## Lectures 1-3 - Probability

Text: WMM, Chapter 1 (All Sections)
Chapter 2, Sections 2.1-2.7.

**Sample Space**

The set of all possible outcomes of a statistical experiment is called the sample space, S. (Definition 2.1, p. 11). Example: When you flip a coin once the sample space is heads or tails, S={H,T}. When you toss a die, the sample space is a number from 1 to 6, S={1,2,3,4,5,6}. When you flip two coins, the sample space is S={HH,HT,TH,TT}. Braces denote a **set**. Each entry in the set is called an **element**.

Instead of listing all of the elements of a big set, you can use a **rule** or **statement**. An example: (The pipe is read as “such that”)

$$S = \{(x, y) | x^2 + y^2 \le 4\}$$

**Event**

An event is a subset of a sample space. (Definition 2.2, p. 14). Example, if S={HH,HT,TH,TT}, possible subsets include: B={Ø}, B={HH}, B={HH,TT}, B=S.

**Complement**

The complement, A’, of an event A with respect to the sample space, S, is the subset of all elements of S that are not in A. (Definition 2.3, p. 14). Example, if S={HH,HT,TH,TT}, possible subsets and complements include: B={Ø} and B’=S; B={HH,TT} and B’={HT,TH}.

**Intersection**

The intersection of two events A and B, denoted by the symbol, $A \cap B$, is the event containing all elements in both A and B. (Definition 2.5, p. 15).Example:

if B={HH,TT} and A={HH,HT,TH} then $A \cap B$ = {HH}

if B={HH,TT} and A={HT,TH} then $A \cap B$ = {Ø}

**Mutually exclusive**, or **disjoint**

Two events A and B are mutually exclusive if $A \cap B$ = {Ø}, that is, if A and B have no common elements. (Definition 2.6, p. 15) Example: $B \cap B'$ = {Ø}, that is, an event and its complement are by definition mutually exclusive events.

