Essential Mathematics for the Political and Social Research

JEFF GILL
Cambridge University Press

Lecture Slides, Chapter 7: Probability Theory

Chapter 7 Objectives

- Probability provides a way of systematically and rigorously treating uncertainty.
- Humans often think in probabilistic terms (even when not gambling).
- Probability theory is a precursor to understanding statistics and various fields of applied mathematics.
- Probability theory could be described as "mathematical models of uncertain reality" because it supports the use of uncertainty in many fields.

Two Interpretations of Probability

- *Subjective probability* is *individually* defined by the conditions under which a person would make a bet or assume a risk in pursuit of some reward.
- *Objective probability* is defined as a limiting relative frequency: the long-run behavior of a non-deterministic outcome or just an observed proportion in a population.

Counting Rules and Permutations

- Number of permutations of n objects: $n!$.
- Number of subsets of a set of n elements: $2n$.

Counting Rules and Permutations

- The number of ways in which n individual units can be ordered is governed by the use of the factorial function:
$$n(n-1)(n-2)\cdots(2)(1) = n!.$$
- There are n ways to select the first object in an ordered list, $n-1$ ways to pick the second, and so on, until we have one item left and there is only one way to pick that one item.
- For example, consider the set $\{A, B, C\}$.
  - There are three (n) ways to pick the first item: A, B, or C.
  - Once we have done this, say we picked C to go first, then there are two ways ($n-1$) to pick the second item: either A or B.
  - After that pick, assume A, then there is only one way to pick the last item ($n-2$): B.

The Fundamental Theorem of Counting

- If there are k distinct decision stages to an operation or process,
- each with its own n_k number of alternatives,
- then there are $\prod_{i=1}^{k} n_k$ possible outcomes.

What this formal language says is that if we have a specific number of individual steps, each of which has some set of alternatives, then the total number of alternatives is the product of those at each step. So for $1, 2, \ldots, k$ different characteristics we multiply the corresponding $n_1, n_2, \ldots, n_k$ number of features.

Counting Example WIth Cards

- Suppose we consider cards in a deck in terms of suit ($n_1 = 4$) and whether they are face cards ($n_2 = 2$).
- There are 8 possible countable outcomes defined by crossing [Diamonds, Hearts, Spades, Clubs] with [Face, NotFace]:

$$\begin{array}{cc} & \begin{array}{cccc} \textbf{D} & \textbf{H} & \textbf{S} & \textbf{C} \end{array} \\ \begin{array}{c} \textbf{F} \\ \textbf{NF} \end{array} & \begin{pmatrix} F, D & F, H & F, S & F, C \\ NF, D & NF, H & NF, S & NF, C \end{pmatrix} \end{array}.$$

Objectives

- We are interested in the number of ways to draw a subset from a larger set.
- Like how many five-card poker hands can be drawn from a 52-card deck?
- This counting is done along two criteria: *with or without tracking the order of selection*, and with or *without replacing chosen units back into the pool for future selection.*
- In this way, the general forms of choice rules combine *ordering* with *counting*.

Ordered With Replacement

- If we have n objects and we want to pick $k < n$ from them, and replace the choice back into the available set each time, then it should be clear that on each iteration there are always n choices.
- By the Fundamental Theorem of Counting, the number of choices is the product of k values of n alternatives:
$$n \times n \times \cdots n = n^k,$$
- This works since if the factorial ordering rule did not decrement.

Ordered Without Replacement

- Suppose again we have n objects and we want to pick $k < n$ from them.
- There are n ways to pick the first object, $n-1$ ways to pick the second object, $n-2$ ways to pick the third object, and so on until we have k choices.
- This decrementing of choices differs from the last case because we are not replacing items on each iteration.
- So the general form of ordered counting, without replacement using the two principles is
$$n \times (n-1) \times (n-2) \times \cdots \times (k+1) \times k = \frac{n!}{(n-k)!},$$
- The factorial notation saves us a lot of trouble because we can express this list as the difference between $n!$ and the factorial series that starts with $k-1$.
- So the denominator, $(n-k)!$, strips off terms lower than k in the product.

Unordered Without Replacement

- This is just like ordered without replacement, except that we cannot see the order of picking.
- For example, if we were picking colored balls out of an urn, then *red,white,red* is equivalent to *red,red,white* and *white,red,red*.
- Therefore, there are $k!$ fewer choices than with ordered, without replacement since there are $k!$ ways to express this redundancy. So we need only to modify the previous form according to

$$\frac{n!}{(n-k)!k!} = \binom{n}{k}.$$

- This is the "choose" notation from earlier.
- The abbreviated notation is handy because unordered, without replacement is an extremely common sampling procedure.

Unordered Without Replacement

- We can derive a useful generalization of this idea by first observing that
$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$
- This form suggests successively peeling off $k-1$ iterates to form a sum:
$$\binom{n}{k} = \sum_{i=0}^{k} \binom{n-1-i}{k-i}.$$
- If we consider J subgroups labeled $k_1, k_2, \ldots, k_J$ with the property that $\sum_{j=1}^{J} k_j = n$, then we get the more general form
$$\frac{n!}{\prod_{j=1}^{J} k_j!} = \binom{n}{k_1}\binom{n-k_1}{k_2}\binom{n-k_1-k_2}{k_3}\cdots\binom{n-k_1-k_2-\cdots-k_{J-2}}{k_{J-1}}\binom{k_J}{k_J},$$
which can be denoted $\binom{n}{k_1,k_2,\ldots,k_J}$.

Unordered With Replacement

- The best way to think of this is that unordered, without replacement needs to be adjusted upward to reflect the increased number of choices.

- This form is best expressed using choose notation:

$$\frac{(n+k-1)!}{(n-1)!k!} = \frac{(n+k-1-k)!}{(n-1-k)!k!} = \binom{n+k-1}{k}.$$

Survey Sampling

- Suppose we want to perform a small survey with 15 respondents from a population of 150.
- How different are our choices with each counting rule?

$$\begin{aligned}
&\text{Ordered, with replacement:} && n^k = 150^{15} = 4.378939 \times 10^{32} \\
&\text{Ordered, without replacement:} && \tfrac{n!}{(n-k)!} = \tfrac{150!}{135!} = 2.123561 \times 10^{32} \\
&\text{Unordered, without replacement:} && \tbinom{n}{k} = \tbinom{150}{15} = 1.623922 \times 10^{20}. \\
&\text{Unordered, with replacement:} && \tbinom{n+k-1}{k} = \tbinom{164}{15} = 6.59974 \times 10^{20}.
\end{aligned}$$

The Binomial Theorem and Pascal's Triangle

- The most common mathematical use for the choose notation is in the following theorem, which relates exponentiation with counting.
- Given any real numbers X and Y and a nonnegative integer n,

$$(X+Y)^n = \sum_{k=0}^{n} \binom{n}{k} X^k Y^{n-k}.$$

- An interesting special case occurs when $X = 1$ and $Y = 1$:

$$2^n = \sum_{k=0}^{n} \binom{n}{k},$$

  which relates the exponent function to the summed binomial.

The Binomial Theorem and Pascal's Triangle

- To show how rapidly the binomial expansion increases in polynomial terms, consider the first six values of n:

$$\begin{aligned}
(X+Y)^0 &= 1\\
(X+Y)^1 &= X+Y\\
(X+Y)^2 &= X^2+2XY+Y^2\\
(X+Y)^3 &= X^3+3X^2Y+3XY^2+Y^3\\
(X+Y)^4 &= X^4+4X^3Y+6X^2Y^2+4XY^3+Y^4\\
(X+Y)^5 &= X^5+5X^4Y+10X^3Y^2+10X^2Y^3+5XY^4+Y^5.
\end{aligned}$$

- Note the symmetry of these forms.

