MATH442601|2 Notebook 1 Fall 2018/2019

prepared by Professor Jenny Baglivo

1 MATH442601|2 Notebook 1

This notebook is concerned with introductory probability concepts. The notes correspond to material in Chapter 1 of the Rice textbook.

1.1 Introduction

Probability is the study of random phenomena. Probability theory can be applied, for example, to study games of chance (e.g. roulette games, card games), occurrences of catastrophic events (e.g. tornados, earthquakes), survival of animal species, the relationship between genetic variation and disease, and changes in stock and commodity markets.

1.1.1 Definitions

1. The term *experiment* (or *random experiment*) is used in probability theory to describe a procedure whose outcome is not known in advance with certainty. Further, experiments are assumed to be repeatable (at least in theory) and to have a well-defined set of possible outcomes.
2. The *sample space* Ω is the set of all possible outcomes of an experiment. An *event* is a subset of the sample space. A *simple event* is an event with a single outcome. Events are usually denoted by capital letters (A, B, C, $\ldots$) and outcomes by lower case letters (x, y, z, $\ldots$). If $x \in A$ is observed, then A is said to have *occurred.* The favorable outcomes of an experiment form the event of interest.
3. Each repetition of an experiment is called a *trial.* *Repeated trials* are repetitions of the experiment using the specified procedure, with the outcomes of the trials having no influence on one another.

Example 1. You toss a fair coin five times and record h (for head) or t (for tail) each time. The sample space is the collection of $32 = 2^5$ sequences of 5 h's or t's:

$$\begin{aligned}\Omega \;= \{&hhhhh,\; hhhht,\; hhhth,\; hhthh,\; hthhh,\; thhhh,\; hhhtt,\; hhtht,\\ &hthht,\; thhht,\; hhtth,\; hthth,\; thhth,\; htthh,\; ththh,\; tthhh,\\ &ttthh,\; ttthh,\; thtth,\; httth,\; tthht,\; ththt,\; httht,\; thhtt,\\ &hthtt,\; hhttt,\; htttt,\; thttt,\; tthtt,\; tttht,\; tttth,\; ttttt\}\end{aligned}$$

If you are interested in getting exactly 5 heads, then the event of interest is the simple event $A = \{hhhhh\}$. If you are interested in getting exactly 3 heads, then the event of interest is

$$A = \{hhhtt, hhtht, hthht, thhht, hhtth, hthth, thhth, htthh, ththh, tthhh\}.$$

Footnote: I let the computer simulate 500 repetitions of the experiment described above. The following table gives the number of times $0, 1, \ldots 5$ heads were observed:

Number of Heads	0	1	2	3	4	5
Number of Times	12	75	159	160	81	13

In particular, exactly 3 heads were observed 32% (160/500) of the time.

Example 2. You toss a fair coin until you get a tail and record the sequence of h's and t's. The sample space is

$$\Omega = \{t,\ ht,\ hht,\ hhht,\ hhhht,\ \ldots\}.$$

If you are interested in obtaining a tail in three or fewer tosses, then the event of interest is

$$A = \{t,\ ht,\ hht\}.$$

Footnote: I let the computer simulate 500 repetitions of the experiment described above. The following table gives the number of heads in each sequence, the number of elements in each sequences, and the number of times each type of sequence was observed.

Number of Heads	0	1	2	3	4	5	6	7	8	9	10	11
Number of Elements	1	2	3	4	5	6	7	8	9	10	11	12
Number of Times	236	118	70	33	21	10	6	2	2	0	1	1

Note that the longest observed sequence had 12 elements. In addition, sequences of length 3 or less were observed 84.8% (424/500) of the time.

Example 3. Consider a wall that is 10 feet wide and 8 feet tall, and a circular dart board of radius 2 feet located at the center of the wall. Assume that each time you throw a dart in the direction of the board, the point of impact is somewhere on the wall. Assume further that your aim is so poor that the point of impact is as likely to be in any subregion of the wall as in any other subregion of the same size.

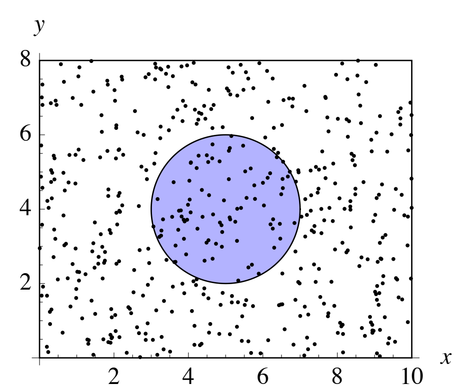
You throw a dart once and record the coordinates of the point of impact, using the coordinate system shown below. The sample space is the collection of points

$$\Omega = \{(x, y) \mid 0 \leq x \leq 10,\ 0 \leq y \leq 8\}.$$

If you are interested in hitting the dart board, then the event of interest is

$$A = \{(x, y) \mid (x - 5)^2 + (y - 4)^2 \leq 4\}.$$

Footnote: I let the computer simulate 500 repetitions of the experiment using these assumptions. The following graphic shows the locations of the "hits" for the 500 repetitions.



In this simulation, the dart hit the dart board 17% (85/500) of the time.

Example 4. Consider choosing a pair of numbers, (x, y), completely at random from the unit square $[0, 1] \times [0, 1]$ and calculating their sum $z = x + y$. If you are interested in getting a sum between 0.50 and 1.50, then the sample space is the collection of points

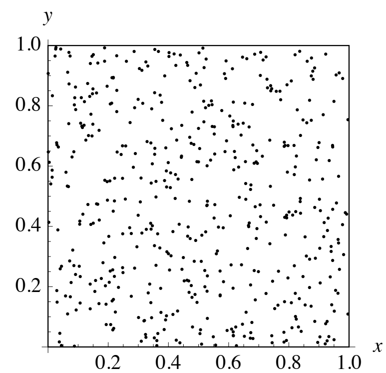
$$\Omega = \{(x, y) \mid 0 \le x \le 1,\ 0 \le y \le 1\}$$

and the event of interest is $A = \{(x, y) \mid 0.50 \le x + y \le 1.50\}$.

