These are rough lecture notes from Math 210 in Fall 2006. It is far from polished or complete. This section mainly covers the material on probability covered in September 2006. All items concerning the computer are missing. Nonetheless, I hope they are helpful.

September 7, 2006

Introduction to Math 210: Mathematics in the Age of Information

Class Website: www.math.upenn.edu/kazdan/210

Course Overview

2. "Day of the Week" Problem
3. Probability
4. Dynamical Systems
5. Secure Encryption
6. Ranking of Sports teams
7. Voting

Introduction to Simple Webpage Design

You can look to SAS, SEAS, WHARTON for help. Note: you must log onto the computer and secure CRT is required.

Security and Permissions: When you have created a webpage, you may want to control who can view and modify it. There are 3 classes of users:

- user (owner)
- group (students)
- others (everybody else)

You can allow different levels of access for each class. There are three levels access:

- Read (R)
- Write (W)
- Execute (E)

To specify the level of access, we use base 2, where R = 4, W = 2 and E = 1. For example, 110 = 6 and 6 means that the user can Read and Write, but not Execute.

Day of the Week Problem
Given the month, day, and year, find what day of the week it is.
Say today is Thursday, September 7, 2006. What day of the week will March 27, 2041 be?

First a simpler question. What day of the week will it be 700 days from now? Since today is
a Thursday and 700 is divisible by 7, 700 days from now will also be a Thursday. Similarly,
703 days from now will be a Sunday. Thus, we see that to find the day of the week, we

1. Pick a reference date, such as Sunday, Jan. 1, 2006.
2. Count the number $N$ of days between the reference date and the specified date.
3. Compute the remainder of dividing $N$ by 7.

Since $365 = 52 \times 7 + 1$, then September 7, 2007 will be a Friday. Leap years have 366
days. The rule for leap years is:

Every fourth year is a leap year, except for multiples of 100, which are not leap years,
unless they are also multiples of 400.
September 12, 2006

Probability

EXAMPLE Say you roll two dice. What is the probability $P$ of rolling:

2 fours? Answer: $\frac{1}{36} \approx .028$

1 two and 1 six? Answer: $P = \frac{2}{36} = \frac{1}{18} \approx .055$

At least 1 is a six, that is, either the first die is a six or the second is a six or both are sixes?

This last question is easier approached by asking ”What is the probability $Q$ that neither is a six?”. In other words, the first die is a not six and the second is not a six?”

Answer: $Q = \frac{5}{6} \times \frac{5}{6} = \frac{25}{36} \approx .694$

To answer the original question, we observe that ”neither is a six” is the complementary event to ”rolling at least 1 six”. Thus the sum of these probabilities is $P + Q = 1$, we compute

$$P = 1 - Q = \frac{25}{36} = \frac{11}{36} \approx .306.$$ 

Remark: Computing probabilities involving and are easier than those involving or.

BIRTHDAY PROBLEM

1. What is the probability that 2 people have the same birthday – that is, they were born on the same day of the year although possibly in different years? Answer: $\frac{1}{365} \approx .00274$.

2. If there are 3 people, what is the probability $P$ that at least 2 of them have the same birthday? [Ignore leap years].

As in an earlier example, this question involves or and is easier to approach if we ask the complementary question: What is the probability $Q$ that nobody shares the same birthday? This version is an and question.

Answer: $Q = \frac{364}{365} \times \frac{363}{365} \approx .9918$. Here $\frac{364}{365}$ is the probability that person 2 does not have the same birthday as person 1 while $\frac{363}{365}$ is the probability that person 3 has a birthday different from both persons 1 and 2. Now, to solve the original question, we simply compute

$$P = 1 - Q = 1 - \left( \frac{364}{365} \right) \left( \frac{363}{365} \right) \approx .00820.$$ 

3. If there are 23 people, what is the probability that at least 2 have the same birthday?
Answer: Based on the approach used to solve part (2), we realize that there will be 22 terms to subtract from 1. That is, the probability is

\[ P = 1 - \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \cdots \left(\frac{343}{365}\right) \approx .5073. \]

Note that I used a perl script to compute .5073, but I could have used Maple or a scientific calculator.

**EXAMPLE: Genetic Relationship Problem**

To whom are you more related, your mom or your brother?

Answer: You are equally related to your mom as you are to your brother. You share 50% of your DNA with your mom, and 50% of your DNA with your brother.