The Binomial Theorem and Pascal's Triangle

- In fact, if we just display the coefficient values and leave out exponents and variables for the moment, we get *Pascal's Triangle*:

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

- This gives a handy form for summarizing binomial expansions (it can obviously go on further than shown here).

Properties of Pascal's Triangle

- Any value in the table is the sum of the two values diagonally above.
- For instance, 10 in the third cell of the bottom row is the sum of the 4 and 6 diagonally above.
- The sum of the kth row (counting the first row as the zero row) can be calculated by $\sum_{j=0}^{k} \binom{k}{j} = 2^k$.
- The sum of the diagonals from left to right: $\{1\}$, $\{1\}$, $\{1,1\}$, $\{1,2\}$, $\{1,3,1\}$, $\{1,4,3\}$,. . . , give the Fibonacci numbers (1,2,3,5,8,13,. . .).
- If the first element in a row after the 1 is a prime number, then every number in that row is divisible by it (except the leading and trailing 1's).
- If a row is treated as consecutive digits in a larger number (carrying multidigit numbers over to the left), then each row is a power of 11:

$$\begin{array}{rrr} 1 = 11^0 & 11 = 11^1 & 121 = 11^2 \\ 1331 = 11^3 & 14641 = 11^4 & 161051 = 11^5, \end{array}$$

  and these are called the "magic 11's."

Primes

- Prime numbers have been interesting to mathematicians for thousands of years.
- How many primes are there?
- How would we find some?

```
find.primes <- function(max)  {
    num.vec <- seq(1,max,by=2)
    if( max > 5) primes <- 3  else(stop("min of max =5"))
    for (i in 3:length(num.vec))  {
        if(min( num.vec[i] %% primes != 0 ))
                primes <- c(primes,num.vec[i])
    }
    return(c(1,primes))
}
```

Primes

```
find.primes(500)
```

```
 [1]   1   3   5   7  11  13  17  19  23  29  31  37  41  43  47  53  59  61
[19]  67  71  73  79  83  89  97 101 103 107 109 113 127 131 137 139 149 151
[37] 157 163 167 173 179 181 191 193 197 199 211 223 227 229 233 239 241 251
[55] 257 263 269 271 277 281 283 293 307 311 313 317 331 337 347 349 353 359
[73] 367 373 379 383 389 397 401 409 419 421 431 433 439 443 449 457 461 463
[91] 467 479 487 491 499
```

Sets

- Sets are holding places.
- A *set* is a bounded collection defined by its contents (or even by its lack thereof) and is usually denoted with curly braces.
- The set of even positive integers less than 10 is

$$\{2, 4, 6, 8\}.$$

- The set of prime numbers between 10 and 20 is:

$$\{11, 13, 17, 19\}.$$

Sets

- We can also define sets without necessarily listing all the contents if there is some criteria that defines the contents.

- For example,
$$\{X\!:\!0 \leq X \leq 10, X \in \mathfrak{R}\}$$
defines the set of all the real numbers between zero and 10 inclusive.

- We can read this statement as "the set that contains all values labeled X such that X is greater than or equal to zero, less than or equal to 10, and part of the real numbers."

- Sets with an infinite number of members need to be described in this fashion rather than listed out as above.

Elements of Sets

- The "things" that are contained within a set are called *elements*, and these can be individual units or multiple units.
- An *event* is any collection of possible outcomes of an experiment, that is, any subset of the full set of possibilities, including the full set itself
- "Event" and "outcome" are used synonymously.
- Define $\{H\}$ and $\{T\}$ as outcomes for a coin flipping experiment, as is $\{H, T\}$.
- Events and sets are typically, but not necessarily, labeled with capital Roman letters: A, B, T, etc.

Elements of Sets

- Events can be abstract in the sense that they may have not yet happened but are imagined, or outcomes can be concrete in that they are observed: "A occurs."
- Events are also defined for more than one individual subelement (odd numbers on a die, hearts out of a deck of cards, etc.).
- Such defined groupings of individual elements constitute an event in the most general sense.
- For example, a single die:
  - Throw a single die.
  - The event that an even number appears is the set $A = \{2, 4, 6\}$.
- Events can also be referred to when they do *not* happen.
- For the example above we can say "if the outcome of the die is a 3, then A did not occur."

Characteristics of Sets

- Suppose we conduct some experiment such as roll a die, toss a coin, or spin a pointer.
- The *sample space* $\mathcal{S}$ of a given experiment is the set that consists of all possible outcomes (events) from this experiment.
- The sample space from flipping a coin is $\{H, T\}$ (provided that we preclude the possibility that the coin lands on its edge).
- A *countable set* is one whose elements can be placed in one-to-one correspondence with the positive integers.
- A *finite set* has a noninfinite number of contained events.
- Countability and finiteness (or their opposites) are not contradictory characteristics.

Examples of Sets

- **Countably Finite Set**. A single throw of a die is a countably finite set,

$$S = \{1, 2, 3, 4, 5, 6\}.$$

- **Multiple Views of Countably Finite**. Tossing a pair of dice is also a countably finite set, but we can consider the sample space in three different ways. If we are just concerned with the sum on the dice (say for a game like craps), the sample space is

$$S = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

  If the individual values matter, then the sample space is extended to a large set:

$$\begin{array}{cccccc}
\{1,1\}, & \{1,2\}, & \{1,3\}, & \{1,4\}, & \{1,5\}, & \{1,6\}, \\
 & \{2,2\}, & \{2,3\}, & \{2,4\}, & \{2,5\}, & \{2,6\}, \\
 & & \{3,3\}, & \{3,4\}, & \{3,5\}, & \{3,6\}, \\
 & & & \{4,4\}, & \{4,5\}, & \{4,6\}, \\
 & & & & \{5,5\}, & \{5,6\}, \\
 & & & & & \{6,6\}.
\end{array}$$

Examples of Sets

- Also, if we have some way of distinguishing between the two dice, such as using different colors, then the sample space is even larger because it is now possible to distinguish order:

$$\begin{array}{cccccc}
\{1,1\}, & \{1,2\}, & \{1,3\}, & \{1,4\}, & \{1,5\}, & \{1,6\}, \\
\{2,1\}, & \{2,2\}, & \{2,3\}, & \{2,4\}, & \{2,5\}, & \{2,6\}, \\
\{3,1\}, & \{3,2\}, & \{3,3\}, & \{3,4\}, & \{3,5\}, & \{3,6\}, \\
\{4,1\}, & \{4,2\}, & \{4,3\}, & \{4,4\}, & \{4,5\}, & \{4,6\}, \\
\{5,1\}, & \{5,2\}, & \{5,3\}, & \{5,4\}, & \{5,5\}, & \{5,6\}, \\
\{6,1\}, & \{6,2\}, & \{6,3\}, & \{6,4\}, & \{6,5\}, & \{6,6\}.
\end{array}$$

- Note that however we define our sample space here, that definition does not affect the probabilistic behavior of the dice. That is, they are not responsive in that they do not change physical behavior due to the game being played.

Examples of Sets

- **Countably Infinite Set**. The number of coin flips until two heads in a row appear is a countably infinite set:

$$\mathcal{S} = \{1, 2, 3, \ldots\}.$$

- **Uncountably Infinite Set**. Spin a pointer and look at the angle in radians. Given a hypothetically infinite precision measuring instrument, this is an uncountably infinite set:

$$\mathcal{S} = [0{:}2\pi).$$

Cardinality of a Set

- *Cardinality* of a set is just the number of elements in the set.
- The finite set A has cardinality given by $n(A)$, $\bar{A}$, or $\|A\|$, where the first form is preferred.
- For finite sets the cardinality is an integer value denoting the quantity of events (exclusive of the null set).
- The cardinality of a countably infinite set is denoted by $\aleph_0$ (the Hebrew aleph character with subscript zero).
- The cardinality of an uncountably infinite set is denoted similarly by $\aleph_1$.

