Notes 1 for Honors Probability and Statistics

Ernie Croot

August 26, 2003

1 Set Theory

1.1 Basic Definitions

In mathematics a set is a collection of *elements* or *objects*. We also allow S to have no objects, and we call this special kind of set the *empty set*, and denote it by $\emptyset$.

If S denotes a set, the symbol $\in$ is used to express when an object is a member of S: The string of symbols "$a \in S$" is shorthand for the statement "a is a member of the set S". Some common sets that you are no doubt familiar with include $\mathbb{Z}$, the set of integers; $\mathbb{N}$, the set of positive integers; $\mathbb{Q}$, the set of rational numbers; $\mathbb{R}$, the set of real numbers; and $\mathbb{C}$, the set of complex numbers. We also have the notion of a subset and superset: Given sets A and B, if we have that every element $a \in A$ is also an element of B – that is, $a \in B$ – then we say that A is a subset of B, and we denote it by $A \subseteq B$. We say that the sets A and B are *equal*, and we denote it by $A = B$, if $A \subseteq B$ and $B \subseteq A$; also, if A is not equal to B, we write this as $A \neq B$. If A is a subset of B and $A \neq B$, then we say that A is properly contained in B or that A is a proper subset of B, and we write this as $A \subset B$. We also use the notation $B \supseteq A$ and $B \supset A$, and it is hoped that you can deduce its meaning. One final thing to note is that the empty set is always a subset of A; that is, $\emptyset \subseteq A$. If A is not the empty set, then $\emptyset$ is a proper subset of A, which means we would use the notation $\emptyset \subset A$.

One common way to define a set is to list out the elements contained within it: For example, we could say that $S = \{1, 2, 3, 4\}$; that is, S is the set of all integers between 1 and 4, inclusive. For sets that have too many elements to list out, but which can be algorithmically generated, one uses the notation

$$S = \{s : ...\},$$

where the "..." identifies the properties satisfied by s. This notation for S is usually read out loud as "S is the set of all elements s such that ...". For example, if we let

$$S = \{n \in \mathbb{Z} : n/2 \in \mathbb{Z}\},$$

one reads this as "S is the set of integers n such that $n/2$ is also an integer". It is obvious that this definition of S is a roundabout way of saying that S is the set of even integers.

Sets do not have to be just numbers, but can be any collection of objects we desire to work with; for example, the set $\mathbb{Z}[x]$ usually means the set of polynomials in the variable x, so this set is all polynomials of the form $a_k x^k + a_{k-1} x^{k-1} + \cdots + a_0$, where the a_i's are integer and k can be any non-negative integer. Notice here that not every element of $\mathbb{Z}[x]$ is a number.

There is a problem when one allows *too much* freedom for how one defines sets algorithmically, and this is perhaps best illustrated by something called Russell's paradox, in honor of the famous British Philospher and mathematician Bertrand Russell: First note that the elements of a set S can themselves be sets; for example the set $\{\{1, 2\}, \{1, 2, \{3, 4\}\}, 5\}$ has three "elements", namely the sets $\{1, 2\}$, $\{1, 2, \{3, 4\}\}$, and the number 5. Notice here that the second element is itself a set. The paradox comes in when we define sets that contain themselves.

Russell's Paradox. Let S be the set of all sets that are not contained in themselves; that is,

$$S = \{T \text{ a set } : T \notin T\}.$$

This looks like a perfectly valid mathematical statement, albeit a complicated one. Another way of stating S is that

$$T \in S \text{ if and only if } T \notin T.$$

But now ask yourself whether S is in itself or not; that is, set $T = S$. Then you get

$$S \in S \text{ if and only if } S \notin S,$$

which is clearly not possible. Thus, our definition for S is not logically possible. Incidentally, another way of stating the same paradox, and in a way that it a little easier to understand, is in terms of the "Barber of Seville": The barber of Seville only shaves those people who don't shave themselves. If the barber of Seville shaved himself, then that would be a contradiciton since he only shaves people who don't shave themself; and if he didn't shave himself, then he would have to shave himself. So, like with Russell's paradox, one can conclude that there is no such person as the Barber of Seville.

There are various systems of axioms for set theory which avoid this paradox, by disallowing S above to be deemed a set, be we will not concern ourselves further with it. Incidentally, the brand of set theory prior to Russell's work is called "naive set theory".