**BUS PROBLEM**

Two buses run every 10 minutes on the same road. Their starting times at the beginning of the day is random. What is the average waiting time?

We can do a computer simulation to check our theory: Using a random number generator (such as in perl) we can determine the average waiting time. See the class website for http://www.math.upenn.edu/~kazdan/210/computer/perl/one-bus-simulation.pl

For two buses with different waiting times, you simply take the bus that comes first. For example, the perl code:

```perl
$bus1 = rand(10)
$bus2 = rand(10)
$min = $bus1
if ($bus2 < $bus1) then {
$min = $bus2;
}
```

**THEORY BEHIND THE BUS PROBLEM**

Waiting time \( T \) is a “random variable”:

\[ F(t) = \text{Prob} (T < t) = \text{(probability bus arrives before time } t) \]

so \( F(t) \) is the probability that the waiting time will be less than \( t \). \( F(t) \) is called the distribution function. It is an increasing function with the additional properties

\[ F(-\infty) = 0 \quad \text{and} \quad F(+\infty) = 1. \]

In this problem, \( F(t) = 0 \) for all \( t \leq 0 \), while after 10 minutes, \( F(10) = 1 \). Consequently \( F(t) = 1 \) for all \( t \geq 10 \).

Clearly

\[ \text{Prob} (a < T < b) = F(b) - F(a). \]
It is useful to introduce the \textit{density function} \( f(t) \). It has the property that

\[
F(t) = \int_{-\infty}^{t} f(s) \, ds. \tag{1}
\]

Using this, the probability that the bus arrives between time \( a \) and \( b \) is then

\[
\text{Prob} \ (a < T < b) = \int_{a}^{b} f(s) \, ds.
\]

How can we compute the density \( f(t) \) from the distribution function \( F(t) \)? This is straightforward if we apply the fundamental theorem of calculus to (1) and deduce that

\[
f(t) = \frac{dF(t)}{dt}.
\]
There are 2 types of Probability:

- Discrete, as tossing a coin or dice.
- Continuous, as the waiting time for a bus.

Say we have a set $S$, our sample space, with subsets $A$ and $B$. It is useful to think of these sets as regions in the plane, $R^2$. Let $P(A)$ be the probability that a point of $S$ is in the subset $A$. It should be plausible that

$$P(A) = \frac{\text{Area}(A)}{\text{Area}(S)}.$$ 

Probability always satisfies:

$$0 \leq P(A) \leq 1, \quad P(\emptyset) = 0, \quad \text{and} \quad P(S) = 1.$$ 

Here $\emptyset$ is the empty set.

**A few remarks about sets**

$A^c$ = complement of $A = S - A$

$(A \cup B)^c = A^c \cap B^c$

$(A \cap B)^c = A^c \cup B^c$

If $A$ and $B$ are mutually exclusive (so $A \cap B = \emptyset$), then

$$P(A \cup B) = P(A) + P(B)$$

Also,

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

More generally:

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

so $P(A \cup B) \leq P(A) + P(B)$. If $B \subseteq A$, then $A \cap B = B$.

**Example** Spinal fluid test for TB and bacteria: 80% of the population test negative for both TB and bacteria. What is the probability that a random sample tests positive for at least one of either TB or bacteria?

**Answer:** To compute $P(A \cup B) = P(A \text{ or } B)$, we see that this is an or question, so we look to the opposite and situation: What is the probability that a random sample tests negative for both TB and bacteria? So now, we see that:

$$P(A \cup B) = 1 - P(A^c \cap B^c) = 1 - 0.80 = 0.20$$
**EXPECTATION AND STANDARD DEVIATION**

Question: Why is the average (expectation) good?

\[ E(c) = \sqrt{(a_1 - c)^2 + (a_2 - c)^2 + \cdots + (a_n - c)^2} \]

How do we pick \( c \) to minimize \( E \)? To get rid of the square root, we let \( Q(A) = E^2(a) \). Now take the derivative of \( Q \):

\[
0 = \frac{dQ}{dc} = 2[(a_1 - c) + (a_2 - c) + \cdots + (a_n - c)] = [a_1 + a_2 + \cdots + a_n] - nc
\]

so

\[ c = \frac{a_1 + \cdots + a_n}{n} \]

Thus, we find that the average minimizes error.

The standard deviation tells how much the data is dispersed.