Footnote: I let the computer simulate 500 repetitions of the experiment using these assumptions. Here are the numerical and graphical results:

$0 \le z \le 0.50$	63
$0.50 \le z \le 1.00$	196
$1.00 \le z \le 1.50$	193
$1.50 \le z \le 2.00$	48

In this simulation, the sum of coordinates was in the $[0.50, 1.50]$ interval 77.8% (389/500) of the time.



1.2 Probability Distributions

The basic rules (or axioms) of probability were introduced by A. Kolmogorov in the 1930's.

1.2.1 Kolmogorov Axioms

Let $A \subseteq \Omega$ be an event, and $P(A)$ be the probability that A will occur.

A *probability measure*, or simply a *probability*, on a sample space Ω is a specification of numbers $P(A)$ satisfying the following axioms:

1. ***Full Space Rule:*** $P(\Omega) = 1$.
2. ***Range Rule:*** If A is an event, then $0 \le P(A) \le 1$.
3. ***Disjoint Union Rule:*** If A_1 and A_2 are *disjoint events* (that is, if $A_1 \cap A_2 = \emptyset$), then

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$

.........continued on the next page

3′. ***Pairwise Disjoint Union Rule:*** More generally, if A_1, A_2, ... are *pairwise disjoint events* (that is, $A_i \cap A_j = \emptyset$ when $i \neq j$), then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots .$$

If the sequence of events is infinite, then the right hand side is understood to be the sum of a convergent infinite series.

Since Ω is the set of all possible outcomes, an outcome in Ω is certain to occur; the probability of an event that is certain to occur must be 1 (axiom 1). Probabilities must be between 0 and 1 (axiom 2), and probabilities must be additive when events are pairwise disjoint (axiom 3).

1.2.2 Properties of Kolmogorov Axioms

The following properties can be proven using the Kolmogorov axioms.

A. ***Complement Rule:*** Let A^c be the complement of A in Ω, $A^c = \Omega \setminus A$. Then

$$P(A^c) = 1 - P(A).$$

B. ***Empty Set Rule:*** $P(\emptyset) = 0$.

C. ***Subset Rule:*** If A is a subset of B, then $P(A) \leq P(B)$.

D. ***Inclusion-Exclusion Rule:*** If A and B are events, then

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

Exercise. Use the Kolmogorov axioms to prove the complement and empty set rules.

Exercise. Use the Kolmogorov axioms to prove the subset and inclusion-exclusion rules.

Exercise. Use the axioms and properties to answer each question below.

(a) Suppose that $P(A) = 0.3$, $P(B) = 0.9$, and $A \cup B = \Omega$.
Compute $P(A \cap B)$, $P(A^c)$, $P(B^c)$, and $P(A^c \cap B^c)$.

(b) Suppose that $P(A) = 0.3$, $P(B) = 0.6$ and $P(A \cup B^c) = 0.5$.
Compute $P(A \cap B)$ and $P(A \cup B)$.

1.2.3 Long Run Interpretation of Probability

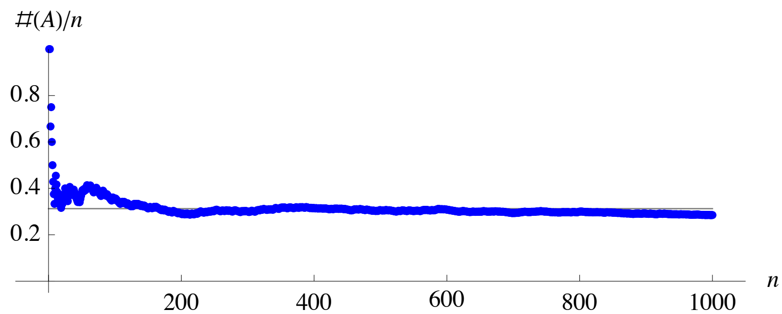
The probability of an event can be interpreted as the "limit" of relative frequencies as the number of repetitions grows large:

$$\frac{\#(A)}{n} \longrightarrow P(A), \text{ as } n \to \infty,$$

where $\#(A)$ is the number of occurrences of event A in n repeated trials of the experiment.

(In the long run, the observed proportion of time event A is observed is close to $P(A)$.)

Example 1, continued. Consider tossing a fair coin 5 times and recording h (for head) or t (for tail) each time, and let A be the event that you observe exactly 3 heads in the five tosses. I let the computer simulate the experiment 1000 times. The following plot shows the sequence of relative frequencies for this simulation:



Exercise. Consider again the four examples in Section 1.1.1 (page 3). Use the descriptions of the experiments to guess the value of $P(A)$ in each case.

1.2.4 Equally Likely Outcomes, Classical Definition of Probability

Suppose that Ω is a finite set with N elements, and $A \subseteq \Omega$ is an event with n elements.

If each outcome is equally likely, then the probability of A is

$$P(A) = \frac{|A|}{|\Omega|} = \frac{n}{N}.$$

In this formula, $|S|$ represents the number of elements in the finite set S.

Exercise. Eight chips numbered $1, 2, \ldots, 8$ are placed in an urn.

(*a*) *Sample Two Chips With Replacement:* An experiment consists of thoroughly mixing the contents of the urn, removing one chip and recording the number on the chip, replacing the chip to the urn, thoroughly mixing the contents again, and removing one chip and recording the number on the chip.