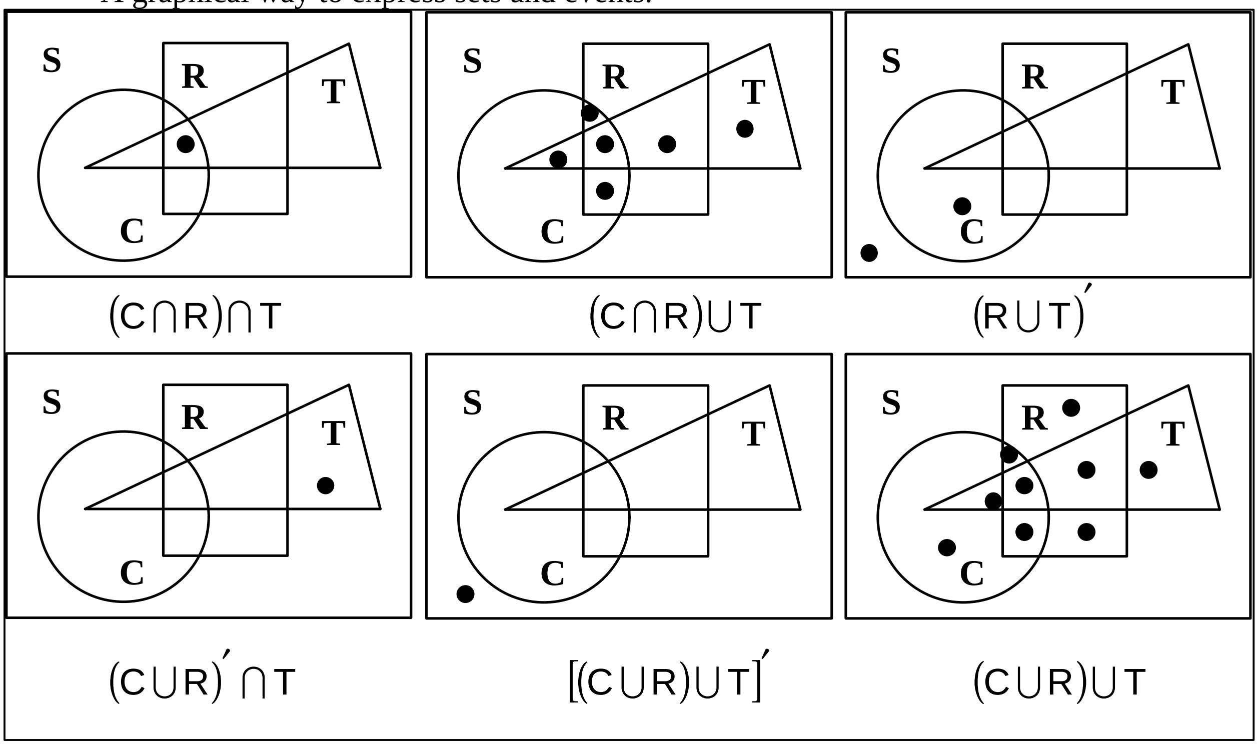
**Union**

The union of two events A and B, denoted by the symbol, $A \cup B$, is the event containing all elements in either A or B. (Definition 2.6, p. 15).Example:

$B \cup B'$ = S, that is, the union of an event and its complement is the sample space

**Venn Diagrams**

A graphical way to express sets and events.




**Counting Rules:**

We need counting rules in probability because the probability of an event A is the ratio of the number of elements in A over the number of elements in the sample space S.

$$P(A) = \frac{\#\text{ of elements in A}}{\#\text{ of elements in S}}$$

Therefore, we need to know how to count the number of elements in A and S. We will study three counting rules:

- generalized multiplication rule
- permutations of distinct objects rule
- combinations of distinct objects rule

**Generalized Multiplication rule**

If an operation can be performed in $n_1$ ways, and if for each of these, a second operation can be performed in $n_2$ ways, and for each of the first two, a third operation can be performed in $n_3$ ways, and so forth, the sequence of k operations can be performed in $n_1 n_2 n_3 \ldots n_k$ ways. (page 20) Example: If you consider the set of elements composed of one coin toss, followed by one die roll, followed by drawing from a hat containing m names. The number of elements in the set is 2*6*m.

**Permutation**

A permutation is an arrangement of all or part of a set of objects. (Definition 2.7, page 22). A permutation is a grouping of elements arranged in a particular way. Example: consider flipping a coin three times. The permutation {H,T,T} is different from the permutation {T,H,T} or {T,T,H}. All of these events are elements of the subset that contains the result one heads, and two tails, but, when you are interested in the order, they become distinct, and we call them permutations. When are permutations important? It depends on the individual situation.

The number of permutations of n distinct objects is n! That is "n factorial".

$$n! = n(n-1)(n-2)(n-3)\ldots 3 * 2 * 1. \tag{1.1}$$

The number of permutations of n distinct objects taken r at a time, where r $\leq$ n, is

$${}_nP_r = \frac{n!}{(n-r)!} \tag{1.2}$$

When is the formula applicable? This formula applies when the order of a result is important. For example: How many ways can a group schedule 3 different meetings on any of five possible dates? The answer is ${}_5P_3$ = 60. How did we know to use equation (1.2)? The key tip-off was the word "different". This means the meetings are distinguishable and order matters.