The Empty Set

- The *empty set*, or *null set* is a set with no elements.
- This seems a little paradoxical since if there is nothing in the set, should not the set simply go away?
- We *need* the idea of an empty set to describe certain events that do not exist and, therefore the empty set is convenient.
- Usually the empty set is denoted with the Greek letter phi: ϕ.
- When there is only one item left in the set, the set is called a *singleton*.

Subsets

- We can perform basic operations on sets that define new sets or provide arithmetic and boolean (true/false) results.
- Set A is a *subset* of set B if every element of A is also an element of B.
- We also say that A is contained in B and denote this as $A \subset B$ or $B \supset A$:

  $$A \subset B \iff \forall X \in A, X \in B,$$

  which reads "A is a subset of B if and only if all values X that are in A are also in B."
- The set A here is a *proper subset* of B if it meets this criteria *and* $A \neq B$.

Equality of Sets

- Equality of sets means that they must contain exactly the same elements.
- To formally assert that two sets are equal we need to claim, however, that both $A \subset B$ and $B \subset A$ are true so that the contents of A exactly match the contents of B:

$$A = B \Longleftrightarrow A \subset B \text{ and } B \subset A.$$

Unions of Sets

- Sets can be "unioned," meaning that they can be combined to create a set that is the same size or larger.
- The *union* of the sets A and B, $A \cup B$, is the new set that contains all of the elements that belong to either A or B.
- The key word in this definition is "or," indicating that the new set is inclusive.
- The union of A and B is the set of elements X whereby

$$A \cup B = \{X : X \in A \text{ or } X \in B\}.$$

Unions of Sets

- The union operator is certainly not confined to two sets, and we can use a modification of the "$\cup$" operator that resembles a summation operator in its application:

$$A_1 \cup A_2 \cup \ldots \cup A_n = \bigcup_{i=1}^{n} A_i.$$

- It is sometimes convenient to specify ranges, say for $m < n$, with the union operator:

$$A_1 \cup A_2 \cup \ldots \cup A_m = \bigcup_{i \leq m} A_i.$$

Unions of Sets

- There is a relationship between unions and subsets: An individual set is always a subset of the new set defined by a union with other sets:

$$A_1 \subset A \Longleftrightarrow A = \bigcup_{i=1}^{n} A_i,$$

  and this clearly works for other constituent sets besides A_1.

- We can also talk about nested subsets:

$$A_n \uparrow A \Longrightarrow A_1 \subset A_2 \subset \dots A_n, \text{ where } A = \bigcup_{i=1}^{n} A_i$$

$$A_n \downarrow A \Longrightarrow A_n \subset A_{n-1} \subset \dots A_1, \text{ where } A = \bigcup_{i=1}^{n} A_i.$$

Intersection of Sets

- The *intersection* of sets contains only those elements found in both (or all for more than two sets i of interest).
- So $A \cap B$ is the new set that contains all of the elements that belong to A and B.
- Now the key word in this definition is "and," indicating that the new set is exclusive.
- So the elements of the intersection do not have the luxury of belonging to one set or the other but must now be a member of both.
- The intersection of A and B is the set elements X whereby

$$A \cap B = \{X : X \in A \text{ and } X \in B\}.$$

Intersection of Sets

- Like the union operator, the intersection operator is not confined to just two sets:

$$A_1 \cap A_2 \cap \ldots \cap A_n = \bigcap_{i=1}^{n} A_i.$$

- It is convenient to specify ranges, say for $m < n$, with the intersection operator:

$$A_1 \cap A_2 \cap \ldots \cap A_m = \bigcap_{i \leq m} A_i.$$

Intersection of Sets

- Sets also define complementary sets by the definition of their existence.
- The *complement* of a given set is the set that contains all elements not in the original set.
- The set $A^{\complement}$ (sometimes denoted A' or $\bar{A}$) is defined by:

$$A^{\complement} = \{X : X \notin A\}.$$

- The complement of the null set is the sample space, and vice versa:

$$\phi^{\complement} = \mathcal{S} \quad \text{and} \quad \mathcal{S}^{\complement} = \phi.$$

- The complement of the set with everything has nothing, and the complement of the set with nothing has everything.

Differences Operator

- The *difference operator* defines which portion of a given set is *not* a member of another.
- The difference of A relative to B is the set of elements X whereby

$$A \setminus B = \{X : X \in A \text{ and } X \notin B\}.$$

  The difference operator can also be expressed with intersection and complement notation:

$$A \setminus B = A \cap B^{\complement}.$$

- The difference operator as defined here is not symmetric: it is not necessarily true that $A \setminus B = B \setminus A$.

Symmetric Difference Operator

- The *symmetric difference* further restricts the resulting set, requiring the operator to apply in both directions.

- The symmetric difference of A relative to B and B relative to A is the set

$$A \bigtriangleup B = \{X\!:\!X \in A \text{ and } X \notin B \text{ or } X \in B \text{ and } B \notin A\}.$$

- Because of this symmetry we can also denote the symmetric difference as the union of two "regular" differences:

$$A \bigtriangleup B = (A \setminus B) \cup (B \setminus A) = (A \cap B^{\complement}) \cup (B \cap A^{\complement}).$$

Single Die Experiment

- Throw a single die. For this experiment, define the following sample space and accompanying sets:

$$\begin{aligned} \mathcal{S} &= \{1,2,3,4,5,6\} \\ A &= \{2,4,6\} \\ B &= \{4,5,6\} \\ C &= \{1\}. \end{aligned}$$

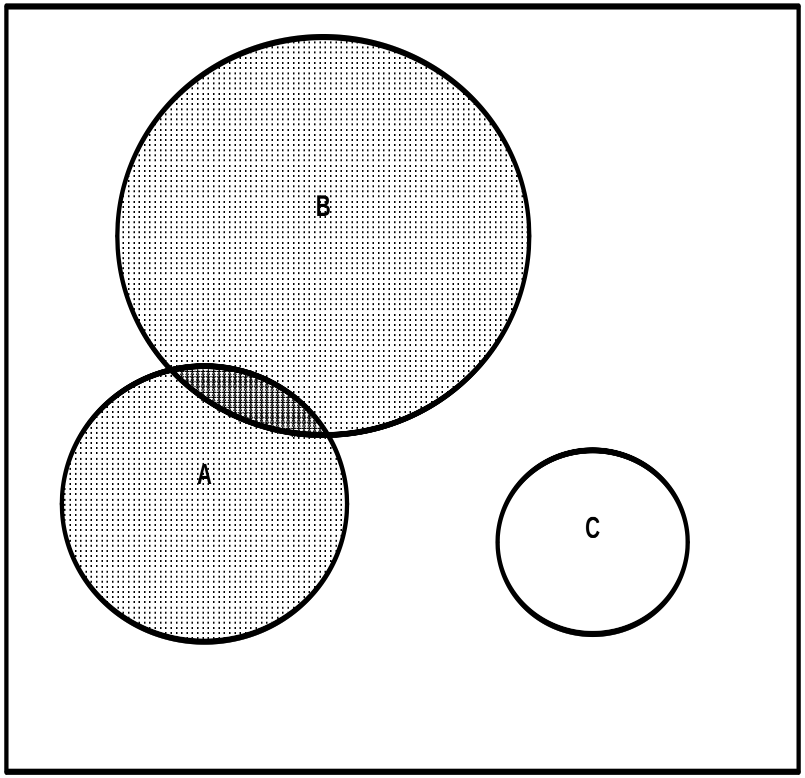
- So A is the set of even numbers, B is the set of numbers greater than 3, and C has just a single element.