1.2 Intersections, Unions, Complements, Cartesian Products, and Powersets

There are various ways of building up sets from other sets by taking intersections, unions, complements, cartesian products, and powersets, and we will presently discuss these constructions.

Given two sets A and B, the set of all elements that are in both A and B is called the intersection of A and B, and is denoted by $A \cap B$. The set of elements of that are either in the set A or the set B is called the union of A and B, and is denoted by $A \cup B$. Finally, if A is a subset of some universal set U, then the complement of A with respect to U is denoted by $\overline{A}$, and is the set of all elements of U that *are not* contained in A. In most cases, the set U will not be stated explicitly and so must be gleaned from the context of the discussion: For example, if one is reading a mathematical paper where the object of study is the set of integers, then U will be the set of integers, unless otherwise stated.

Here are some basic laws concerning set intersection, union, and complementation:

I. If $A \cap \overline{A} = \emptyset$.
II. $A \cup \overline{A} = U$.
III. $\overline{\overline{A}} = A$.
IV. If $A \subseteq B$, then $\overline{B} \subseteq \overline{A}$.
V. If $A \subseteq B$ and $C \subseteq B$, then $(A \cup C) \subseteq B$.
VI. $\overline{A} \cup \overline{B} = \overline{A \cap B}$.
VII. $\overline{A} \cap \overline{B} = \overline{A \cup B}$.

Parts I through V are easy to prove. Here is a proof of part VI, and proof of part VI is similar (or can be proved by assuming VI, and taking complements):

Proof of VI. We will show that $\overline{A} \cup \overline{B} \subseteq \overline{A \cap B}$ and that $\overline{A \cap B} \subseteq \overline{A} \cup \overline{B}$. To prove the first part, we note that since $A \cap B \subseteq A$ and $A \cap B \subseteq B$, from property IV we get $\overline{A} \subseteq \overline{A \cap B}$ and $\overline{B} \subseteq \overline{A \cap B}$; and then, from property V we deduce $(\overline{A} \cup \overline{B}) \subseteq \overline{A \cap B}$.

To prove the second part, let $x \in \overline{A \cap B}$. If $x \in \overline{A}$, then we are done, since then $x \in \overline{A} \cup \overline{B}$. So, we may assume that $x \notin \overline{A}$, which is the same as saying that $x \in A$. We will presently deduce that $x \in \overline{B}$, which would imply $x \in \overline{A} \cup \overline{B}$, and therefore finish the proof. To make this deduction, suppose that, for proof by contradiction, $x \in B$. Then, since x is also in A we would have $x \in A \cap B$, which is impossible since we are given also that $x \in \overline{A \cap B}$, which would mean that $x \in \emptyset = \overline{A \cap B} \cap (A \cap B)$. Thus x is not in B, which implies that $x \in \overline{B}$. ■

Given two sets A and B, the cartesian product of A and B, denoted by $A \times B$, is the set of all ordered pairs [1] (a, b), where $a \in A$ and $b \in B$. If B is the same set as A, then the cartesian product $A \times B$ is sometimes denoted as A^2. One can define the cartesian product of any number of sets (instead of just two sets), and the definition of such a set (cartesian product) is what you think it should be.

Given a set A, the *powerset* of A, denoted by 2^A is the set of all subsets of A. So, for example, if $A = \{1, 2, 3\}$, then

$$2^A = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}.$$

Notice here that we use $\{1\}$ to denote the subset of A consisting of the single number 1, rather than just 1. This is because *elements* and *subsets* are two different kinds of objects: It would be correct to say that $\{1\} \in 2^A$, but incorrect to say $1 \in 2^A$. Also notice that the empty set is in A, since it is indeed a subset of A. Now, if A has k elements, then 2^A will have 2^k elements; and so, one can see why the notation 2^A for the powerset was chosen.

[1] The phrase "ordered pair" means that we distinguish between (a, b) and (b, a); for example, the ordered pair $(1, 2)$ is not the same as $(2, 1)$. There are some contexts where it is *not* important to distinguish between the two orderings; for example, if we say that S is the set $\{1, 2\}$, then it doesn't matter if we list the elements of S like that, or as $\{2, 1\}$.

1.3 Maps and the Cardinality of a Set

Before we can proceed further with our discussion of sets, we will need to introduce the notion of surjective, injective, and bijective functions.