**Quick Calculation of Permutations by hand**

When n becomes large but r is small, it can be difficult to compute the permutation of ${}_nP_r$. Consider the case where n=200 and r=2. Then

$${}_nP_r = \frac{200!}{(200-2)!} = \frac{200!}{198!}$$

Our calculators cannot computer the factorial of 200 or 198. The numbers are too large. However, we can still obtain the number of permutations, if we consider that

$$200! = 200 * 199 * 198!$$

Then we have

$${}_nP_r = \frac{200!}{(200-2)!} = \frac{200 * 199 * 198!}{198!} = 200 * 199 = 39800$$

**Implementation of Permutations in MATLAB**

This simple code, perm.m, illustrates how one would numerically compute a permutation. It doesn't use the cancellation trick shown above. It computes the factorial of n, then computes the factorial of (n-x), then returns the quotient. If you want to write a better permutation code, you should include the cancellation trick.

```
%
%  FUNCTION perm(n,x)
%
%  This function computes the number of permutations
%  when choosing x distinct objects from n distinct objects
%
%  Author:  David Keffer, dkeffer@utk.edu
%  Department of Chemical Engineering, University of Tennessee, Knoxville
%  Last updated:  January 12, 2000
%
function f = perm(n,x)
fac1 = 1.0;
if (n > 1)
   for i = n:-1:2
      fac1 = fac1*i;
   end
end
fac2 = 1.0;
if (n-x > 1)
   for i = (n-x):-1:2
      fac2 = fac2*i;
   end
end
f = fac1/fac2;
```

**Combination**

A combination is a grouping of elements without regard to order. (p. 25) The number of combinations of n distinct objects taken r at a time, where r ≤ n, is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \tag{1.3}$$

**Quick Calculation of Combinations by hand**

When n becomes large but r is small, it can be difficult to compute the combination, $\binom{n}{r}_r$.

Consider the case where n=200 and r=2. Then

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{200!}{2!(200-2)!} = \frac{200!}{2! \cdot 198!}$$

Our calculators cannot computer the factorial of 200 or 198. The numbers are too large. However, we can still obtain the number of permutations, if we consider that

$$200! = 200 * 199 * 198!$$

Then we have

$$\binom{n}{r} = \frac{200!}{2! \cdot (200-2)!} = \frac{200 * 199 * 198!}{2! \cdot 198!} = \frac{200 * 199}{2} = 19900$$

## Implementation of Combinations in MATLAB

This simple code, comb.m, illustrates how one would numerically compute a combination. It computes the factorial of n, then computes the factorial of (n-x), then computes the factorial of x, then returns the n!/((n-x)!*x!). It doesn't use the cancellation trick shown above. If you want to write a better combination code, you should include the cancellation trick.

```
%
%  FUNCTION comb(n,x)
%
%  This function computes the number of combinations
%  when choosing x indistinct objects from n indistinct objects
%
%  Author:  David Keffer, dkeffer@utk.edu
%  Department of Chemical Engineering, University of Tennessee, Knoxville
%  Last updated:  November 19, 1999
%
function f = comb(n,x)
fac1 = 1.0;
if (n > 1)
   for i = n:-1:2
      fac1 = fac1*i;
   end
end
fac2 = 1.0;
if (n-x > 1)
   for i = (n-x):-1:2
      fac2 = fac2*i;
   end
end
fac3 = 1.0;
if (x > 1)
   for i = x:-1:2
      fac3 = fac3*i;
   end
end
f = fac1/(fac2*fac3);
```

## The difference between permutations and combinations

Example One.

What are the number of ways of arranging the letters A, B, C, when order matters?

$_3P_3 = 6$. They are {ABC, ACB, BAC, BCA, CAB, CBA}.

What are the number of ways of choosing A,B,C when order doesn't matter?

$$\binom{3}{3} = 1$$ The answer is {ABC}. Order doesn't matter, so that's the only set.

Example Two.

What are the number of ways of arranging 2 of the letters A, B, C, when order matters?

${}_3P_2$ =6. They are {AB, AC, BA, BC, CA, CB}.

What are the number of ways of choosing 2 of the letters A, B, C, when order doesn't matter?

$$\binom{3}{2} = 3$$ The answer is {AB,BC, AC}.