The sample space for the experiment can be represented as follows:

$$\begin{array}{lllllllllllllllll}
\Omega = \{11, & 12, & 13, & 14, & 15, & 16, & 17, & 18, & 21, & 22, & 23, & 24, & 25, & 26, & 27, & 28, \\
31, & 32, & 33, & 34, & 35, & 36, & 37, & 38, & 41, & 42, & 43, & 44, & 45, & 46, & 47, & 48, \\
51, & 52, & 53, & 54, & 55, & 56, & 57, & 58, & 61, & 62, & 63, & 64, & 65, & 66, & 67, & 68, \\
71, & 72, & 73, & 74, & 75, & 76, & 77, & 78, & 81, & 82, & 83, & 84, & 85, & 86, & 87, & 88\}
\end{array}$$

Let A be the event that the sum of the numbers is 6. Then

$$A = \{\underline{\hspace{6cm}}\},$$

and $P(A) =$ __________________________.

(*b*) *Sample Two Chips Without Replacement:* An experiment consists of thoroughly mixing the contents of the urn, removing one chip and recording the number on the chip, thoroughly mixing the remaining contents, and removing one chip and recording the number on the chip.

The sample space for the experiment can be represented as follows:

$$\begin{array}{llllllllllllll}
\Omega = \{12, & 13, & 14, & 15, & 16, & 17, & 18, & 21, & 23, & 24, & 25, & 26, & 27, & 28, \\
31, & 32, & 34, & 35, & 36, & 37, & 38, & 41, & 42, & 43, & 45, & 46, & 47, & 48, \\
51, & 52, & 53, & 54, & 56, & 57, & 58, & 61, & 62, & 63, & 64, & 65, & 67, & 68, \\
71, & 72, & 73, & 74, & 75, & 76, & 78, & 81, & 82, & 83, & 84, & 85, & 86, & 87\}
\end{array}$$

Let A be the event that the sum of the numbers is 6. Then

$$A = \{\underline{\hspace{6cm}}\},$$

and $P(A) =$ __________________________.

1.3 Counting Methods

1.3.1 Multiplication Rule, Sampling With/Without Replacement

Methods for counting the number of elements in a sample space or event are important in probability. The multiplication rule is the basic counting method.

Theorem (Multiplication Rule). If an operation consists of r steps of which

- the first can be done in n_1 ways,
- for each of these the second can be done in n_2 ways,
- for each of the first and second steps the third can be done in n_3 ways,
- and so forth,

then the entire operation can be done in $n_1 \times n_2 \times \cdots \times n_r$ ways.

Two important special cases of the multiplication rule are

1. ***Sampling With Replacement:*** For a set of size n and a sample of size r, there are a total of
$$n^r = n \times n \times \cdots \times n$$
ordered samples, if duplication is allowed.
2. ***Sampling Without Replacement:*** For a set of size n and a sample of size r, there are a total of
$$\frac{n!}{(n-r)!} = n \times (n-1) \times \cdots \times (n-r+1)$$
ordered samples, if duplication is not allowed.

Note that the symbol "$n!$" (read "n factorial") is defined as

$$n! = n(n-1)(n-2)\ldots 1 \text{ when } n \text{ is a positive integer and } 0! = 1.$$

Exercise. A menu in a restaurant reads like this:

1. Choice of one: chicken soup, tomato juice, fruit cocktail
2. Choice of one: beef hash, roast ham, fried chicken, spaghetti with meat balls
3. Choice of one: potatoes, brocolli, lima beans
4. Choice of one: chocolate ice cream, apple pie
5. Choice of one: coffee, tea, milk

(a) Find the number of complete dinners, where a complete dinner consists of one choice from each category.

(b) Suppose that you can't choose both spaghetti with meatballs and potatoes. How many complete dinners are there now?

(c) Suppose that the choice "none" is added to each category. Find the number of dinners with at least one item. Find the number of dinners with at least two items.

Exercise. The members of the executive committee of your club must decide among themselves who will serve as President, Vice President, Secretary and Treasurer; the remaining individuals serve as just committee members. An individual can hold at most one office.

How many different assignments are there

(a) if there are a total of 4 people on the executive committee?

(b) if there are a total of 8 people on the executive committee?

Exercise. Suppose that you are in charge of scheduling speakers for the student association. There are five dates. Ten people are interested in speaking:

Bob, Alice, Ted, Jim, Mary, Fred, Oscar, Beth, Charlie, Chris.

A schedule is a matching of five different people to the five dates.

(a) Find the number of schedules.

(b) Find the number of schedules if Bob can only come on one of the first three dates.

Example: The Birthday Problem. Suppose that there are r unrelated people in a room, none of whom was born on February 29th of a leap year. You would like to determine the probability that at least 2 people have the same birthday.

(i) You ask, and record, each person's birthday. There are 365^r possible outcomes, where an outcome is a sequences of r responses.

(ii) Consider the event "everyone has a different birthday." The number of outcomes in this event is $\frac{365!}{(365-r)!} = 365 \times 364 \times \cdots \times (365 - r + 1)$.

(iii) Suppose that each sequence of birthdays is equally likely. The probability that at least two people have a common birthday is 1 minus the probability that everyone has a different birthday, or

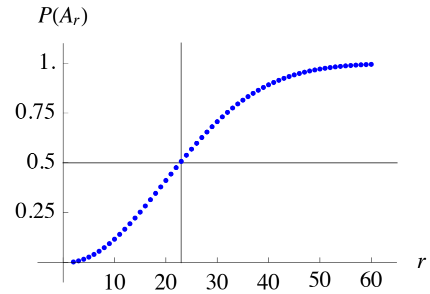
$$P(A_r) = 1 - \frac{365 \times 364 \times \cdots \times (365 - r + 1)}{365^r},$$

where A_r is the event that at least two people have a common birthday when there are r people in the room.