**Example:** Referring back to the Bus Problem: \( F(t) = \text{Prob}(T < t) \)

- \( F(t) = 0 \) when \(-\infty < t \leq 0\)
- \( F(t) = 1 \) when \( t \geq 10 \)

\[
F(t) = \int_{-\infty}^{t} f(s) \, ds,
\]

\( f(s) \, ds \) is the density, with \( \frac{dF}{dt} = f(t) \).

What is \( F(t) \)? We use that the probability of a bus coming in a time interval, \( 0 \leq t \leq 10 \), depends only on the length of this interval, not on its endpoints:

\[
P(t_0 \leq T \leq t_0 + h) = P(t \leq T \leq t + h),
\]

so

\[
\frac{F(t_0 + h) - F(t_0)}{h} = \frac{F(t + h) - F(t)}{h}
\]

(here we are assuming that \( t_0, t_0 + h, t, \) and \( t + h \) all lie in the interval \( 0 \leq t \leq 10 \)). To use this, divide both sides by \( h \) and let \( h \to 0 \). By the definition of the derivative, this gives

\[
F'(t_0) = F'(t) = f(t),
\]

so the density \( f(t) \) is constant, say \( f(t) = a \). Integrating we find \( F(t) = at + b \). Now we use that \( F(0) = 0 \) and \( F(10) = 1 \) to conclude that \( F(t) = \frac{t}{10} \) for \( 0 \leq t \leq 10 \) and \( f(t) = 1/10 \).

One Bus:

Expected value \( E(t) = \int_{-\infty}^{\infty} sf(s) \, ds = \int_{0}^{10} s \frac{1}{10} \, ds = \left. \frac{s^2}{20} \right|_{0}^{10} = 5 \).
Two Buses: say $T_1$ is the waiting time for bus #1, $T_2$ for bus #2, and $T$ for the actual waiting time. Thus there are three random variables, $T_1$, $T_2$, and $T$. Clearly $T$ is either $T_1$ or $T_2$, whichever is smaller. From our solution of the one bus problem, we know both $P(T_1 < t)$ and $P(T_2 < t)$. We want $F(t) = P(T < t)$. Since this is an or problem, we recast it as an and problem. Now $T > t$ means that both buses come after $t$ minutes: $T_1 > t$ and $T_2 > t$. But for $0 \leq t \leq 10$ the probability

$$P(T_1 \geq t) = 1 - P(T_1 < t) = 1 - \frac{t}{10}$$

Because the arrival times of bus #1 and bus #2 are independent events, we multiply their probabilities:

$$P(T \geq t) = [1 - P(T_1 \geq 1)][1 - P(T_2 \geq t)] = \left(1 - \frac{t}{10}\right)^2 = 1 - \frac{2}{10}t + \frac{1}{100}t^2.$$ 

Consequently,

$$F(t) = P(T < t) = 1 - P(T \geq t) = \frac{1}{100}(20t - t^2).$$

Using this we find the corresponding density function

$$f(t) = F'(t) = \frac{2}{10}(1 - \frac{t}{10})$$

and can compute the expected value:

$$\text{Expected value } E(T) = \int_0^{10} s\left(\frac{2}{10}\right)(1 - \frac{s}{10}) ds = \frac{10}{3}.$$ 

**Method 2:** Here is another method of finding the distribution function $F(t) = P(T < t)$ for the two bus problem. The only interesting times are $0 \leq t \leq 10$. We are picking a random pair of points $(T_1, T_2)$ in the square $0 \leq T_1 \leq 10$, $0 \leq T_2 \leq 10$ and want the probability that either $0 < T_1 < t$ or $0 < T_2 < t$, that is, the point $(T_1, T_2)$ is in one of the gray rectangles at the bottom or left side. The probability is the area of these rectangles divided by the area of the large square. The areas of these rectangles can be computed in several ways. One is to use the area of the big box and subtract the area $(10-t)^2$ of the white square in the upper right corner. Using this, we find

$$F(t) = P(T < t) = \frac{100 - (10-t)^2}{100} = \frac{1}{100}(20t-t^2).$$
just as before. Now use the previous calculation to compute the expected value.

**METHOD 3:** This computes the expected value of the waiting time without first finding the density function. Realize that there are 3 independent random numbers:

- time bus #1 comes,
- time bus #2 comes,
- time you come.