Given two sets A and B, we say that f is a function or map from A to B, and denote it by $f : A \to B$ if and only if to each $a \in A$ there corresponds one and only element $f(a) \in B$. This definition is the same as the usual definition of a function. A more abstract way to define functions is as ordered pairs in $A \times B$. More specifically, one can define a function f as a subset of $A \times B$, where for each $a \in A$ there is exactly one $b \in B$ such that the ordered pair $(a, b) \in f$. I admit this is a rather strange way to define functions, and I am not sure myself why some mathematicians even bother using this definition. Often when a candidate function f fails to be a genuine function, it is easy to spot why; however, sometimes it is not so obvious. The basic problem that often occurs is that there are different names for the same element of A, and if f is indeed a function then it had better send the element defined by these different names to the same target element in B; that is, in order for f to be a function, it must be "well-defined", which means that if $x, y \in A$, with $x = y$, then $f(x) = f(y)$. Here is an example of a candidate f that is *not* well-defined, and so is *not* a function: For each integer j, we let $j + 2\mathbb{Z}$ denote the set $\{j + 2z \ : \ z \in \mathbb{Z}\}$. It is obvous that $j + 2\mathbb{Z} = k + 2\mathbb{Z}$ if and only if j and k are of the same parity; that is, you get equality here if j and k are both even or both odd. Let A and B both be the two element set $\{1 + 2\mathbb{Z},\ 0 + 2\mathbb{Z}\}$, and consider candidate function f from A to B given by

$$f(j + 2\mathbb{Z}) \ = \ \begin{cases} 0 + 2\mathbb{Z}, & \text{if } j = 0; \\ 1 + 2\mathbb{Z}, & \text{if } j \neq 0. \end{cases}$$

If we restrict ourselves to $j = 0, 1$, then this does indeed give a function from A to $B = A$; however, if we allow j to be any integer, then this will not give a function from A to B, because it is not well-defined. That is, for example,

$$f(0 + 2\mathbb{Z}) \ = \ 0 + 2\mathbb{Z} \ \neq \ 1 + 2\mathbb{Z} \ = \ f(2 + 2\mathbb{Z}),$$

and yet $0 + 2\mathbb{Z} = 2 + 2\mathbb{Z}$, which is a case where we have two different names for the same element of A.

Associated to the function f we have the image of f, denoted by $\mathrm{im}(f) = \{f(a) \ : \ a \in A\}$, which is the set of all $b \in B$ that f "lands on". Another collection of sets associated to f are the *fibres* or *inverse images*: Given

$f : A \to B$, and given a set $C \subseteq B$, the set $D = \{a \in A \ : \ f(a) \in C\}$ is called the inverse image of C, and is denoted by $f^{-1}(C)$. For a given element $b \in B$, we let $f^{-1}(b)$ denote the set $f^{-1}(\{b\})$. [2] According to this definition we do not have to have that C lies entirely in $\mathrm{im}(f)$; indeed, we may have that $f^{-1}(C)$ is empty, and this occurs precisely when $C \cap \mathrm{im}(f) = \emptyset$. If two sets $C, D \subseteq B$ are disjoint, that is $C \cap D = \emptyset$, then $f^{-1}(C)$ and $f^{-1}(D)$ are also disjoint. In fact, the collection of sets $f^{-1}(b)$, where b runs through the elements of B, form what is called a *disjoint partition* of A: A collection of sets $A_1, A_2, \ldots$ form a disjoint partition of A if $A = A_1 \cup A_2 \cup \cdots$, and if the A_i's are pairwise disjoint, by which we mean $A_i \cap A_j = \emptyset$ for $i \neq j$. Since we know that the sets $f^{-1}(b)$ are disjoint, to show that they form a parition of A, we just need to see that their union equals A. This too is easy to see, since each $a \in A$ is in at least one of the sets $f^{-1}(b)$, namely $b = f(a)$.

If $\mathrm{im}(f) = B$, then we say that f is surjective or onto; that is, f surjective means that every element of B gets "landed on" by f. We say that f is injective or one-to-one if $f(a) = f(c)$, $a, c \in A$, implies $a = c$. Another way of saying this is that $f^{-1}(b)$ is either the empty set or contains exactly one element. Finally, if f is both surjective and injective, then we say that f is a one-to-one correspondance, or is bijective, or is a bijection.