It is surprising how quickly the probability that at least two people have a common birthday grows under these simplistic assumptions.

For example,

- $P(A_r) > 0.50$ when $r \geq 23$.
- $P(A_r) > 0.99$ when $r \geq 57$.



Exercise. Let A_r be the event that at least one of r unrelated individuals in a room was born on November 7th. Using the same simplifying assumptions as above, find a general formula for $P(A_r)$. In addition, find the smallest value of r so that $P(A_r) \geq 0.50$.

1.3.2 Permutations, Combinations

- A *permutation* is an ordered subset of r distinct objects out of a set of n objects.
- A *combination* is an unordered subset of r distinct objects out of the n objects.

Counting rules. By the multiplication rule, there are a total of

$$_nP_r = \frac{n!}{(n-r)!} = n \times (n-1) \times \cdots \times (n-r+1)$$

permutations of r objects out of n objects.

Since each unordered subset corresponds to $r!$ ordered subsets (the r chosen elements are permuted in all possible ways), there are a total of

$$_nC_r = \frac{_nP_r}{r!} = \frac{n!}{(n-r)!\, r!} = \frac{n \times (n-1) \times \cdots \times (n-r+1)}{r \times (r-1) \times \cdots \times 1}$$

combinations of r objects out of n objects.

For example, there are ${}_6P_3 = 6{\times}5{\times}4 = 120$ ordered subsets of 3 elements from $\{1, 2, 3, 4, 5, 6\}$. The 120 ordered subsets can be represented as follows:

123	124	125	126	132	134	135	136	142	143	145	146	152	153	154
156	162	163	164	165	213	214	215	216	231	234	235	236	241	243
245	246	251	253	254	256	261	263	264	265	312	314	315	316	321
324	325	326	341	342	345	346	351	352	354	356	361	362	364	365
412	413	415	416	421	423	425	426	431	432	435	436	451	452	453
456	461	462	463	465	512	513	514	516	521	523	524	526	531	532
534	536	541	542	543	546	561	562	563	564	612	613	614	615	621
623	624	625	631	632	634	635	641	642	643	645	651	652	653	654

Each unordered subset corresponds to $3! = 6$ ordered subsets, implying that there are 20 unordered subsets: $\{1,2,3\}$, $\{1,2,4\}$, $\{1,2,5\}$, $\{1,2,6\}$, $\{1,3,4\}$, $\{1,3,5\}$, $\{1,3,6\}$, $\{1,4,5\}$, $\{1,4,6\}$, $\{1,5,6\}$, $\{2,3,4\}$, $\{2,3,5\}$, $\{2,3,6\}$, $\{2,4,5\}$, $\{2,4,6\}$, $\{2,5,6\}$, $\{3,4,5\}$, $\{3,4,6\}$, $\{3,5,6\}$, $\{4,5,6\}$.

Number of combinations. The notation $\binom{n}{r}$ (read "n choose r") is used to denote the total number of combinations. Special cases are as follows:

$$\binom{n}{0} = \binom{n}{n} = 1 \quad \text{and} \quad \binom{n}{1} = \binom{n}{n-1} = n.$$

Further, since choosing r elements to form a subset is equivalent to choosing the remaining $n - r$ elements to form the complementary subset,

$$\binom{n}{r} = \binom{n}{n-r} \quad \text{for } r = 0, 1, \ldots, n.$$

1.3.3 Binomial Coefficients, Binomial Theorem

The quantities $\binom{n}{r}$, $r = 0, 1, \ldots, n$, are often referred to as the *binomial coefficients*.

Theorem (Binomial Theorem). For all numbers x and y and each positive integer n,

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}.$$

Idea of the proof: The product on the left can be written as a sequence of n factors:

$$(x+y)^n = (x+y) \times (x+y) \times \cdots \times (x+y).$$

The product expands to 2^n summands, where each summand is a sequence of n letters (one from each factor). For each r, exactly $\binom{n}{r}$ sequences have r copies of x and $n - r$ copies of y.

When $n = 3$, for example,

$$(x+y)(x+y)(x+y) = xxx + xxy + xyx + xyy + yxx + yxy + yyx + yyy$$

has 1 term with zero x's, 3 terms with one x, 3 terms with two x's and 1 term with three x's.

Exercise. Use the binomial theorem to demonstrate that a set with n elements has 2^n subsets.

Exercise. You toss a fair coin n times and record h (for head) or t (for tail) each time. Let A_r be the event that there are exactly r heads in the sequence. Find a general formula for

$$P(A_r), \; r = 0, 1, \ldots, n.$$

In addition, construct a table of probabilities $P(A_r)$ for $r = 0, 1, 2, \ldots, 6$ when $n = 6$.

1.3.4 Simple Urn Model, Maximum Likelihood Estimation

Suppose there are M special objects in an urn containing a total of N objects. In a subset of size n chosen from the urn, exactly m are special.

(*i*) *Unordered Subsets:* There are a total of

$$\binom{M}{m} \times \binom{N-M}{n-m}$$

unordered subsets with exactly m special objects (and exactly $n-m$ ordinary objects). If each choice of subset is equally likely, then for each m

$$P(m \text{ special objects}) = \frac{\binom{M}{m} \times \binom{N-M}{n-m}}{\binom{N}{n}}.$$

(*ii*) *Ordered Subsets:* There are a total of

$$\binom{n}{m} \times {}_{M}P_{m} \times {}_{N-M}P_{n-m}$$

ordered subsets with exactly m special objects. (The positions of the special objects are selected first, followed by the special objects to fill these positions, followed by the ordinary objects to fill the remaining positions.) If each choice of subset is equally likely, then for each m

$$P(m \text{ special objects}) = \frac{\binom{n}{m} \times {}_{M}P_{m} \times {}_{N-M}P_{n-m}}{{}_{N}P_{n}}.$$

Interestingly, $P(m$ special objects) is the same in both cases. To see this, we write the first formula in terms of factorials, collect terms, and show that it is equivalent to the second formula (please complete):

Exercise. A shipment of 20 laptop computers contains 4 defective and 16 good machines. You select 3 machines at random and test them.