- Using the described operators, we find that

$$\begin{array}{lll} A^{\complement} = \{1,3,5\} & A \cup B = \{2,4,5,6\} & A \cap B = \{4,6\} \\ B \cap C = \phi & (A \cap B)^{\complement} = \{1,2,3,5\} & A \setminus B = \{2\} \\ B \setminus A = \{5\} & A \bigtriangleup B = \{2,5\} & (A \cap B) \cup C = \{1,4,6\}. \end{array}$$

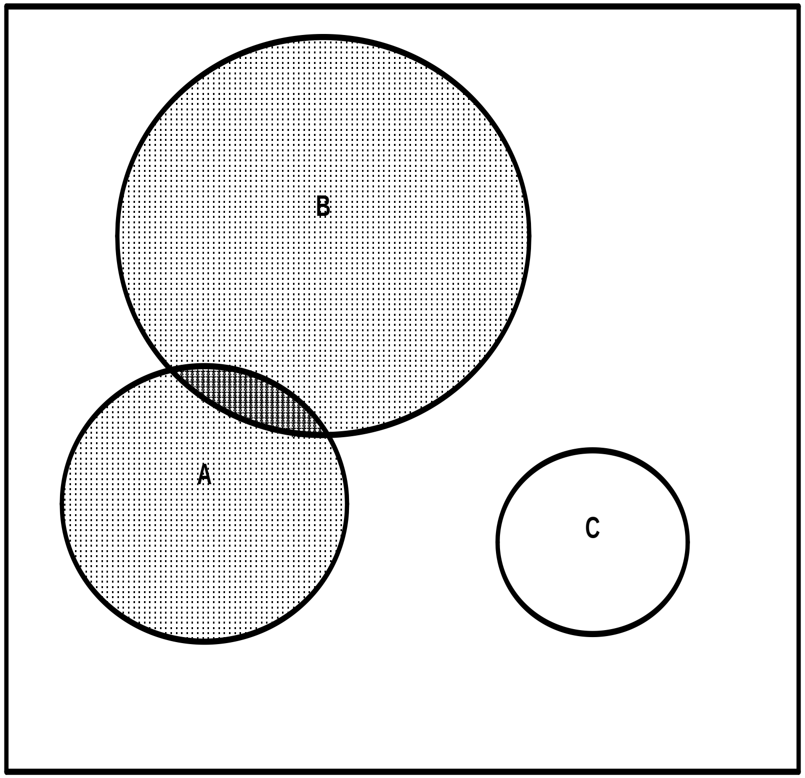
Venn Diagram

- Consider three sets where the "universe" of possible outcomes ($\mathcal{S}$) is given by the surrounding box.
- The intersection of A and B is the dark region that belongs to both sets, whereas the union of A and B is the lightly shaded region that indicates elements in A or B (including the intersection region).



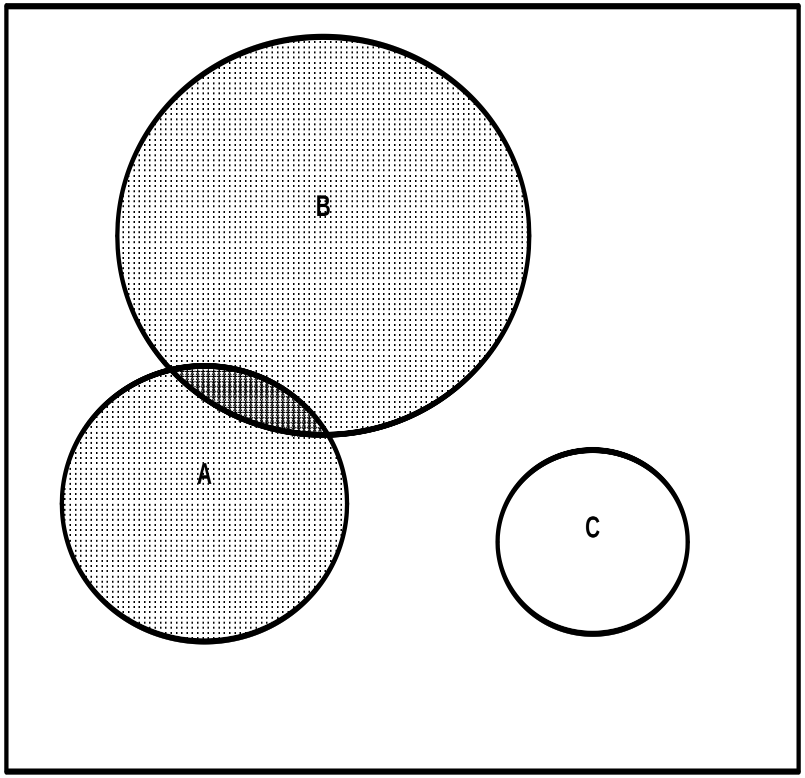
Venn Diagram

- The intersection of A or B with C is ϕ, since there is no overlap.
- The complement of $A \cup B$ is all of the nonshaded region, including C.



Venn Diagram

- The region $(A \cap B)^{\mathsf{C}}$ is every part of $\mathcal{S}$ *except* the intersection, which could also be expressed as those elements that are in the complement of A *or* the complement of B, thus ruling out the intersection.
- The portion of A that does not overlap with B is denoted $A \backslash B$, and we can also identify $A \bigtriangleup B$ in the figure, which is either A or B (the full circles) but not both.



Formal Properties of Set Operations

Properties For Any Three Sets A, B, and C, in $\mathcal{S}$

$\rightarrowtail$ Commutative Property $A \cup B = B \cup A$

$A \cap B = B \cap A$

$\rightarrowtail$ Associative Property $A \cup (B \cup C) = (A \cup B) \cup C$

$A \cap (B \cap C) = (A \cap B) \cap C$

$\rightarrowtail$ Distributive Property $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

$\rightarrowtail$ de Morgan's Laws $(A \cup B)^{\mathsf{C}} = A^{\mathsf{C}} \cap B^{\mathsf{C}}$

$(A \cap B)^{\mathsf{C}} = A^{\mathsf{C}} \cup B^{\mathsf{C}}$

Example Proof

- First show $A \cup (B \cap C) \subset (A \cup B) \cap (A \cup C)$ by demonstrating that some element in the first set is also in the second set:

$$\begin{aligned}
&\text{Suppose } X \in A \cup (B \cap C), \text{ so } X \in A \text{ or } X \in (B \cap C) \\
&\text{If } X \in A, \text{ then } X \in (A \cup B) \text{ and } X \in (A \cup C) \\
&\qquad \therefore X \in (A \cup B) \cap (A \cup C) \\
&\text{Or if } X \notin A, \text{ then } X \in (B \cap C) \text{ and } X \in B \text{ and } X \in C \\
&\qquad \therefore X \in (A \cup B) \cap (A \cup C)
\end{aligned}$$

Example Proof

- Now show $A \cup (B \cap C) \supset (A \cup B) \cap (A \cup C)$ by demonstrating that some element in the second set is also in the first set:

$$\begin{aligned}
&\text{Suppose } X \in (A \cup B) \cap (A \cup C) \text{ so } X \in (A \cup B) \text{ and } X \in (A \cup C)\\
&\text{If } X \in A, \text{ then } X \in A \cup (B \cap C)\\
&\text{Or if } X \notin A, \text{ then } X \in B \text{ and } X \in C,\\
&\qquad\qquad\qquad\text{since it is in the two unions but not in } A.\\
&\therefore X \in A \cup (B \cap C)
\end{aligned}$$

- Since $A \cup (B \cap C) \subset (A \cup B) \cap (A \cup C)$ *and* $A \cup (B \cap C) \supset (A \cup B) \cap (A \cup C)$, then every element in the first set is in the second set, and every element in the second set is in the first set. So the sets must be equal.