By “symmetry”, we expect that these three events will be equally spaced in the interval $0 \leq t \leq 10$ and therefore are separated by times $\frac{10}{3}$ minutes.

**Monte Carlo Method**

You have an area of a set $S$. You pick lots of random points in the box. See how many $\alpha\%$ are in $S$. Then, area $S \approx \alpha \times$ (area of box).

*Example:* Estimate area of box: $x^2 + y^2 < 1$. Pick random $(x, y)$. Is $x^2 + y^2 < 1$? If so, increase $Q = Q + 1$. Do this $N$-times. Then:

\[
\text{Area of big box} = 4 \text{ and } 4\left(\frac{Q}{N}\right) \approx \text{area of circle}.
\]

**Simple Example with Discrete Probability**

1 die: 1, 2, ..., 6

Probability of getting each of these numbers is $\frac{1}{6}$.

Expected value: $(1)(\frac{1}{6}) + (2)(\frac{1}{6}) + \cdots + (6)(\frac{1}{6}) = 3\frac{1}{2}$

Variance: Standard Deviation $= \sqrt{\text{variance}}$
Expected Value and Standard Deviation

Example 1: For a standard die, the probability that you get one of the numbers 1, 2, 3, 4, 5, or 6 is $\frac{1}{6}$. Let the random variable $X$ be the number that appears when you toss the die.

\[
\text{expected value } E(X) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{21}{6} = \frac{7}{2}
\]

\[
V(X) = \text{Variance} = E[(X - E(X))^2] = \frac{1}{6}[(1 - \frac{7}{2})^2 + (2 - \frac{7}{2})^2 + \cdots + (6 - \frac{7}{2})^2] = \frac{35}{12} \approx 2.9
\]

\[
SD(X) = \text{Standard Deviation} = \sqrt{V(X)} \approx 1.7
\]

Example 2: You have a box with the following numbers: 0, 2, 3, 4, 6. It is equally likely that you can get any one of those numbers. In other words, the probability of getting any one of those numbers is $\frac{1}{5}$.

\[
E(X) = 0 \times \frac{1}{5} + 2 \times \frac{1}{5} + \cdots + 6 \times \frac{1}{5} = \frac{15}{5} = 3
\]

\[
V(X) = \frac{1}{5}[(0 - 3)^2 + (2 - 3)^2 + (3 - 3)^2 + (4 - 3)^2 + (6 - 3)^2] = \frac{1}{5}[9 + 1 + 0 + 1 + 9] = 4
\]

\[
SD(X) = \sqrt{V(X)} = \sqrt{4} = 2
\]

General Properties

Assume that $X, Y$ are independent random variables.

1. $E(X + c) = E(X) + c$
2. $E(aX) = aE(X)$
3. $E(X + Y) = E(X) + E(Y)$

The variance of a random variable $X$, $V(X)$, is the measure of how spread out things are. The definition is: $V(X) = E[(X - E(X))^2]$. Since $|X - E(X)|^2 = X^2 - 2E(X)X + E(X)^2$, then

\[
V(X) = E(X^2) - 2E(X)E(X) + E(X)^2 = E(X^2) - (E(X))^2
\]

Note that $V(X + c) = Var(X)$ because $(X + c) - E(X + c) = X - E(X)$ so the constant, $c$, is cancelled. It follows that standard deviation $SD(X + c) = SD(X)$.

4. $V(aX) = a^2V(X)$. This should be evident.
5. $SD(aX) = |a|SD(X)$
6. $V(X + Y) = E(X + Y)^2 - [E(X) + E(Y)]^2$. For this it is critical that $X$ and $Y$ be independent random variables.

$$= V(X) + V(Y) - [E(X)^2 + 2E(X)E(Y) + E(Y)^2]$$

Since $X$ and $Y$ are independent, then $E(XY) = E(X)E(Y) = 0$. So in this case, $V(X + Y) = V(X) + V(Y)$ and $SD(X + Y) = \sqrt{SD(X)^2 + SD(Y)^2}$

**Example.** You toss two standard dice. What is the expected value of the sum of points you get?

Let the random variable $X_1$ be the number of points on the first die and $X_2$ on the second die. Since $E(X_1) = E(X_2) = \frac{7}{2}$, then $E(X_1 + X_2) = E(X_1) + E(X_2) = 2(\frac{7}{2}) = 7$.