Given a set C we denote the identity map on C by 1_C, which is a function from C to C such that $1_C(c) = c$ for every $c \in C$. We say that f has an inverse g if g is a function from B to A such that the composition $(f \circ g)(b) = f(g(b))$ is the identity on B and the composition $(g \circ f)(b)$ is the identity on A; that is, $f \circ g = 1_B$ and $g \circ f = 1_A$. It turns out that, and is not difficult to prove, f has an inverse g if and only if f is a bijection.

Given two sets A and B we write $|A| = |B|$ if there is a bijection $f : A \to B$. This is a generalization of the usual notion of $|A|$ and $|B|$, which is the number of elements of A and B, which only has meaning when A and B are finite sets, and certainly if two finite sets have the same number of elements if and only if there is a bijection between them (the sets).

We can also put a partial ordering on sets as follows:[3] Given sets A and B, we say that A is *less than or equal* to B, and we write this as either $A \leq B$ or $|A| \leq |B|$, if there exists an injection $f : A \to B$. If $A \leq B$, but $|A| \neq |B|$,

[2] The reason for this extra definition is that $f^{-1}(C)$ is defined for subsets $C \subseteq B$, not for *elements* of B.

[3] Actually, sets can be "well-ordered", by which we mean that given any two sets A, B, either $|A| \leq |B|$ or $|B| \leq |A|$. It turns out that the fact that sets can be well-ordered is equivalent to the axiom of choice.

then we write $A < B$.

Here are some further properties of functions and sets:

I. (Inclusion Preserving) If $f : A \to B$, and $C \subseteq D \subseteq B$, then $f^{-1}(C) \subseteq f^{-1}(D)$.

II. (Transitivity of Cardinalities) If $|A| = |B|$ and $|B| = |C|$, then $|A| = |C|$.

III. If $A \leq B$ and $B \leq A$, then $|A| = |B|$. This statement is equivalent to saying that if there exists injections from A to B and from B to A, then there exists a bijection from A to B (and B to A).

IV. $|A| < |2^A|$; that is, there is an injection from A to 2^A but there cannot be a bijection from A to 2^A. This is not obvious to prove in the case where A is infinite, and the usual proof uses Cantor's diagonalization argument, as we'll see.

Of these properties, the one that would give you the most trouble if you tried to prove it is property IV. Therefore, I will give the proof here.

Proof of IV. Suppose, for proof by contradiction that there exists a bijection $\phi : A \to 2^A$. The idea is to now use Cantor's trick (i.e. diagonalization) to show that ϕ cannot be surjective, and therefore couldn't have been a bijection, which would imply that $A < 2^A$. To show this (that ϕ is not surjective), we cook up an element in 2^A that cannot get mapped to by ϕ as follows: Let $B \subseteq A$ be defined as follows

$$b \in B \text{ if and only if } b \notin \phi(b). \tag{1}$$

Notice how similar this looks to the definition of S in Russell's paradox. Now, if ϕ were surjective there would have to be an element $b \in A$ such that $\phi(b) = B$, but this is impossible since upson setting $\phi(b) = B$ in (1) we would have

$$b \in B \text{ if and only if } b \notin B = \phi(b).$$

Thus, as claimed, ϕ is not surjective, and the claim follows. ∎

This proof is perhaps a little difficult to understand if you haven't seen it before; also, even if you have seen Cantor's diagonal trick, you might not recognize in the form presented above. Thus, I will give you a little of the intuition behind the proof, by considering the case where $A = \mathbb{N}$ (the positive integers): Suppose you had a bijection ϕ from A to 2^A. You can think of this function by imagining that you have a sheet of paper with an infinite number of rows and an infinite number of columns. Each row will be filled with an infinite number of 0's and 1's, and the pattern of 0's and 1's in the jth row will tell you which subset of A the set $\phi(j)$ happens to be: There will be a 1 in the ith column (and jth row) if the number $i \in \phi(j)$, and there will be a 0 in this column (again, jth row) if the number $i \notin \phi(j)$.