Disjoint Sets

- Two sets A and B are *disjoint* when their intersection is empty: $A \cap B = \phi$.
- This is generalizable: the k sets $A_1, A_2, \ldots, A_k$ are *pairwise disjoint*, also called *mutually exclusive*, if $A_i \cap A_j = \phi \ \forall i \neq j$.
- If $A_1, A_2, \ldots, A_k$ are pairwise disjoint and we add the condition that $\bigcup_{i=1}^{k} A_i = \mathcal{S}$ (they cover the sample space completely), then we say that the $A_1, A_2, \ldots, A_k$ are a *partition* of the sample space.
- The outcomes $\{1, 2, 3, 4, 5, 6\}$ form a partition of $\mathcal{S}$ for throwing a single die because they are pairwise distinct and nothing else can occur.
- $A_1, A_2, \ldots, A_k$ are a partition of $\mathcal{S}$ iff
  - $A_i \cap A_j = \phi \ \ \forall i \neq j$.
  - $\bigcup_{i=1}^{k} A_i = \mathcal{S}$.

Overlapping Group Memberships

- Sociologists are often interested in determining connections between distinct social networks.
- Bonacich (1978) used set theory to explore overlapping group memberships such as sports teams, clubs, and social groups in a high school.
- The data are given by the following cross-listing of 18 community members and 14 social events they could possibly have attended.
- An “X” indicated that the individual on that row attended the social event indexed by the column.
- The idea is that the more social events two selected individuals have in common, the closer they are in a social network.

Overlapping Group Memberships

	Event													
Ind.	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	X	X	X	X	X	X		X	X					
2	X	X	X		X		X	X						
3		X	X	X	X	X	X	X	X					
4	X		X	X	X	X	X	X						
5			X	X	X		X							
6			X		X	X		X						
7					X	X	X	X						
8						X		X	X					
9					X		X	X	X					
10							X	X	X			X		
11								X	X	X		X		
12								X	X	X		X	X	X
13							X	X	X	X		X	X	X
14						X	X		X	X	X	X	X	X
15							X	X		X	X	X		
16								X	X					
17									X		X			
18									X		X			

Overlapping Group Memberships

- There are two relatively distinct groups here, with reasonable overlap to complicate things.
- Counting the social events as sets, we can ask some specific questions and make some observations.
- First, observe that only $M = N$; they have the same members: $\{12, 13, 14\}$.
- A number of sets are disjoint, such as (A, J), (B, L), (D, M), and others.
- Yet, the full group of sets, $A\!:\!N$, clearly does not form a partition due to the many nonempty intersections.
- There is no subset of the social events that forms a partition:
  - either I or K would have to be included in the formed partition because they are the only two that include individuals 17 and 18.
  - the set K lacks individual 16, necessitating inclusion of H or I, but these overlap with K elsewhere.
  - I lacks individual 15, but each of the five sets that include this individual overlap somewhere with I.
  - thus no subset can be configured to form a partition.

Overlapping Group Memberships

- Let's now test out the first de Morgan's Law, $(A \cup B)^{\complement} = A^{\complement} \cap B^{\complement}$, for the sets E and G.

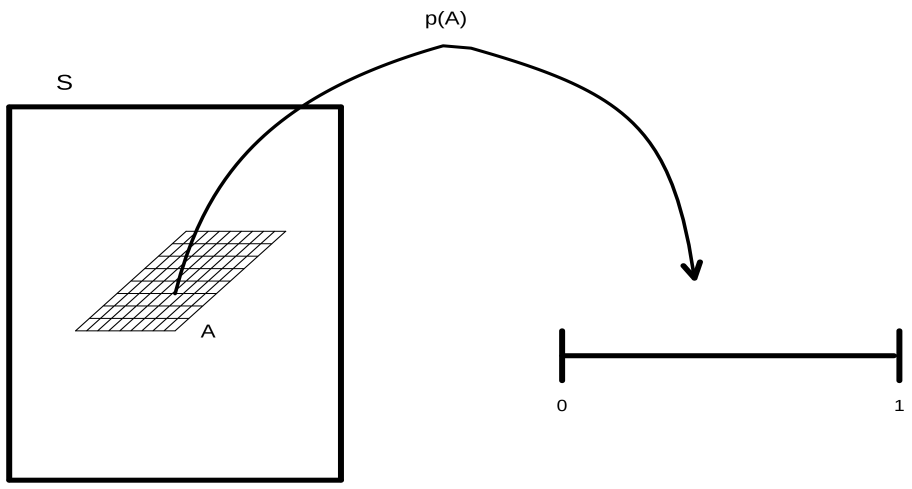
$$\begin{aligned}
(E \cup G)^{\complement} &= (\{1,2,3,4,5,6,7,9\} \cup \{2,3,4,5,7,9,10,13,14,15\})^{\complement} \\
&= (\{1,2,3,4,5,6,7,9,10,13,14,15\})^{\complement} \\
&= \{8,11,12,16,17,18\} \\
E^{\complement} \cap G^{\complement} &= \{1,2,3,4,5,6,7,9\}^{\complement} \cap \{2,3,4,5,7,9,10,13,14,15\}^{\complement} \\
&= \{8,10,11,12,13,14,15,16,17,18\} \\
&\qquad\qquad\qquad\qquad \cap \{1,6,8,11,12,16,17,18\} \\
&= \{8,11,12,,16,17,18\},
\end{aligned}$$

which demonstrates the property.

The Probability Function

- This is a mapping from a defined event (or events) onto a metric bounded by zero (it cannot happen) and one (it will happen with absolute certainty).

- A probability function enables us to discuss various *degrees of likelihood of occurrence* in a systematic and practical way.



Sigma Algebras

- A collection of subsets of the sample space $\mathcal{S}$ is called a *sigma-algebra* (also called a *sigma-field*), and denoted $\mathfrak{F}$ (a fancy looking "F"), if it satisfies the following three properties:
  1. **Null Set Inclusion.** It contains the null set: $\phi \in \mathfrak{F}$.
  2. **Closed Under Complementation.** If $A \in \mathfrak{F}$, then $A^{\complement} \in \mathfrak{F}$.
  3. **Closed Under Countable Unions.** If $A_1, A_2, \ldots \in \mathfrak{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathfrak{F}$.
- So if A is any identified subset of $\mathcal{S}$, then an associated (minimal size) sigma-algebra is $\mathfrak{F} = \{\phi, A, A^{\complement}, \mathcal{S}\}$.

Sigma Algebras

- $\mathfrak{F}' = \{\phi, A, A^{\complement}, A, A, \mathcal{S}, A^{\complement}\}$ because there is no requirement that we not repeat events in a sigma-algebra.
- But this is not terribly useful, so it is common to specify the minimal size sigma-algebra as we have originally done.
- Such a sigma-algebra has a particular name: a *Borel-field*.
- These definitions are of course inherently discrete in measure and they do have corresponding versions over continuous intervals, although the associated mathematics get much more involved.

Sigma Algebras

- For this experiment, flip a coin once.
- This produces

$$\mathcal{S} = \{H, T\}$$
$$\mathfrak{F} = \{\phi, H, T, (H, T)\}.$$

  using the notation.
- Given a sample space $\mathcal{S}$ and an associated sigma-algebra $\mathfrak{F}$, a *probability function* is a mapping, p, from the domain defined by $\mathfrak{F}$ to the interval $[0{:}1]$.

Kolmogorov Axioms

- The *Kolmogorov probability axioms* specify the conditions for a proper probability function:
  - The probability of any realizable event is between zero and one: $p(A_i) \in [0{:}1] \ \ \forall a_i \in \mathfrak{F}$.
  - Something happens with probability one: $p(\mathcal{S}) = 1$.
  - The probability of unions of n pairwise disjoint events is the sum of their individual probabilities: $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} p(A_i)$ (even if $n = \infty$).
- The *triple* (also called a probability space or a probability measure space) consists of $(\mathcal{S}, \mathfrak{F}, P)$, to fully specify the sample space, sigma-algebra, and probability function applied.