**Permutations of indistinct objects**

In some cases, you have both distinct and indistinct objects. In this case, you must combine the permutation and combination formulae. The number of distinct permutations of n things taken r at a time where there are $n_1$ of one kind, $n_2$ of the second kind, up to $n_k$ of the kth kind is:

$$\text{permutations\_of\_indistinct\_objects} = \frac{n!}{n_1!n_2!\ldots n_k!}$$

You can see that if all the objects are distinct, then every factor in the denominator is one and this reduces to the original permutation formula of 1.2.

For example, how many ways can you arrange all elements of the following set {A,AA,B,B,C}?

$$\frac{6!}{3!2!1!} = \frac{720}{1} * \frac{1}{12} = 60$$

**Probability**

The probability of an event A is the sum of the weights of all sample points in A. (definition 2.8, p. 28) Therefore,

$$0 \le P(A) \le 1$$
$$P(\varnothing) = 0$$
$$P(S) = 1$$

Example: On an ordinary die, any number is equally likely to turn up. The probability of getting any particular number is 1/6.

Example: If, you have an illegal 6 sided die that is weighted to preferentially yield 6. For example, instead of yielding each number between 1 and 6 1/6 of the time, it yields 6 half the time, and the rest of the numbers 1/10 of the time. The sum of the weights 5*(0.1) + 0.5 = 1. Therefore, the probability of getting a 6, P(6) = 0.5. P(1) = 0.1.

**Additive Rules**

The probability of Union (theorem 2.10, page 31):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.4}$$

Example: The sample space consists of the letter number pairs, {A1, A2, B1}.
The probability of getting a pair with an A or a pair with a 1 is

$$\begin{aligned}
&P(A \cup 1) = P(A) + P(1) - P(A \cap 1) \\
&P(A) = 2/3 \\
&P(1) = 2/3 \\
&P(A \cap 1) = 1/3 \\
&P(A \cup 1) = 2/3 + 2/3 - 1/3 = 1
\end{aligned}$$

Corollary One: if A and B are mutually exclusive:

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

Corollary Two: if $A_1$ $A_2$….$A_n$ are mutually exclusive:

$$P(A_1 \cup A_2 \cup ... \cup A_n) = \sum_{i=1}^{n} P(A_i)$$

Corollary Three: If $A_1$ $A_2$….$A_n$ are mutually exclusive and include all of the sample space,

$$P(A_1 \cup A_2 \cup ... \cup A_n) = \sum_{i=1}^{n} P(A_i) = P(S) = 1$$

Extension to three events

$$\begin{aligned}
P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) \\
& + P(A \cap B \cap C)
\end{aligned}$$

Example: Consider the set of ten objects {cat, dog, wolf, tiger, oak, elm, maple, opal, ruby, pearl}

What is the probability that you would randomly select a word that is (A) an animal OR that (B) has 4 letters OR that (C) starts with a vowel.) Two methods of solution. First, You can pick these out by hand. There are 4 animals, 2 trees that start with vowels, and 2 minerals that have 4 letters. Therefore the probability is 0.8 or 80%. The other method of solution is to use the equation given above to find the probability of the union.

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

$$P(A \cup B \cup C) = 4/10 + 3/10 + 3/10 - 1/10 - 0 - 1/10 + 0 = 8/10$$

**Conditional Probability**

The probability of an event B occurring when it is know that some event A has already occurred is called a conditional probability, P(B|A), and is read, "the probability of B given A", and is defined by:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \qquad \text{for } P(A) > 0 \tag{1.5}$$

*Example:* Using the above example of ten words. What is the probability that we choose a (B) three letter word given that we know that we have chosen an (A) animal.

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{2/9}{4/9} = 1/2$$

**Independence/Multiplicative Rules**

Two events are independent if and only if

$$P(B \mid A) = P(B) \qquad \text{and} \qquad P(A \mid B) = P(A) \tag{1.6}$$

Substituting eqns (1.6) into equation (1.5) we have:

Two events are independent if and only if

$$P(A \cap B) = P(A) * P(B)\,. \tag{1.7}$$

*Extension of the intersection rule to more than two events.* (Theorem 21.5, page 40)

If in an experiment, $A_1, A_2, A_3 \ldots A_k$ can occur then:

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \ldots P(A_k \mid A_1 \cap A_2 \cap \ldots \cap A_{k-1})$$

If the events are independent then the following is also true:

$$P(A_1 \cap A_2 \cap \ldots \cap A_k) = \prod_{i=1}^{k} P(A_i)$$

These equations are simply repeated applications of equations (1.5) and (1.6).

