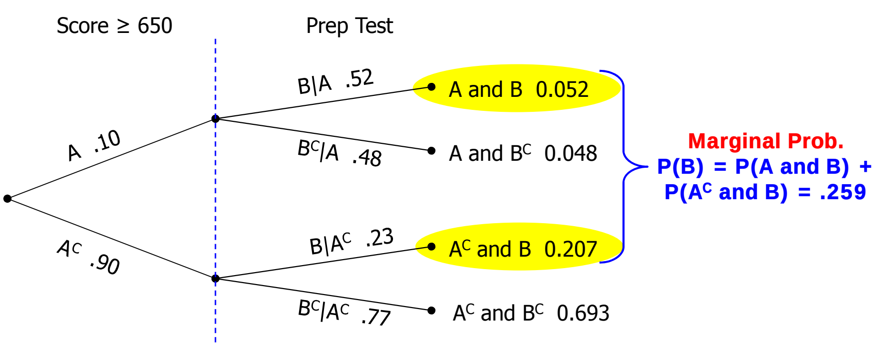


Example 6.9 – Continued…

In order to go from

$P(B \mid A) = 0.52$ to $P(A \mid B) = ??$

we need to apply Bayes' Law. **Graphically:**



Example 6.9 – FYI

Thus,

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{.052}{.259} = .201$$

The probability of scoring 650 or better doubles to 20.1% when the prep course is taken.

Bayesian Terminology…

The probabilities P(A) and P(A^C) are called ***prior probabilities*** because they are determined ***prior*** to the decision about taking the preparatory course.

The conditional probability P(A | B) is called a ***posterior probability*** (or revised probability), because the prior probability is revised ***after*** the decision about taking the preparatory course.