The idea of Cantor's argument is to cook up a string of 0's and 1's (that is, a subset of A) such that this string differs from all the other strings in all the rows of your infinite sheet of paper in at least one column entry. This string will thus correspond to a subset of A which does not get mapped to by ϕ, because if it were mapped to by ϕ, then it would have to have the same pattern of 0's and 1's as one of the rows on your sheet of paper. One such string can be gotten by letting its ith entry be different from the ith column entry of the string in the ith row of your piece of paper. If you call this string B, then the ith entry of B is a 1 if the entry in the ith row, ith column is a 0; and the ith entry of B is a 0 if the entry in the ith row, ith column is a 1. Clearly this B has the property we are looking for, and so we deduce ϕ cannot be surjective. To see that this really is the same proof as we gave above (at least when $A = \mathbb{N}$), note that the condition that the ith entry of B is different than the entry in the ith row, ith column is equivalent to the condition (when we now think of B as a subset of A, rather than as a string of 0's and 1's)

$$i \in B \ \text{ if and only if } \ i \notin \phi(i).$$

Here are two more facts about cardinalities:

I. There is an infinite sequence of infinite sets, no two of which are in bijection. For example,

$$|\mathbb{N}| \ < \ |2^{\mathbb{N}}| \ < \ \left|2^{2^{\mathbb{N}}}\right| \ < \ \cdots.$$

All sets with the same carinality as the integers as said to be $\aleph_0$ (read "aleph null"; aleph is the first letter of the Hebrew alphabet). The set $\aleph_1$ is defined to be the smallest infinite set larger than $\aleph_0$, and we will not concern ourselves further with it here.

II. Any set S with $|S| = |\mathbb{N}|$ is said to be *countable*; and if $|S| \neq |\mathbb{N}|$ and S infinite, then S is said to be *uncountable*. The sets $\mathbb{N}, \mathbb{Z}$, and $\mathbb{Q}$ are all countable sets; that is,

$$|\mathbb{N}| \ = \ |\mathbb{Z}| \ = \ |\mathbb{Q}|.$$

It turns out that

$$|\mathbb{C}| \;=\; |\mathbb{R}| \;=\; |[0,1]| \;=\; |2^{\mathbb{N}}|,$$

and so are all uncountable.

2 Sample Space, events, and measures

NOTE: Before I embark on the discussion in this section, I first want to point out that we will take a more abstract and more general approach to things than what appears at the beginning of chapter 2 of Meyer. I will point out later in these notes how the discussion differs from the one presented there.

Formally, a probability space is a triple (S, Σ, P), where S is called the sample space, Σ is call the set of measurable events, and P is the probability measure. We will now see what these three things mean.

2.1 Basic Definitions

When we perform an experiment, the set of all possible outcomes is called the sample space S. For example, if the experiment is "flip a coin three times", then the possible outcomes are HHH, HHT, HTH, HTT, THH, THT, TTH, and TTT; so, there are 8 possible outcomes. These outcomes are sometimes called "elementary events". In this example we could let $S = \{HHH, ..., TTT\}$ be the sample space. Depending on the way one records the experiment, the sample space could be different; thus, there is no one right sample space for a given experiment.

Once one has identified the sample space S, one next defines the set of events E to be the set of all subsets of S. Thus, $E = 2^S$.

Next, one picks out a special subset $\Sigma \subseteq E$, called a σ-algebra, which satisfies the following properties:

I. $\emptyset \in \Sigma$;

II. If $A \in \Sigma$, then $\overline{A} \in \Sigma$ (here the complement is taken with respect to S); and,

III. If $A_1, A_2, ...$ is a countable sequence of events in Σ, then their union also belongs to Σ.

A σ-algebra is a generalization of an algebra, which has the same defining properties as a σ-algebra, except that for property III, only finite unions are allowed. One obvious corollary of Σ being a σ-algebra is that it is also closed under countable intersections, by which we mean: If $A_1, A_2, \ldots$ is a countable sequence of events in Σ, then their intersection also belongs to Σ. The proof follows since by de Morgan's law we have:

$$\bigcap_i A_i \;=\; \overline{\bigcup_i \overline{A_i}},$$

and the right-hand-side belongs to Σ since each $\overline{A_i}$ belongs to Σ, then the contable union of these $\overline{A_i}$ belongs to Σ, and finally, the complement of this countable union then belongs to Σ.

The σ-algebra Σ that we will work with is not just *any* σ-algebra, but is one for which we have a probability measure P. Formally, a probability measure P is a mapping $P : \Sigma \to [0, 1]$ which satisfies the following properties:

I. If $A_1, A_2, \ldots$ is a countable sequence of disjoint events in Σ (to say they are disjoint means they are disjoint as sets), then

$$P\left(\bigcup_i A_i\right) \;=\; \sum_i P(A_i).$$

II. $P(S) = 1$.

Side Notes: One might worry that $\emptyset$ and S is not in Σ, and so one might would worry that their probabilities could not be computed. However, one sees that $\emptyset$ lies in Σ by definition, and S lies in Σ, being the complement of $\emptyset$.[4] Another question: Why do we only allow a countable number of sets for property I? Why not uncountable? And the answer is: Otherwise, what would the sum of $P(A_i)$ over an uncountable set of sets even mean?