*Example Problems for Probability*
There are four example given below. In each example, let it be clear that **we use three and only three rules!** We use the rules for the union, the conditional probability, and the intersection. To restate these rules, we have:

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

Conditional: $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$ for $P(A) > 0$

Intersection: $P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$

*Example 1:* You flip a coin twice. What is the probability of getting heads on the second flip (B) given that you got heads on the first flip (A)? Are the events independent?

The probability of getting a head on the first flip is 0.5 The probability of getting a head on the second flip is 0.5. The intersection of A and B from the set {HH, HT, TH, TT} is 0.25. We see that equation (1.7) is satisfied, so the equivalent statement of equation (1.6) is also satisfied, yielding

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0.25}{0.5} = 0.5 = P(B)$$

Each flip of the coin is independent.

*Example 2:* You have a bag with 3 lima beans and 2 pinto beans in it. You draw 2 beans from it randomly without replacement. What is the probability that you draw a lima bean (B) given that you already drew a lima bean on the first draw (A)? Are the events independent?

The easiest way to solve this problem is to list all the possible outcomes.

The probability for drawing a lima bean the first time (event A) is 3/5 = 0.6

The probability for drawing a pinto bean the first time is 2/5 = 0.4

If we draw a lima bean the first time, then there are 2 lima beans and 2 pinto beans. In that case the probability for drawing a lima bean the second time is 2/4 = 0.5 and the probability for drawing a pinto bean the second time is 2/4 = 0.5.

If we draw a pinto bean the first time, then there are 3 lima beans and 1 pinto beans. In that case the probability for drawing a lima bean the second time is 3/4 = 0.75 and the probability for drawing a pinto bean the second time is 1/4 = 0.25.

So we have four possible outcomes {LL, LP, PL, PP}.

The probability of each outcome is given by the product of the probabilities of the events in that outcome. (This is the multiplicative rule.)

The probability of LL is 0.6*0.5 = 0.3.

The probability of LP is 0.6*0.5 = 0.3.

The probability of PL is 0.4*0.75 = 0.3.

The probability of PP is 0.4*0.1 = 0.1.

The individual probabilities then for {LL, LP, PL, PP} are {0.3, 0.3, 0.3, 0.1}. With these figures, we can write:

Now event A includes any outcome with L in the first draw, LL and LP.

$$P(A) = P(LL) + P(LP) = 0.3 + 0.3 = 0.6$$

Event B includes any outcome with L in the first draw, LL and PL. The sum of those two probabilities is 0.3+0.3 = 0.6 so

$$P(B) = P(LL) + P(PL) = 0.3 + 0.3 = 0.6$$

The intersection of A and B includes the outcome LL

$$P(A \cap B) = P(LL) = 0.3$$

Given this information, we have that the conditional probability of B given A, (or the probability that we draw a lima bean on the second draw, given that we drew a lima bean on the first draw) is

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{0.3}{0.6} = 0.5$$

Now, to check for independence,

$$P(B \mid A) = 0.5 \neq P(B) = 0.6$$

Therefore, the two experiments are not independent because equation (1.6) is not satisfied. Since equation (1.6) is equivalent to equation (1.7), we can also check equation (1.7) to verify independence.

$$P(A \cap B) = 0.3 \neq P(A) * P(B) = 0.6 * 0.6 = 0.36$$

Equation 1.7 also says that the events are not independent.