Calculations with Probabilities

- The manipulation of probability functions follows logical and predictable rules.
- The probability of a union of two sets is no smaller than the probability of an intersection of two sets.
- These two probabilities are equal if one set is a subset of another.
- It also makes intuitive sense that subsets have no greater probability than the enclosing set:

$$\text{If } A \subset B, \text{ then } p(A) \leq p(B).$$

Calculations with Probabilities

Calculations with Probabilities for A, B, and C, in $\mathcal{S}$

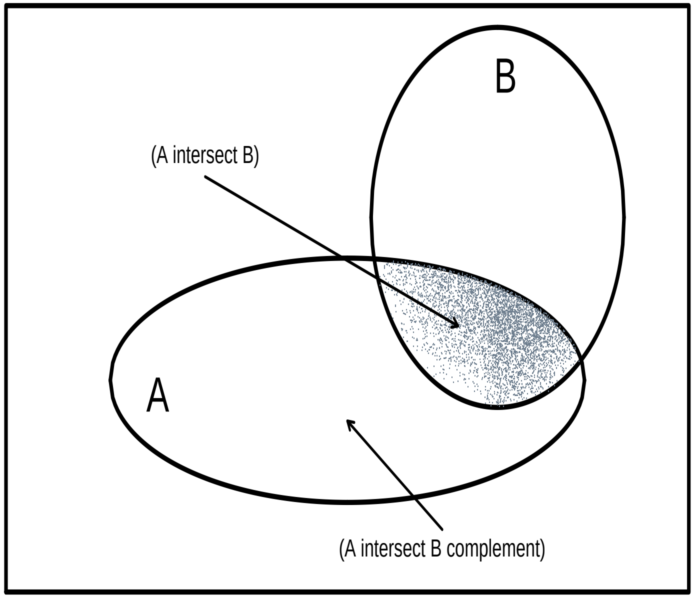
$\rightarrowtail$ Probability of Unions	$p(A \cup B)$ $= p(A) + p(B) - p(A \cap B)$
$\rightarrowtail$ Probability of Intersections	$p(A \cap B)$ $= p(A) + p(B) - p(A \cup B)$ (also denoted $p(A, B)$)
$\rightarrowtail$ Probability of Complements	$p(A^{\complement}) = 1 - p(A)$, $p(A) = 1 - p(A^{\complement})$
$\rightarrowtail$ Probability of the Null Set	$p(\phi) = 0$
$\rightarrowtail$ Probability of the Sample Space	$p(\mathcal{S}) = 1$
$\rightarrowtail$ Boole's Inequality	$p(\bigcup_j A_j) \leq \sum_j p(A_j)$

The Theorem of Total Probability

- Given any events A and B, $p(A) = p(A \cap B) + p(A \cap B^{\complement})$.
- This intuitively says that the probability of an event A can be decomposed into to parts: one that intersects with another set B and the other that intersects with the complement of B.
- If there is no intersection or if B is a subset of A, then one of the two parts has probability zero.
- More generally, if $B_1, B_2, \ldots, B_n$ is a partition of the sample space, then

$$p(A) = p(A \cap B_1) + p(A \cap B_2) + \ldots + p(A \cap B_n).$$

The Theorem of Total Probability



Probabilistic Analysis of Supreme Court Decisions

- Jeffrey Segal (1984) looked at U.S. Supreme Court decisions to review search and seizure cases from lower courts.
- He constructed a model using data from all 123 Fourth Amendment cases from 1962 to 1981 to explain why the Court upheld the lower court ruling versus overturning it.
- The objective was to make probabilistic statements about Supreme Court decisions given specific aspects of the case and therefore to make predictive claims about future actions.
- Since his multivariate statistical model simultaneously incorporates all these variables, the probabilities described are the effects of individual variables holding the effects of all others constant.

Probabilistic Analysis of Supreme Court Decisions

- One of his first findings was that a police search has a 0.85 probability of being upheld by the Court if it took place at the home of another person and only a 0.10 probability of being upheld in the detainee's own home.
- This is a dramatic difference in probability terms and reveals considerable information about the thinking of the Court.
- Another notable difference occurs when the search takes place with no property interest versus a search on the actual person: 0.85 compared to 0.41.
- Relatedly, a "stop and frisk" search case has a 0.70 probability of being upheld whereas a full personal search has a probability of 0.40 of being upheld.
- These probabilistic findings point to an underlying distinction that justices make in terms of the personal context of the search.

Probabilistic Analysis of Supreme Court Decisions

- Segal also found differences with regard to police possession of a warrant or probable cause.
- A search sanctioned by a warrant had a 0.85 probability of being upheld but only a 0.50 probability in the absence of such prior authority.
- The probability that the Court would uphold probable cause searches (where the police notice some evidence of illegality) was 0.65, whereas those that were not probable cause searches were upheld by the Court with probability 0.53.
- This is not a great difference, and Segal pointed out that it is confounded with other criteria that affect the overall reasonableness of the search.
- If the search is performed subject to a lawful arrest, then there is a (quite impressive) 0.99 probability of being upheld, but only a 0.50 probability if there is no arrest, and all the way down to 0.28 if there is an unlawful arrest.

Conditional Probability

- *Conditional probability* provides a means of systematically including other information by changing "$p(A)$" to "$p(A|B)$" to mean the probability that A occurs given that B has occurred.
- Suppose a single die is rolled but it cannot be seen.
- The probability that the upward face is a four is obviously one-sixth, $p(x = 4) = \frac{1}{6}$.
- Further suppose that you are told that the value is greater than three.
- Would you revise your probability statement?
- Obviously it would be appropriate to update since there are now only three possible outcomes, one of which is a four.
- This gives $p(x = 4|x > 3) = \frac{1}{3}$, which is a substantially different statement.

Conditional Probability

- Given two outcomes A and B in $\mathcal{S}$, the probability that A occurs given that B occurs is the probability that A and B both occur divided by the probability that B occurs:

$$p(A|B) = \frac{p(A \cap B)}{p(B)},$$

provided that $p(B) \neq 0$.

Example of Conditional Probability

- In rolling two dice labeled X and Y, we are interested in whether the sum of the up faces is four, given that the die labeled X shows a three.

- The unconditional probability is given by

$$p(X+Y=4) = p(\{1,3\},\{2,2\},\{3,1\}) = \frac{1}{12},$$

  since there are 3 defined outcomes here out of 36 total.

- The conditional probability, however, is given by

$$\begin{aligned} p(X+Y=4|X=3) &= \frac{p(X+Y=4, X=3)}{p(X=3)} \\ &= \frac{p(\{3,1\})}{p(\{3,1\},\{3,2\},\{3,3\}\{3,4\},\{3,5\},\{3,6\})} \\ &= \frac{1}{6}. \end{aligned}$$

Bayes' Law

- We can rearrange $p(A|B) = \frac{p(A \cap B)}{p(B)}$ to get $p(A|B)p(B) = p(A \cap B)$.
- Similarly, for the set $B^{\complement}$, we get $p(A|B^{\complement})p(B^{\complement}) = p(A \cap B^{\complement})$.
- For any set B we know that A has two components, one that intersects with B and one that does not (although either could be a null set).
- So the set A can be expressed as the sum of conditional probabilities:

$$p(A) = p(A|B)p(B) + p(A|B^{\complement})p(B^{\complement}).$$

- Thus the Theorem of Total Probability can also be reexpressed in conditional notation, showing that the probability of any event can be decomposed into conditional statements about any other event.