**Example 3:** You have a box with $n$ - entries of $a$ and $k$ - entries of $b$.

1. What is the probability of drawing an $a$?
   
   Answer: $P(X = a) = \frac{n}{n+k}$ Let’s call this value $p$.

1a. What is the probability of drawing a $b$?
   
   Answer: $P(X = a) = \frac{n}{n+k}$ Equivalently, it is simply $1 - p = 1 - \frac{n}{n+k}$.

2. If $a$ and $b$ are numbers, what is the expected value?
   
   Answer: $E(x) = a\frac{n}{n+k} + b\frac{k}{n+k} = \frac{na+kb}{n+k}$

3. What is the variance?
   
   Answer: $V(X) = \left[a - \left(\frac{na+kb}{n+k}\right)\right]^2 \frac{n}{(n+k)} + \left[b - \left(\frac{na+kb}{n+k}\right)\right]^2 \frac{k}{(n+k)} = (b-a)^2 \frac{nk}{(n+k)^2} = (b-a)^2 p(1-p)$

4. What is the standard deviation?
   
   Answer: $SD(x) = \sqrt{V(x)} = \sqrt{\left(\frac{n}{n+k}\right)\left(\frac{k}{n+k}\right) \times |b-a|} = |b-a| \sqrt{p(1-p)}$

**Example 4a:** You have a group of 50 people and you are testing for some disease that has probability “$p$” in the population. There are 2 possible methods of testing:

1. Test each person separately.

2. Perform batch testing: You do one test if the result is negative, you do 51 tests if the result is positive.

On average, how many tests will be required?

Answer: Let the random variable $X$ be number of tests needed if samples are pooled. Thus, there are 2 possible values of $X : X = 1$ or $X = 51$.

Clearly $X = 1$ if nobody is infected, so $P(X = 1) = (1 - p)^{50}$ (assuming these are independent events.) Thus, $P(X = 51) = 1 - P(X = 1) = 1 - (1 - p)^{50}$
The expected value is

\[ E(X) = 1 \cdot (1 - p)^{50} + 51 \cdot [1 - (1 - p)^{50}] \]

The following table gives \( E(x) \) for various values of \( p \).

<table>
<thead>
<tr>
<th>( p )</th>
<th>( E(X) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>51.0</td>
</tr>
<tr>
<td>0.1</td>
<td>50.7</td>
</tr>
<tr>
<td>0.01</td>
<td>20.8</td>
</tr>
<tr>
<td>0.001</td>
<td>3.4</td>
</tr>
<tr>
<td>0.0001</td>
<td>1.2</td>
</tr>
</tbody>
</table>

**Example 4b** Say you are testing \( N \) people in batches of \( k \) people. What is the optimal batch size?

For simplicity, assume \( N \) is divisible by \( k \) (this is completely unimportant). Then there are \( N/k \) batches. In each batch you do either 1 or \( k+1 \) tests so in each batch, the expected number of tests is

\[ E(X) = 1 \cdot (1 - p)^k + (k+1) \cdot [1 - (1 - p)^k]. \]

Since we repeat this \( N/k \) times, the expected total number of tests is

\[ \text{Tests}(k) = \frac{N}{k} \left[ 1 \cdot (1 - p)^k + (k+1) \cdot [1 - (1 - p)^k] \right] = N \left[ 1 - (1 - p)^k + \frac{1}{k} \right]. \]

To get the optimal batch size, we want to pick \( k \) to minimize \( \text{Tests}(k) \). Thus the first derivative is zero:

\[ \frac{d}{dk} \text{Tests}(k) = 0. \]

Before doing so, we use that \( p \) is presumably small so by the binomial theorem

\[ (1 - p)^k = 1 - kp + \frac{k(k-1)}{2} p^2 + \text{smaller terms} \approx 1 - kp. \]

Then

\[ \text{Tests}(k) \approx N \left[ 1 - (1 - kp) + \frac{1}{k} \right]. \]

Consequently, we want to pick \( k \) so that

\[ 0 = \frac{d}{dk} \text{Tests}(k) \approx N \left[ p - \frac{1}{k^2} \right]. \]

so the optimal \( k \approx 1/\sqrt{p} \).
September 21, 2006

Copier Problem

The Dean of the Math Department decides to assign copy codes based on the last four digits of each faculty member’s Social Security Number. What is the probability of collision (multiple persons with the same copy code) if there are a) 2 people? 3 people? b) 100 people?