*Example 3:* (example 2.33 in WMM): One bag contains 4 White and 3 black marbles. A second bag contains three white and five black marbles. One marble is randomly drawn from the first bag and stuck in the second bag, unseen (A). What is the probability of drawing a black ball from the second bag (B)? Now A is composed of the event of drawing a white ball (W1) or a black ball (B1) from the first bag.

You want: $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)} = \dfrac{P((W1 \cap B) \cup (B1 \cap B))}{P(W1 \cup B1)}$

The probability of A is 1. You draw a marble; it is either black or white. Continuing and using equation (1.4)

$$P(B \mid A) = P((W1 \cap B) \cup (B1 \cap B)) = P(W1 \cap B) + P(B1 \cap B) - P((W1 \cap B) \cap (B1 \cap B))$$

The intersection of drawing a white ball first or a black ball first is 0.0, so the last term drops. Then substituting equation (1.5) into the above equation, we have:

$$P(B \mid A) = P(W1 \cap B) + P(B1 \cap B) = P(W1)P(B \mid W1) + P(B1)P(B \mid B1)$$

P(W1) = 4/7, P(B1)=3/7. Once you draw a white marble from bag one and put it in bag two, the probably of drawing a black marble from bag two is 5/9. Similarly, once you draw a black marble from bag one and put it in bag two, the probably of drawing a black marble from bag two is 6/9. Plugging these numbers into the equation above:

$$P(B \mid A) = P(W1)P(B \mid W1) + P(B1)P(B \mid B1)$$
$$= \left(\frac{4}{7}\right)\left(\frac{5}{9}\right) + \left(\frac{3}{7}\right)\left(\frac{6}{9}\right) = \frac{38}{63}$$

*Example 4.*

In sampling a population for the presence of a disease, the population is of two types: **I**nfected and **U**ninfected. The results of the test are of two types: **P**ositive and **N**egative. In rare disease detection, a high probability for detecting a disease can still lead to more false positives than true positives. Consider a case where a disease infects 1 out of every 100,000 individuals. The probability for a positive test result given that the subject is infected is 0.99. The probability for a negative test result given that the subject is uninfected is 0.999.

(1) For testing a single person, define the complete sample space.

(2) What is the probability of a false negative test result (a negative test result given that the subject is infected)?

(3) What is the probability of being uninfected AND having a negative test result?

(4) What is the probability of testing positive?

(5) Determine rigorously whether testing positive and having the disease are independent.

(6) Determine the percentage of people who test positive who are really uninfected.

(7) In a population of 250 million, with the infection rate given, how many people would you expect to be
(a) Infected-test Positive, (b) Infected-test Negative, (c ) Uninfected-test Positive, (d) Uninfected-test negative.

**Solution:**

We are told:

$$P(I) = 10^{-5}$$

$$P(N|U) = \frac{P(N \cap U)}{P(U)} = 0.999$$

$$P(P|I) = \frac{P(P \cap I)}{P(I)} = 0.99$$

(1) For testing a single person, define the complete sample space.

The sample space is $S = \{IP, IN, UP, UN\}$ where I = Infected, U=Uninfected, P=positive test result, N=negative test result.

The Venn Diagram looks like this:

| | |
|---|---|
| Infected $\cap$ Positive | Infected $\cap$ Negative |
| Uninfected $\cap$ Positive | UnInfected $\cap$ Negative |

When you have a simple sample space like this, you can see some additional constraints on the system, in addition to the union, conditional, and intersection rules. You will need some of these additional constraints to solve the problems below.

For example, if a person tests positive, they are either infected or uninfected. Therefore, using the union rule we have:

$$P(P) = P[(P \cap I) \cup (P \cap U)] = P(P \cap I) + P(P \cap U) - P[(P \cap I) \cap (P \cap U)]$$

There is no intersection between being infected and uninfected, therefore:

$$P(P) = P(P \cap I) + P(P \cap U)$$

We can write three other analogous contraints:

$$P(N) = P(N \cap I) + P(N \cap U)$$
$$P(U) = P(U \cap P) + P(U \cap N)$$
$$P(I) = P(I \cap P) + P(I \cap N)$$

Remember, these four rules only work for simple (but common) Venn Diagrams like the one shown above.