Bayes' Law

- Suppose now that we are interested in decomposing $p(A|C)$ with regard to another event, B and $B^{\complement}$.
- We start with the definition of conditional probability, expand via the most basic form of the Theorem of Total Probability, and then simplify:

$$\begin{aligned} p(A|C) &= \frac{p(A \cap C)}{p(C)} \\ &= \frac{p(A \cap B \cap C) + p(A \cap B^{\complement} \cap C)}{p(C)} \\ &= \frac{p(A|B \cap C)p(B \cap C) + p(A|B^{\complement} \cap C)p(B^{\complement} \cap C)}{p(C)} \\ &= p(A|B \cap C)p(B|C) + p(A|B^{\complement} \cap C)p(B^{\complement}|C). \end{aligned}$$

Bayes' Law

- We can manipulate the conditional probability statements in parallel:

$$p(A|B) = \frac{p(A \cap B)}{p(B)} \qquad\qquad p(B|A) = \frac{p(B \cap A)}{p(A)}$$

$$p(A \cap B) = p(A|B)p(B) \qquad\qquad p(B \cap A) = p(B|A)p(A).$$

- We know that $p(A \cap B) = p(B \cap A)$, so we can equate

$$\begin{aligned} p(A|B)p(B) &= p(B|A)p(A) \\ p(A|B) &= \frac{p(A)}{p(B)} p(B|A) \\ &= \frac{p(A)p(B|A)}{p(A)p(B|A) + p(A^{\complement})p(B|A^{\complement})}, \end{aligned}$$

  where the last step uses the Total Probability Theorem.

- This means that we have a way of relating the two conditional probability statements.

Bayes' Law

- Any joint probability can be decomposed into a series of conditional probabilities followed by a final unconditional probability using the *multiplication rule*.
- This is a generalization of the definition of conditional probability.
- The joint distribution of k events can be reexpressed as

$$\begin{aligned} p(A_1, A_2, \ldots, A_k) = p(A_k | A_{k-1}, A_{k-2}, \ldots, A_2, A_1) \\ \times p(A_{k-1} | A_{k-2}, \ldots, A_2, A_1) \cdots p(A_3 | A_2, A_1) p(A_2 | A_1) p(A_1). \end{aligned}$$

- So we can "reassemble" the joint distribution form starting from the right-hand side using the definition of conditional probability, giving

$$p(A_2 | A_1) p(A_1) = p(A_1, A_2)$$

Example of Inverse Probability

- The order of conditionality can be really important.
- suspected probability of AIDS in risk group: $P(A) = 0.02$

  probability of correct positive classification: $P(C|A) = 0.95$

  probability of correct negative classification: $P(C^c|A^c) = 0.97$
- Suppose we want $P(A|C)$, from: $$P(A|C) = \frac{P(A)}{P(C)}P(C|A)$$
- Getting the unconditional: $$\begin{aligned} P(C) &= P(C \cap A) + P(C \cap A^c) \\ &= P(C|A)P(A) + P(C|A^c)]P(A^c) \\ &= P(C|A)P(A) + [1 - P(C^c|A^c)]P(A^c) \\ &= (0.95)(0.02) + (1 - 0.97)(0.98) \cong 0.05 \end{aligned}$$

Example of Inverse Probability

- The order of conditionality can be really important.
- suspected probability of AIDS in risk group: $P(A) = 0.02$

  probability of correct positive classification: $P(C|A) = 0.95$

  probability of correct negative classification: $P(C^c|A^c) = 0.97$
- Suppose we want $P(A|C)$, from: $$P(A|C) = \frac{P(A)}{P(C)} P(C|A)$$
- Getting the unconditional: $$\begin{aligned} P(C) &= P(C \cap A) + P(C \cap A^c) \\ &= P(C|A)P(A) + P(C|A^c)]P(A^c) \\ &= P(C|A)P(A) + [1 - P(C^c|A^c)]P(A^c) \\ &= (0.95)(0.02) + (1 - 0.97)(0.98) \cong 0.05 \end{aligned}$$
- So now we can calculate:

$$P(A|C) = \frac{P(A)}{P(C)} P(C|A) = \frac{0.02}{0.05}(0.95) = 0.38$$

Simpson's Paradox

- Sometimes conditioning on another event actually provides *opposite* results from what would normally be expected.
- Suppose, for example, a state initiated a pilot job training program for welfare recipients with the goal of improving skill levels to presumably increase the chances of obtaining employment for these individuals.
- The investigators assign half of the group to the job placement program and leave the other half out as a control group.
- The results for the full group and a breakdown by sex are provided in Table **??**.

Illustration of Simpson's Paradox: Job Training

		Job	No Job	Placement Rate
Full Group, $n = 400$	Training	100	100	50%
	No Training	80	120	40%
Men, $n = 200$	Training	90	60	60%
	No Training	35	15	70%
Women, $n = 200$	Training	10	40	20%
	No Training	45	105	30%

Simpson's Paradox

- Looking at the full group, those receiving the job training are somewhat more likely to land employment than those who did not.
- Yet when we look at these same people divided into men and women, the results are the opposite.
- Now it appears that it is better not to participate in the job training program for both sexes.
- This surprising result is called *Simpson's Paradox*.

Simpson's Paradox

- How can something that is good for the full group be bad for all of the component subgroups?
- A necessary condition for this paradox to arise is for Training and Job to be correlated with each other, *and* Male to be correlated with both Training and Job.
- So more men received job training and more men got jobs.
- Treatment (placement in the program) is confounded with sex.
- Therefore the full group analysis "masks" the effect of the treatment by aggregating the confounding effect out.
- This is also called *aggregation bias* because the average of the group averages is not the average of the full population.

Simpson's Paradox

- We can also analyze this using the conditional probability version of the Total Probability Theorem.
- Label the events J for a job, T for training, and M for male.
- Observe that the $p(J|T) = 0.5$ since a total of 200 individuals got the training and 100 acquired jobs.
- Does this comport with the conditioning variable?

$$
\begin{aligned}
p(J|T) &= p(J|M \cap T)p(M|T) + p(J|M^{\complement} \cap T)p(M^{\complement}|T) \\
&= (0.6)\left(\frac{90+60}{100+100}\right) + (0.2)\left(\frac{10+40}{100+100}\right) \\
&= 0.5.
\end{aligned}
$$

- Thus the explanation for why $p(J|T) > p(J|T^{\complement})$ but $p(J|M \cap T) < p(J|M \cap T^{\complement})$ and $p(J|M^{\complement} \cap T) < p(J|M^{\complement} \cap T^{\complement})$ is the unequal weighting between men and women.

Independence

- In the last section we found that certain events will change the probability of other events and that we are advised to use the conditional probability statement as a way of updating information about a probability of interest.

- How do we treat the first case when it does not change the subsequent probability of interest?

- If all we are interested in is the probability of voting Republican in the next election, then it is obviously reasonable to ignore the information about car color and continue to use whatever probability we had originally assigned.

- But suppose we are interested in the probability that an individual votes Republican (event A) *and* owns a blue car (event B)? This joint probability is just the product of the unconditional probabilities and we say that A and B are *independent* if

$$p(A \cap B) = p(A)p(B).$$

Independence

- A set of events $A_1, A_2, \ldots, A_k$ is *pairwise independent* if

$$p(A_i \cap A_j) = p(A_i)p(A_j) \;\; \forall i \neq j.$$

- This means that if we pick any two events out of the set, they are independent of each other.

- Pairwise independence does *mean* the same thing as general independence, though it is a property now attached to the pairing operation.