Answer: Let’s define some variables:

\( N = 10,000 \) (this is the number of unique 4-digit combinations)

\( k \) = the number of faculty members

\( Q_k \) is the probability of having no collision for any \( k \) number of people

a) For two people, the probability of having no collision is clearly \( \frac{9,999}{10,000} \). Similarly, for three people, the probability of having no collision is:

\[
Q_3 = \frac{9,999}{10,000} \times \frac{9,998}{10,000} \times \frac{9,997}{10,000} = \frac{(N-1)!}{N^{k-1}(N-k)!}
\]

For any \( k \) number of people, the probability

\[
Q_k = \frac{9,999}{10,000} \times \frac{9,998}{10,000} \times \cdots \times \frac{10,000-(k-1)}{10,000} \times \frac{(N-1)!}{N^{k-1}(N-k)!}
\]

For large \( n \), we can use **Sterling’s Formula**:

\[
n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n
\]

As \( n \to \infty \), \( \frac{a_n}{b_n} \to 1 \). Using this, \( Q_k = \frac{\sqrt{2\pi n} \left(\frac{n}{e}\right)^n}{N^k \sqrt{2\pi (N-k) \left(\frac{N}{e}\right)^{N-k}}} \). Then,

\[
Q_k \approx \frac{1}{e^k \left(\frac{n}{n-k}\right)^{n-k+\frac{1}{2}}}
\]

\[
\ln Q_k = (N-k+\frac{1}{2})\ln\frac{N}{N-k}
\]

Thus, for part b),

\[
\ln Q_{100} = (9900.5)\ln\left(\frac{10,000}{9,900}\right) = -0.4966
\]

\[
Q_{100} = e^{-0.4966} = 0.608 \approx 61\%
\]

Therefore, \( P_k \) = probability of a collision \( = 1 - Q_k = 39\% \).

**Conditional probability**

*Example [Life Expectancy]* From experimental data in a certain country, from a group of 100,000 women about 90% can expect to live to age 60 but only 55% are expected to live...
to age 80. If a woman is 60 years old, what is the probability that she will live to be age 80?

This is an example of conditional probability. The original sample space was all of these women. But now we have more information: this woman has already lived 60 years. This is the new sample space. It has $90,000 = 90\% \times 100,000$ women, and contains the subset of $55\% \times 100,000 = 55,000$ women who are expected to live to age 80. Thus, the probability that a woman who is now 60 will live to age 80 is $55,000/90,000 \approx .611 \approx 61\%$.

**Definition** [Conditional Probability]. The probability of a point being in a set $A$ knowing the point is in $B$ is written $P(A \mid B)$ and defined in an obvious manner:

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

Instead of using the language of sets, it is often to think of $A$ and $B$ as “events”.

**Example** [Ethnic Blood Types] The following probability table concerns the blood types of three ethnic groups.

<table>
<thead>
<tr>
<th></th>
<th>Type O</th>
<th>Type A</th>
<th>Type B</th>
<th>Type AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group X</td>
<td>.082</td>
<td>.106</td>
<td>.008</td>
<td>.004</td>
</tr>
<tr>
<td>Group Y</td>
<td>.135</td>
<td>.141</td>
<td>.018</td>
<td>.006</td>
</tr>
<tr>
<td>Group Z</td>
<td>.215</td>
<td>.200</td>
<td>.065</td>
<td>.020</td>
</tr>
<tr>
<td>Total</td>
<td>.432</td>
<td>.447</td>
<td>.091</td>
<td>.030</td>
</tr>
</tbody>
</table>

If someone is in Group Z, what is the probability they have blood type A?

We want to compute $P(A \mid Z)$. Using (2) we find $P(A \mid Z) = P(A \cap Z)/P(Z)$. Clearly: $P(A) = .447$, $P(Z) = .5$, $P(A \cap Z) = .2$ so

$$P(A \mid Z) = P(A \cap Z)/P(Z) = .2/.5 = 40\%.$$ 

Similarly,

$$P(Z \mid A) = P(Z \cap A)/P(A) = .2/.447 \approx 44.7\%.$$ 

If an individual does not have blood type B, what is the probability this individual is in Group X?