In the case where S is a finite set, say S has k elements, there is an obvious choice for Σ and P which makes (S, Σ, P) into a probability space: Namely, you take Σ to be the power set on S, and you define P so that every subset A of S having ℓ elements has $P(A) = \ell/k$. Consider for example the coin toss

[4] In class I wrote "$\Omega \in \Sigma$" for property I of σ-algebras. Recall here that Ω is another common name for the sample space S. Once one has "closure under complements" property, then "$\Omega \in \Sigma$" is equivalent to "$\emptyset \in \Sigma$".

experiment where $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. Then, if Σ is the powerset on S, it will have $2^8 = 256$ elements. Now, consider the event "You roll at least one head during your three rolls". If you call this event A, then $A = \{HHH, HHT, HTH, HTT, THH, THT, TTH\}$, and so has $\ell = 7$ elements. If you choice of probability measure is as described above, then this event will have probability $P(A) = 7/8$.

More generally, if one has that S has k elements, the set of all probability measures on 2^S are constructed in the following way: Let $S = \{s_1, ..., s_k\}$, and start with probabilities $P(\{s_1\}) = a_1, P(\{s_2\}) = a_2, ..., P(\{s_k\}) = a_k$, where $1 = a_1 + \cdots + a_k$. This sum of probabilities condition is necessary since

$$1 \ = \ P(S) \ = \ P(\{s_1\}) + \cdots + P(\{s_k\}).$$

Now, it follows that $P(\{s_{i_1}, s_{i_2}, ..., s_{i_\ell}\}) = a_{i_1} + \cdots + a_{i_\ell}$. One can check that this indeed gives a probability measure on 2^S.

These more general measures are actually needed to describe certain experiments; for example, suppose you roll a loaded die where the probability of getting a 1 is 1/2, and the probability of getting 2,3,4,5,6 are each 1/10. Then, $S = \{1, 2, 3, 4, 5, 6\}$, and not all the elementary events $\{j\}$ have the same probability.

Before we move on to some subtleties, here are some basic properties of probability measures on a σ-algebra $\Sigma \subset 2^S$:

I. $P(\emptyset) = 0$. Since $S \cup \emptyset = S$, where S and $\emptyset$ are disjoint, we have by property II above that

$$1 \ = \ P(S) \ = \ P(S \cup \emptyset) \ = \ P(S) + P(\emptyset) \ = \ 1 + P(\emptyset)$$

implies $P(\emptyset) = 0$.

II. If $E \in \Sigma$, then $P(\overline{E}) = 1 - P(E)$. To see this we note that $\overline{E} \in \Sigma$ by properties of a σ-algebra. Then, since E and $\overline{E}$ are disjoint we deduce

$$1 \ = \ P(S) \ = \ P(E \cup \overline{E}) \ = \ P(E) + P(\overline{E}),$$

and the result follows.

III. If $A, B \in \Sigma$ with $A \subseteq B$, then $P(A) \leq P(B)$. To see this, observe that $B = A \cup (B \cap \overline{A})$; also, note that A and $B \cap \overline{A}$ are disjoint. Then,

$$P(B) \ = \ P(A \cup (B \cap \overline{A})) \ = \ P(A) + P(B \cap \overline{A}) \ \geq \ P(A),$$

whence the claim follows.

IV. If $A_1, ..., A_k \in \Sigma$ then

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) \ \leq \ P(A_1) + \cdots + P(A_k).$$

This is one of your homeworks, so I won't give the proof here. However, the idea of the proof is to first show $P(A \cup B) \leq P(A) + P(B)$ for events $A, B \in \Sigma$, which can be done by observing that $A \cup B = A \cup (B \cap \overline{A})$. Then, noting that these last two unioned sets are disjoint one can deduce...

V. If $A, B \in \Sigma$ then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. This is sometimes called "inclusion exclusion", and its name will become obvious when we do a more general version of it later on in the course. One can also regard this formula as a generalization of property IV. Since this property was another one of your homeworks, as with property IV, I will not bother to write down the proof.