(a) How many subsets of 3 contain exactly k defective machines, for $k = 0, 1, 2, 3$?

(b) If each choice of subset is equally likely, find the probability that the subset you choose contains k defective machines, for $k = 0, 1, 2, 3$.

Estimating unknown quantities; maximum likelihood estimation. In statistics, information from a sample is used when complete information is impossible to obtain or would take too long to gather. One of the most important techniques is the method of *maximum likelihood*, developed by the British statistician R.A. Fisher.

An interesting application of Fisher's method uses probabilities derived from the simple urn model to estimate the total size of a population (N) based on information from two overlapping but incomplete sources:

1. *First source:* Information from a first main source is used to identify a subpopulation of interest, whose size is M.
2. *Second source:* Information from a second independent source is used to identify a sample from the population, whose size is n.
3. *Overlap:* Individuals identified using both sources form the overlap. Let m be the size of the overlap.

Let $Lik(N)$ (read "likelihood of N") be the probability of observing m individuals in the overlap if the true population size is N:

$$Lik(N) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}.$$

The likelihood function is maximized when $N = \left\lfloor M\left(\frac{n}{m}\right)\right\rfloor$.

Example (Regal & Hook, Biometrics *(1999), 55:1241-46).* Spina bifida is a rare spinal column defect that can be treated but not cured. Treatments for spina bifida include surgery, medication and physical therapy. This example considers estimating the number of babies born with spina bifida in upstate New York between 1969 and 1974, using information from birth and death certificates as the primary source of information on the disorder, and rehabilitation files as the secondary source.

The researchers identified 566 cases of spina bifida by examining the birth and death certificates of all live births in upstate New York between 1969 and 1974. Rehabilitation files included information on 188 cases, with an overlap of 128 cases. Since $566 - 128 = 438$ cases were identified from birth and death certificates only and $188 - 128 = 60$ cases were identified from rehabilitation files only, the sources give incomplete information about the total number of spina bifida cases. This information can be displayed as an incomplete 2-by-2 table:

128	438	566
60	$N-626$	$N-566$
188	$N-188$	N

$$Lik(N) = \frac{\binom{566}{128}\binom{N-566}{60}}{\binom{N}{188}}$$

1. *Left:* The rows in the table correspond to individuals who were identified or not identified by the primary identification method (birth and death certificates), and the columns in the table correspond to individuals who were identified or not identified by the secondary identification method (rehabilitation files).
2. *Right:* Individuals identified using the primary method are the *special* individuals in the urn, while those identified using the secondary method form the *sample* taken from the urn. The likelihood function is shown on the right above.

If we knew N, then we could compute the probability of getting exactly 128 special and 60 ordinary individuals in the sample of size 188. Since we don't know N, we will try different values consistent with the table. Here are three choices, including their probabilities:

128	438	566
60	158	218
188	596	784

$Lik(784) = 0.026$

128	438	566
60	176	236
188	614	802

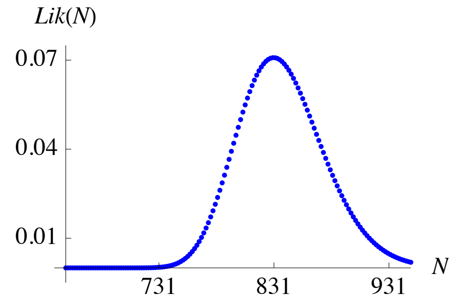
$Lik(802) = 0.050$

128	438	566
60	240	300
188	678	866

$Lik(866) = 0.047$

The plot on the right shows $Lik(N)$ for values of N between 650 and 950. The function is maximized when

$$N = \left\lfloor 566\left(\frac{188}{128}\right)\right\rfloor = 831.$$



Capture-recapture method. Using the simple urn model setup to estimate an unknown total population size is often referred to as the *capture-recapture method*. The method's name comes from its use in the field of ecology.

For this type of problem, we can show that the likelihood function increases up to a point (the point of maximum likelihood), and decreases afterwards. In the spina bifida example, the unique point of maximum likelihood occurs when $N = \lfloor 566(188/128)\rfloor = 831$. Although the number of different urns is unlimited (each $N \geq 626$ defines a possible urn), the urn model that is most consistent with our data is the one in which N equals 831.

The simple assumptions used were that the two methods of identifying individuals were independent of each other, and that each choice of subset was equally likely.

Likelihood ratio method. Determining a single number best guess is just a first step in estimation; a range of plausible values for an unknown quantity is much more useful.

The *likelihood ratio method* is a commonly used method for finding a range of plausible values. To construct the range, we consider all models where the ratio of the likelihood function to its maximum possible value is at least a fixed proportion p.

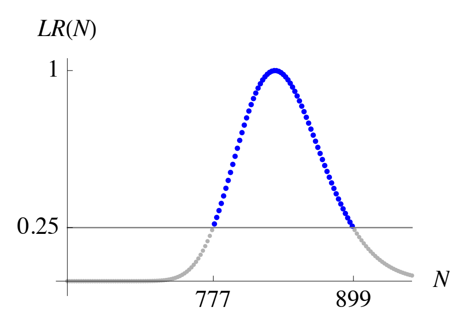
A commonly used proportion is $p = 0.25$. To illustrate the technique using the spina bifida example,

let

$$LR(N) = \frac{Lik(N)}{Lik(831)}$$

be the likelihood ratio function. (Remember that the likelihood was maximized at 831.)