We want: $P(X \mid -B) = P(X \cap -B)/P(-B)$. But $P(-B) = 1 - P(B) = 1 - .091 = .909$ and $P(X \cap -B) = .82 + .106 + .004 = .192$ so:

$$P(X \mid -B) = .192/.909 \approx .211 = 21.1\%.$$
In equation (2) interchanging the roles of $A$ and $B$ we get

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}. \quad (3)$$

However, $A \cap B = B \cap A$. Thus we conclude the useful Bayes’ formula

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}. \quad (4)$$

**Definition:** Two events $A$ and $B$ are independent if $A$ and $B$ have positive probability and $P(A \mid B) = P(A)$. By Bayes’ formula (4), this is equivalent to both $P(B \mid A) = P(B)$ and $P(A \cap B) = P(A)P(B)$.

**Example** [Cancer Test]. Say a small percentage, $p\%$ of the population has a certain cancer. If you take the test, and you test positive, how upset should you be? In other words, What is the probability that you have this cancer?

To answer this we need data about the reliability of the test:

- If you have cancer, what is the probability $P(+) \mid C)$ of testing positive? Equivalently, If you have cancer, what is the probability $P(\neg C) = 1 - P(+) \mid C)$ of testing negative? These are false negatives.

- If you don’t have cancer, what is the probability $P(+) \mid \neg C)$ of testing positive? These are false positives. Equivalently, If you have don’t cancer, what is the probability $P(\neg C) = 1 - P(+) \mid \neg C)$ of testing negative?

It should be evident that both false positives and false negatives are undesirable. If there are false positives and you test positive, then the next step is often a more complicated (and painful) test to see if you really do have the cancer.

Here are some sample data about a certain test:

- $p = 1\%$
- $P(+) \mid \neg C) = 21\%$, so there are false positives. This is sometimes phrased (misleadingly) as “the test is 79% effective”
- $P(\neg C) = 0$, so there are no false negatives.

Since in a group of 100 people only 1 will have this cancer but roughly 20 will test positive, intuitively, we see that you have about a $1/20 = 5\%$ chance of having this cancer.

Here are the details of the computation. You have tested positive. We want to compute the likelihood that you have cancer, that is, $P(C \mid +)$. Bayes’ formula (4) states:

$$P(C \mid +) = \frac{P(+) \mid C)P(C)}{P(+)}.$$
Since there are no false negatives we know \( P(\cdot \mid C)P(C) = 1(0.01) = 0.01 \). Also \( P(\cdot \mid \neg C)P(\neg C) = (0.21)((0.99) = 0.2079 \) since there are false positives. Thus
\[
P(\cdot) = P(\cdot \cap C) + P(\cdot \cap \neg C)
\]
which, using (2), gives
\[
P(\cdot) = P(\cdot \mid C)P(C) + P(\cdot \mid \neg C)P(\neg C) = 1(0.01) + (0.21)((0.99) = 0.2179,
\]
so
\[
P(C \mid \cdot) = \frac{0.01}{0.2179} \approx 0.046 \approx 5%.
\]

To understand conditional probability problems, it often helps to use a tree diagram. Looking at the right side of the tree we can see that those who who test positive are either because they have cancer, 1.0 \times 1\% of the population, or because they don’t have cancer but have a false positive, 0.21 \times 99\% of the population. The ratio of those who test positive and also have cancer to all those who test positive is
\[
P(C \mid +) = \frac{1 \times 0.01}{1 \times 0.01 + (0.21 \times 0.99)} = \frac{0.01}{0.2179} \approx 0.046,
\]
just as before.

Remark: If there are false negatives, then one should compute \( P(C \mid \neg \cdot) \).

I took this example from K. Devlin’s web page http://www.maa.org/devlin/devlin2_00.html
He has many other interesting articles on mathematical topics.

In this cancer test example, the outcomes had only had 2 states: test positive or negative. Also, the possibilities are only that you either have cancer, or not. There are situations that arise when there are more mutually exclusive possibilities \( A_1, \ldots, A_n \), where \( \cup_{j=1}^n A_j = everything \). Since the \( A_j \)'s are mutually disjoint: \( A_i \cap A_j = \theta \) if \( i \neq j \). Then
\[
P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{P(B)}
\]
But,
\[
P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n).
\]
Therefore, using (2) for each term on the right,

\[ P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)} \]

**Monte Hall Problem**

There are 3 doors. Two of the doors hide goats; one of the doors hides a car. Monte Hall knows which doors hide goats and which door hides the car. You pick a number. It’s a goat. Do you change your mind, or have your odds stayed the same?