Independence

- Consider the following three events for two tosses of a fair coin:
  - **Event A:** Heads appears on the first toss.
  - **Event B:** Heads appears on the second toss.
  - **Event C:** Exactly one heads appears in the two tosses.
- Also we can ascertain that they are each pairwise independent:

$$p(A \cap B) = \frac{1}{4} = p(A)p(B) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$$

$$p(B \cap C) = \frac{1}{4} = p(B)p(C) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$$

$$p(A \cap C) = \frac{1}{4} = p(A)p(C) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right),$$

but they are not independent as a group because

$$p(A \cap B \cap C) = 0 \neq p(A)p(B)p(C) = \frac{1}{8}.$$

Conditional Independence

- Events A and B are *conditionally independent* on event C:

$$p(A \cap B|C) = p(A|C)p(B|C).$$

- Returning to the example above, A and B are not conditionally independent either if the condition is C because

$$p(A \cap B|C) = 0,$$

but

$$p(A|C) = \frac{1}{2}, \qquad p(B|C) = \frac{1}{2},$$

and their product is clearly not zero.

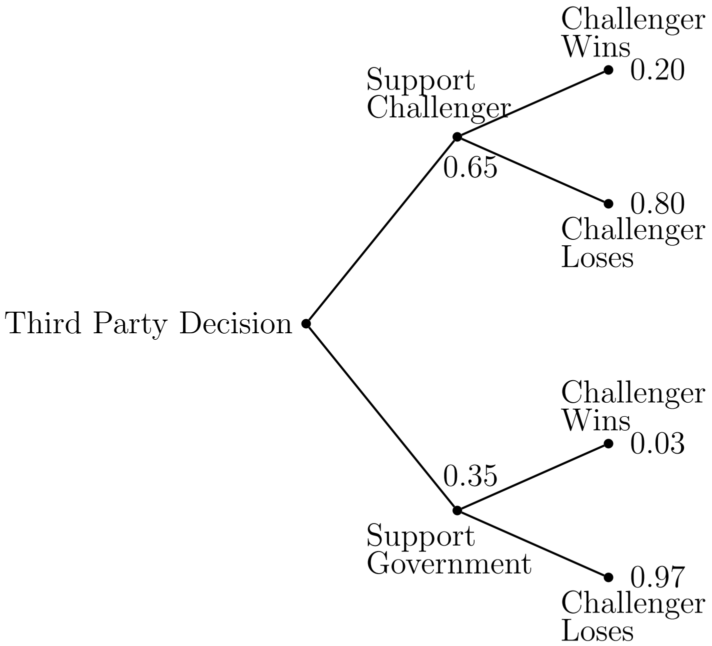
Independence of Functions

- If A and B are independent, then functions of A and B operating on the same domain are also independent.
- As an example, suppose we can generate random (equally likely) integers from 1 to 20.
- Define A as the event that a prime number occurs except the prime number 2:
$$p(A) = p(x \in \{1, 3, 5, 7, 11, 13, 17, 19\}) = \frac{8}{20},$$
and B as the event that the number is greater than 10:
$$p(B) = p(x > 10) = \frac{10}{20}.$$
- Since there are 4 odd primes above and below 10, A and B are independent:
$$p(A \cap B) = \frac{4}{20} = p(A)p(B) = \frac{8}{20}\frac{1}{2}.$$
- For example, use the simple functions $p(g(A)) = p(A^{\complement}) = \frac{12}{20}$ and $p(h(B)) = p(B^{\complement}) = \frac{1}{2}$.
- Then $\frac{12}{20} \times \frac{1}{2} = \frac{3}{10} = p(g(A) \cap (h(B))$.

Analyzing Politics with Decision Trees

- Bueno de Mesquita, Newman, and Rabushka (1985) suggested a method for forecasting political events and applied it to the test case of Hong Kong's reversal to China (PRC).
- The decision tree in Figure **??** shows the possible decisions and results for a third party's decision to support either the challenger or the government in an election, given that they have ruled out doing nothing.
- Suppose we were trying to anticipate the behavior of this party and the resulting impact on the election (presumably the support of this party matters).

Figure 1: Marginal Contribution of Third Parties



Analyzing Politics with Decision Trees

Analyzing Politics with Decision Trees

- Hypothetical probabilities of each event at the nodes are given in the figure.
- The probability that the challenger wins is 0.03 and the probability that the challenger loses is 0.97, *after the third party has already thrown its support behind the government.*
- The probability that challenger wins is 0.20 and the probability that the challenger loses is 0.80, *after the third party has already thrown its support behind the challenger.*
- So these are conditional probabilities exactly as we have studied before but now show diagrammatically.
- In standard notation these are

$$p(C|SC) = 0.20 \qquad p(G|SC) = 0.80$$
$$p(C|SG) = 0.03 \qquad p(G|SG) = 0.97,$$

where we denote C as challenger wins, G as government wins, SG for the third party supports the government, and SC for the third party supports the challenger.

Analyzing Politics with Decision Trees

- As an analyst of future Hong Kong elections, one might be estimating the probability of either action on the part of the third party, and these are given here as $p(SC) = 0.65$ and $p(SG) = 0.35$.
- In other words, our study indicates that this party is somewhat more inclined to support the opposition.
- So what does this mean for predicting the eventual outcome?
- We have to multiply the probability of getting to the first node of interest times the probability of getting to the outcome of interest, and we have to do this for the entire tree to get the full picture.
- From the tree, the probability that the challenger wins when the third party supports them is $(0.20)(0.65) = 0.13$.
- Conceptually we can rewrite this in the form $p(C|SC)p(SC)$, and we know that this is really $p(C, SC) = p(C|SC)p(SC)$.

Odds

- Sometimes probability statements are reexpressed as *odds*.
- In some academic fields this is routine and researchers view these to be more intuitive than standard probability statements.
- Consider a sample space with only two outcomes: success and failure.
- These can be defined for any social event we like: wars, marriages, group formations, crimes, and so on.
- Defining the probability of success as $p = p(S)$ and the probability of failure as $q = 1 - p = p(F)$, the odds of success are

$$\text{odds}(S) = \frac{p}{q}.$$

Odds

- While the probability metric is confined to [0:1], odds are positive but unbounded.
- Often times odds are given as integer comparisons: "The odds of success are 3 to 2" and notated 3:2, and if it is convenient, making the second number 1 is particularly intuitive.
- Converting probabilities to odds does not lose any information: if $odds(S) = \alpha{:}\beta$, then the two probabilities are

$$p(S) = \frac{\alpha}{\alpha + \beta} \qquad\qquad p(F) = \frac{\beta}{\alpha + \beta}.$$

Parental Involvement for Black Grandmothers

- Pearson et al. (1990) researched the notion that black grandparents, typically grandmothers, living in the same household are more active in parenting their grandchildren than their white counterparts.
- The authors were concerned with testing differences in extended family systems and the roles that members play in child rearing.
- They obtained data on 130 black families where the grandmother lived in the house and with reported levels of direct parenting for the grandchildren.
- Three dichotomous (yes/no) effects were of direct interest here:
  - *Supportive behavior* was defined as reading bedtime stories, playing games, or doing a pleasant outing with the child. This was the main variable of interest to the researchers.
  - The first supporting variable was *punishment behavior*, which was whether or not the grandmother punished the child on misbehavior.
  - The second supporting variable was *controlling behavior*, which meant that the grandmother established the rules of behavior for the child.

Parental Involvement for Black Grandmothers

- Pearson et al. looked at a wide range of explanations for differing levels of grandmother involvement, but the two most interesting findings related to these variables.
- Grandmothers who took the punishment behavior role versus not taking on this role had an odds ratio of 2.99:1 for exhibiting supportive behavior.
- Furthermore, grandmothers who took the controlling behavior role versus not doing so had an odds ratio of 5.38:1 for exhibiting supportive behavior.
- Therefore authoritarian behavior strongly predicts positive parenting interactions.