Now, $LR(N) \geq 0.25$ corresponds to values of N in the range $\{777, 778, \ldots, 899\}$.



Thus, we would report the range of integers from 777 to 899 as the range of plausible values for the number of babies born with spina bifida in upstate New York during the 5 years of the study.

1.3.5 Partitioning Sets, Multinomial Coefficients, Multinomial Theorem

The multiplication rule can be used to find the number of *partitions* of a set of n elements into k distinguishable subsets of sizes r_1, r_2, …, r_k.

Theorem (Number of Partitions). The number of ways that n objects can be grouped into k distinguishable subsets of sizes r_1, r_2, …, r_k is

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1!\; r_2!\; \cdots\; r_k!},$$

where the symbol on the left is read "n choose $r_1, r_2, \ldots, r_k$".

Proof: The choices are made in k steps: r_1 of the n elements are chosen for the first subset, r_2 of the remaining $n - r_1$ elements are chosen for the second subset, and so forth. The result is the product of the numbers of ways to perform each step:

$$\binom{n}{r_1} \times \binom{n-r_1}{r_2} \times \cdots \times \binom{n - r_1 - \cdots - r_{k-1}}{r_k} = \frac{n!}{r_1!\; r_2!\; \cdots\; r_k!}$$

after simplification.

Exercise. A class of 15 students is to be split into 3 recitation sections of 5 students each, led by Sally, Mary and Joe, respectively. The recitation sections are distinguished by their group leaders, who are not members of the class. The total number of ways in which this can be done is

$$\binom{15}{5,5,5} = \frac{15!}{5!\; 5!\; 5!} = 756,756.$$

Recalculate this total using the multiplication rule as described in the proof above.

Permutations of indistinguishable objects. The formula above also represents the number of ways to permute n objects, where the first r_1 are indistinguishable, the next r_2 are indistinguishable, ..., the last r_k are indistinguishable. The computation is done as follows: r_1 of the n positions are chosen for the first type of object, r_2 of the remaining $n - r_1$ positions are chosen for the second type of object, and so forth.

Exercise. Find the number of permutations of the eleven letters in the word *MISSISSIPPI*. How many permutations are there whose first and last letters are "*I*"?

.

Multinomial coefficients. The quantities $\binom{n}{r_1, r_2, \ldots, r_k}$ are often referred to as the *multinomial coefficients* because of the following theorem.

Theorem (Multinomial Theorem). For all numbers $x_1, x_2, \ldots, x_k$ and each positive integer n,

$$(x_1 + x_2 + \cdots + x_k)^n \;=\; \sum_{(r_1, r_2, \ldots, r_k)} \binom{n}{r_1, r_2, \ldots, r_k} x_1^{r_1} x_2^{r_2} \cdots x_k^{r_k},$$

where the sum is over all k-tuples of non-negative integers with $\sum_i r_i = n$.

Idea of proof: The product on the left can be written as a sequence of n factors

$$(x_1 + x_2 + \cdots + x_k)^n = (x_1 + x_2 + \cdots + x_k) \times$$

$$(x_1 + x_2 + \cdots + x_k) \times \cdots \times (x_1 + x_2 + \cdots + x_k).$$

The product expands to k^n summands, where each summand is a sequence of n letters (one from each factor). For each $r_1, r_2, \ldots, r_k$, exactly $\binom{n}{r_1, r_2, \ldots, r_k}$ sequences have r_1 copies of x_1, r_2 copies of x_2, etc.

1.4 Conditional Probability

Assume that A and B are events, and that $P(B) > 0$. Then the *conditional probability* of A given B is defined as follows:

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

Notes:

- Event B is often referred to as the *conditional sample space*.
- The conditional probability $P(A|B)$ is the relative "size" of A within B.
- If Ω is a finite set and each outcome is equally likely, then $P(A|B) = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{|A \cap B|}{|B|}$.

Exercise. 1000 adults are classified by two criteria: (1) whether the person smokes (S) or not, and (2) whether the person has a respiratory disease (R) or not, as shown on the right.

	R	R^c	*Sum:*
S	45	355	400
S^c	5	595	600
Sum:	50	950	1000

The name of one person is chosen at random. If each choice of individual is equally likely, then

$P(S) =$ ____________________ $P(R) =$ ____________________

$P(S|R) =$ ____________________ $P(R|S) =$ ____________________

$P(S|R^c) =$ ____________________ $P(R|S^c) =$ ____________________

Exercise. Answer the questions on the right based on the following table of intersection probabilities ($P(A \cap B)$, $P(A \cap B^c)$, $P(A^c \cap B)$, $P(A^c \cap B^c)$), and their sums:

	B	B^c	*Sum:*
A	0.10	0.40	0.50
A^c	0.20	0.30	0.50
Sum:	0.30	0.70	1.00

$P(A|B) =$ ____________________

$P(B|A) =$ ____________________

$P(A|A \cup B) =$ ____________________

$P(A|A \cap B) =$ ____________________

$P(A \cap B|A \cup B) =$ ____________________

1.4.1 Multiplication Rule For Probability

Assume that A and B are events with positive probability. Then the definition of conditional probability implies that the probability of the intersection, $A \cap B$, can be written as a product of probabilities in two different ways:

$$P(A \cap B) = P(B) \times P(A|B) = P(A) \times P(B|A).$$

More generally,

Theorem (Multiplication Rule for Probability). If $A_1, A_2, \ldots, A_k$ are events and $P(A_1 \cap A_2 \cap \cdots \cap A_{k-1}) > 0$, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) =$$
$$P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 \cap A_2) \times \cdots \times P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}).$$

Exercise. Fifteen percent of the people in a certain adult population have respiratory disease.