Also consider that the probability of being infected given a person is positive plus the probability of being uninfected given a person is positive is 1. A person is either infected or uninfected, regardless of whether they tested positive or negative. We can write this as.

$$P(I\,|\,P) + P(U\,|\,P) = 1$$

This is just a restatement of the fact that the sums of the probabilities must be equal to 1. Again, three analogous statements can be made:

$$P(I\,|\,N) + P(U\,|\,N) = 1$$

$$P(P|U) + P(N|U) = 1$$
$$P(P|I) + P(N|I) = 1$$

Remember, these four rules only work for simple (but common) Venn Diagrams like the one shown above.

In solving the problems, below, remember we have this group of rules. There are many ways to solve some of the problems. We just go looking for the one that seems easiest.

(2) What is the probability of a false negative test result (a negative test result given that the subject is infected)?

We want: $P(N|I) = \frac{P(N \cap I)}{P(I)}$ so we need the two factors on the right hand side.

We have been given the denominator. In order to find the numerator, we must use the other given:

$$P(P|I) = \frac{P(P \cap I)}{P(I)} = 0.99$$

which rearranges for the intersection of P AND I

$$P(P \cap I) = P(I) \cdot P(P|I) = \left(10^{-5}\right)(0.99) = 0.99 \cdot 10^{-5}$$

We must realize that the probability of I is the union of IP and IN groups.
So using the definition of the Union, we have:

$$P(I) = P[(I \cap P) \cup (I \cap N)] = P(I \cap P) + P(I \cap N) - P[(I \cap P) \cap (I \cap N)]$$

The result cannot be both positive and negative:

$$P[(I \cap P) \cap (I \cap N)] = 0$$

So,

$$P(I \cap N) = P(I) - P(I \cap P) = \left(10^{-5}\right) - 0.99 \cdot 10^{-5} = 10^{-7}$$

Then we can plug into our original equation:

$$P(N|I) = \frac{P(N \cap I)}{P(I)} = \frac{10^{-7}}{10^{-5}} = 0.01$$

OR, an alternative solution, relies on us recognizing:

$P(N|I) + P(P|I) = 1$ because every test comes out positive or negative.

$$P(N|I) = 1 - P(P|I) = 1 - 0.99 = 0.01$$

(3) What is the probability of being uninfected AND having a negative test result?

We want $P(N \cap U)$

we can obtain this from either:

(a) the UNION RULE:

$$P(N) = P[(N \cap I) \cup (N \cap U)]$$
$$P(N) = P(N \cap I) + P(N \cap U) - P[(N \cap I) \cap (N \cap U)]$$
$$P[(N \cap I) \cap (N \cap U)] = 0$$

$$P(N) = P(N \cap I) + P(N \cap U)$$
$$P(N \cap U) = P(N) - P(N \cap I)$$

but we don't know $P(N \cap I)$ and we don't know $P(N)$

or (b) the conditional probability rule:

$$P(U|N) = \frac{P(N \cap U)}{P(N)}$$
$$P(N \cap U) = P(N) \cdot P(U|N)$$

but we don't know $P(U|N)$ and we don't know $P(N)$

or (c) the conditional probability rule:

$$P(N|U) = \frac{P(N \cap U)}{P(U)} = 0.999$$
$$P(N \cap U) = P(U) \cdot P(N|U) = P(U) \cdot 0.999$$

I like choice (c) because we are given $P(N|U) = 0.999$ and we know

$P(U) = 1 - P(I) = 1 - 10^{-5} = 0.99999$ so

$$P(N \cap U) = P(U) \cdot P(N|U) = (0.99999)(0.999) = 0.99899001$$

(4) What is the probability of testing positive?