- Among those with respiratory disease, 45% are current smokers, 35% are past smokers, and 20% have never smoked cigarettes.
- Among those with no respiratory disease, 20% are current smokers, 30% are past smokers, and 50% have never smoked cigarettes.

A person's name is chosen. If each choice of individual is equally likely, use the information above to complete the following table of intersection probabilities and their sums:

	Current Smoker	Past Smoker	Never Smoked	*Sum:*
Respiratory Disease				
No Respiratory Disease				
Sum:				

Exercise. On a first try, a sharpshooter can hit the bull's eye 90% of the time. Given a hit, the probability that she hits on the second try is 0.92. Given a miss, the probability that she hits on the second try is 0.88. Fill in probabilities for each of the following events:

The sharpshooter hits the target on both the first and second tries.	
The sharpshooter hits the target on the first try but not on the second try.	
The sharpshooter hits the target on the second try but not on the first try.	
The sharpshooter misses the target on both the first and second tries.	

Exercise. Four slips of paper are sampled without replacement from a well-mixed urn containing twenty-five slips of paper: fifteen slips with the letter X written on each and ten slips of paper with the letter Y written on each. Then

$$P(XYXX) = P(X) \times P(Y|X) \times P(X|XY) \times P(X|XYX) = \frac{15}{25} \times \frac{10}{24} \times \frac{14}{23} \times \frac{13}{22} = \frac{91}{1012} \approx 0.0899.$$

Fill in probabilities for each of the following events:

The sequence $XXXY$ is chosen	
Exactly 1 Y and 3 X's are chosen	
The sequence $XXYY$ is chosen	
Exactly 2 Y's and 2 X's are chosen	

Fill in probabilities for the following events if sampling is done with replacement:

Exactly 1 Y and 3 X's are chosen	
Exactly 2 Y's and 2 X's are chosen	

1.4.2 Law of Total Probability, Law of Average Conditional Probability

The law of total probability can be used to write an unconditional probability as the weighted average of conditional probabilities. Specifically,

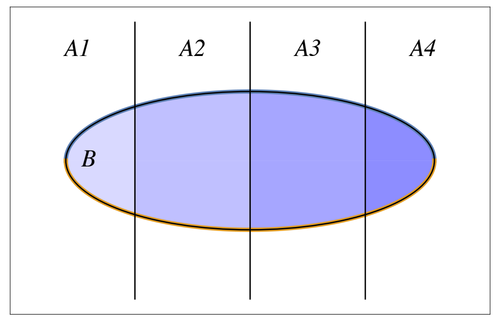
Theorem (Law of Total Probability). Let A_1, A_2, $\ldots$, A_k and B be events with non-zero probability. If A_1, A_2, $\ldots$, A_k are pairwise disjoint with union Ω, then

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_k)P(B|A_k).$$

Proof: If A_1, A_2, $\ldots$, A_k are pairwise disjoint with union Ω, then the distributive law implies that the sets

$$B \cap A_1,\ B \cap A_2,\ \ldots,\ B \cap A_k$$

are pairwise disjoint with union B, as illustrated on the right using $k = 4$ sets to partition Ω.



Thus, the disjoint union rule (axiom 3) and the definition of conditional probability imply that

$$P(B) = \sum_{i=1}^{k} P(B \cap A_i) = \sum_{i=1}^{k} P(A_i)P(B|A_i).$$

Exercise. Fifteen percent of the people in a certain adult population have respiratory disease. Among those with respiratory disease, 45% are current smokers. Among those with no respiratory disease, 20% are current smokers. Find the probability that a person in this population is a current smoker.

Exercise. A city is comprised of four different ethnic groups: 35% of the population is of ethnic group 1, 25% of ethnic group 2, and 20% each of ethnic groups 3 and 4. 30%, 50%, 40% and 45% of ethnic groups 1, 2, 3, and 4, respectively, have blood type O. Find the probability that a person living in the city has blood type O.

Law of average conditional probability. The law of total probability is often called the *law of average conditional probabilities*: $P(B)$ is the weighted average of the conditional probabilities

$$P(B|A_1),\ P(B|A_2),\ \ldots,\ P(B|A_k),$$

where $P(A_1)$, $P(A_2)$, ..., $P(A_k)$ are the weights in the weighted average.

1.4.3 Bayes Rule, Prior and Posterior Probabilities, Diagnostic Tests

Bayes rule, proven by the Reverend T. Bayes in the 1760's, can be used to update probabilities given that an event has occurred. Specifically,

Theorem (Bayes Rule). Let A_1, A_2, ..., A_k and B be events with non-zero probability. If A_1, A_2, ..., A_k are pairwise disjoint with union Ω, then

$$P(A_j|B) \;=\; \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)}, \quad j = 1, 2, \ldots, k.$$

Proof: The numerator of the expression on the right is $P(A_j \cap B)$, and the denominator is $P(B)$ by the law of total probability. Thus, the ratio simplifies to $P(A_j|B)$.

Exercise. A city is comprised of four different ethnic groups: 35% of the population is of ethnic group 1, 25% of ethnic group 2, and 20% each of ethnic groups 3 and 4. 25%, 35%, 35% and 30% of ethnic groups 1, 2, 3, and 4, respectively, have blood type B. If a person has blood type B, find the probabilities that he or she is in ethnic group j, for $j = 1, 2, 3, 4$.

Prior and posterior events. In applications,

- The collection of probabilities $\{P(A_j)\}$ are often referred to as the *prior* probabilities (the probabilities *before* observing an outcome in B), and
- The collection of probabilities $\{P(A_j|B)\}$ are often referred to as the *posterior* probabilities (the probabilities *after* event B has occurred).

Diagnostic screening tests. In diagnostic screening, a test is applied to an individual who has not yet exhibited clinical symptoms of a particular disease with the goal of determining that individual's probability of having the disease. Those who test positive are considered more likely to have the disease and may be recommended to take further tests or to start treatment.

Let D be the event that an individual has the disease, D^c be the event that the individual is disease free, POS be the event the individual tests positive for disease and NEG be the event the individual tests negative for disease. Then

1. ***Sensitivity:*** The *sensitivity* (or *true-positive rate*) of the test is the probability that an individual chosen at random from those with disease tests positive for disease,

   $$\text{Sensitivity} = P(POS|D).$$

2. ***Specificity:*** The *specificity* (or *true-negative rate*) of the test is the probability that an individual chosen at random from those who are disease free tests negative for disease,

   $$\text{Specificity} = P(NEG|D^c).$$

3. ***Prevalence:*** The *prevalence* of the disease is the probability that an individual chosen at random from the study population has the disease,

   $$\text{Prevalence} = P(D).$$

4. ***Predictive Value of Positive Test:*** The *predictive value of a positive test* (or the *positive predictive value of the test*) is the probability that an individual chosen at random from those testing positive for disease actually has the disease,

   $$\text{Positive Predictive Value} = P(D|POS).$$

5. ***Predictive Value of Negative Test:*** The *predictive value of a negative test* (or the *negative predictive value of the test*) is the probability that an individual chosen at random from those testing negative for disease is actually disease free,

   $$\text{Negative Predictive Value} = P(D^c|NEG).$$

Example (Carpal Tunnel Syndrome; Pagano & Gauvreau, 2000, page 157). Carpal tunnel syndrome is an affliction of the wrist, which occurs when the median nerve (a nerve running from the forearm into the hand) becomes pressed or squeezed. The risk of developing carpal tunnel syndrome is not confined to individuals in a single industry, but is especially common in those performing assembly line work.

The National Institute for Occupational Safety and Health (NIOSH) has established a definition of this disorder that incorporates three criteria: symptoms of nerve damage, a history of occupational risk factors, and the presence of physical exam findings. The sensitivity of this definition as a test for the syndrome is 0.67, and the specificity is 0.58. That is,

$$P(POS|D) = 0.67 \text{ and } P(NEG|D^c) = 0.58,$$

where D is the event that a worker has carpal tunnel syndrome, D^c is the event that a worker does not have the disease, POS is the event that the worker satisfies all three criteria established by NIOSH and NEG is the event that the worker does not satisfy all three criteria.

Suppose that 20% of workers in a particular industry have carpal tunnel syndrome. Then the following table summarizes the calculations needed to find the predictive value of a positive test and of a negative test.

Analysis of Carpal Tunnel Syndrome Test
When 20% of Workers Have the Disease

	POS	NEG	Total
D	0.134	0.066	0.20
D^c	0.336	0.464	0.80
Total	0.470	0.530	1.00

Predictive Value of
Positive Test is 0.2851.

Predictive Value of
Negative Test is 0.8755.

(*Question:* Can you see where I got these numbers?)

1.5 Independent Events

Events A and B are said to be *independent* (or *probabilistically independent*) if

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

Otherwise, A and B are said to be *dependent*.

Theorem (Independent Events). Let A and B be events whose probabilities satisfy the following inequalities: $0 < P(A) < 1$ and $0 < P(B) < 1$. Then

$$A \text{ and } B \text{ are independent} \;\Leftrightarrow\; P(A|B) = P(A) \;\Leftrightarrow\; P(B|A) = P(B).$$

Exercise. Demonstrate the first equivalence in the theorem above.

Exercise. One chip is drawn from each of two well-mixed urns:

- *Urn 1:* The first urn contains 5 red, 4 blue and 6 green chips.
- *Urn 2:* The second urn contains 7 red, 3 blue and 10 green chips.

Find the probability that both chips are the same color.

1.5.1 Mutually Independent Events

Events $A_1, A_2, \ldots, A_k$ are said to be *mutually independent* if

- For each pair of distinct indices (i_1, i_2): $P(A_{i_1} \cap A_{i_2}) = P(A_{i_1}) \times P(A_{i_2})$,
- For each triple of distinct indices (i_1, i_2, i_3):

$$P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) = P(A_{i_1}) \times P(A_{i_2}) \times P(A_{i_3}),$$

- and so forth.

Exercise. Two fair six-sided dice (one red and one green) are rolled, and the number on the top face of each die is recorded. The sample space can be represented as follows:

$$\begin{aligned} \Omega = \ & \{11, 12, 13, 14, 15, 16, 21, 22, 23, 24, 25, 26, 31, 32, 33, 34, 35, 36, \\ & 41, 42, 43, 44, 45, 46, 51, 52, 53, 54, 55, 56, 61, 62, 63, 64, 65, 66\} \end{aligned}$$

Let A be the event that the red die shows a 3, 4, or 5; B be the event that the green die shows a 1 or 2; and C be the event that the dice total is 7.

Are A, B and C mutually independent? Why?

1.5.2 Repeated Trials and Mutual Independence

As stated at the beginning of this notebook, the term experiment is used in probability theory to describe a procedure whose outcome is not known in advance with certainty. Experiments are assumed to be repeatable, and to have a well-defined set of outcomes. Repeated trials are repetitions of an experiment using the specified procedure, with the outcomes of the trials having no influence on one another. The results of repeated trials of an experiment are mutually independent.

Exercise. The following procedure is repeated six times: thoroughly shuffle a standard deck of 52 cards, choose a card, note its suit (C, D, H, S), replace the card.

Find the probability of obtaining

- (a) The sequence $CDHDSS$.
- (b) Exactly two diamonds and two spades.
- (c) Exactly two diamonds, two spades, and two hearts.