We want $P(P)$

We can find $P(P)$ either by:

(a) the fact that the sum of the probabilities must be one

$P(P) + P(N) = 1$ but we don't know $P(N)$

$$P(P) = 1 - P(N)$$

(b) the conditional probability distribution:

$P(I|P) = \dfrac{P(P \cap I)}{P(P)}$ but we don't know $P(I|P)$

(c) the conditional probability distribution:

$P(U|P) = \dfrac{P(P \cap U)}{P(P)}$ but we don't know $P(U|P)$

(d) the sum of the probabilities must be one and a different conditional probability:

$$P(P) = 1 - P(N)$$
$$P(U|N) = \frac{P(N \cap U)}{P(N)}$$

$$P(P) = 1 - P(N) = 1 - \frac{P(N \cap U)}{P(U|N)}$$ but we don't know $P(U|N)$

(e) the sum of the probabilities must be one and a different conditional probability:

$$P(P) = 1 - P(N)$$

$$P(I|N) = \frac{P(N \cap I)}{P(N)}$$

$$P(P) = 1 - P(N) = 1 - \frac{P(N \cap I)}{P(I|N)}$$ but we don't know $P(I|N)$

(f) the Union rule:

$$P(P) = P[(P \cap I) \cup (P \cap U)]$$

$$P(P) = P(P \cap I) + P(P \cap U) - P[(P \cap I) \cap (P \cap U)]$$

$$P[(P \cap I) \cap (P \cap U)] = 0$$

$$P(P) = P(P \cap I) + P(P \cap U)$$

combine with conditional probabilities that we do know:

$$P(P) = P(I) * P(P | I) + P(U) * P(P | U)$$

I like choice (f):

$$P(P) = 10^{-5} \cdot 0.99 + 0.99999 \cdot P(P | U)$$

we can get the last factor by considering (as we did in part (2))

$P(P|U) + P(N|U) = 1$ because all tests are either positive or negative.

$$P(P|U) = 1 - P(N|U) = 1 - 0.999 = 0.001$$

so

$$P(P) = 10^{-5} \cdot 0.99 + 0.99999 \cdot 0.001 = 0.00100989$$

(5) Determine rigorously whether testing positive and having the disease are independent.

If $P(P)$ and $P(I)$ are independent:

$$P(P \cap I) = P(P) \cdot P(I)$$

$$0.99 \cdot 10^{-5} = 0.00100989 \cdot 10^{-5}$$

NOT INDEPENDENT.

(6) Determine the percentage of people who test positive but who are really uninfected.

We want: $\frac{P(P \cap U)}{P(P)}$

$$\frac{P(P \cap U)}{P(P)} = \frac{0.99999 \cdot 10^{-3}}{0.00100989} = 0.990196952 = 99\%$$

Despite the high accuracy of the test 99% of those people who test positive are actually uninfected.

(7) In a population of 250 million, with the infection rate given, how many people would you expect to be
(a) Infected-test Positive, (b) Infected-test Negative, (c ) Uninfected-test Positive, (d) Uninfected-test negative.

These are four intersections:

From part (5) we know:

$$P(P \cap I) = 0.99 \cdot 10^{-5}$$

$$P(P \cap U) = 0.99999 \cdot 0.001 = 0.99999 \cdot 10^{-3}$$

From part (2) we know

$$P(N \cap I) = P(I) - P(P \cap I) = \left(10^{-5}\right) - 0.99 \cdot 10^{-5} = 10^{-7}$$

From part (3) we know

$$P(N \cap U) = P(U) \cdot P(N|U) = (0.99999)(0.999) = 0.99899001$$

These should sum to 1.0 and they do.

Out of 250 million people, the number who are infected and test positive are: 2475.
Out of 250 million people, the number who are infected and test negative are: 25.
Out of 250 million people, the number who are uninfected and test positive are: 249,997.5
Out of 250 million people, the number who are uninfected and test negative are: 249.7475 million