SPLASH 2012, Game Theory: Set, Logic, and Probability Review

This summary defines a lot of jargon–that way, if I accidentally use any of these terms without definition during the class, it will (hopefully) be less confusing. I don't know how much will get used, but I figure it's best to just have it here. Don't worry if you don't know everything here, and feel free to stop me in class if something is confusing!

1 Logic notation

Logic notation ties in nicely with set notation, discussed below–so well that it's hard to separate the two. Some key notation:

- **Logical or ($\vee$):** $A \vee B$ is true if least one of A or B happened, and maybe both. Note that this is different than English usage, which tends to use "or" to mean "exclusive or": either A or B happens, but not both. In math, however, $A \vee B$ means that we know that A, B, or (A and B) happened.
- **Logical and ($\wedge$):** $A \wedge B$ is true only if both A and B happened.
- **Implication ($\implies$):** $\implies$ is the "implication symbol". $A \implies B$ translates to "if A happens, then B happens." For example, $Cat(x) \implies Animal(x)$ means that if x is a cat, it is also an animal. Note that implication only works in one direction: if x is an animal, it isn't necessarily a cat.
- **If and only if ($\iff$):** $\iff$, also written as "iff", represents "if and only if". It's basically implication in both directions: $A \iff B$ translates to $(A \iff B) \wedge (B \iff A)$. For example, our definition of equality uses $\iff$. $A = B \iff B = A$. That's why you can move variables to either side of the equation. For another example, $Cat(x) \iff Feliscatus(x)$. "Felis catus" is just the scientific name for a cat, so if x is a cat, it is of the species Felis catus.

2 Sets and Sequences

Set notation is incredibly useful. It provides a really fast and convenient way of writing down complex ideas. A **set** is an unordered collection of unique elements. We indicate that a collection of elements is a set by putting the elements inside curly braces. Sets can either be **finite** or **infinite**.

For example, $\{1, 2, 3\}$ is a set containing three elements. The set of all integers, the infinite set $\{\ldots -3, -2, -1, 0, 1, 2, \ldots\}$ has a special symbol to represent it, $\mathbb{Z}$. The set of all real numbers is not only infinite but continuous–there is no way even to create an ordered list of its elements. (If you're interested in this concept, the reals are an example of 'uncountably infinite sets'–you can look it up online). The reals contain all sorts of numbers, for example, integers like 17, fractions like 3/4, and "irrational numbers" like π. We represent the set of all reals with the special symbol $\mathbb{R}$. Typically we represent sets by a capital letter. For example, we can say that $A = \{1, 2, 3\}$, which means that A represents the collection of elements $\{1, 2, 3\}$.

Order doesn't matter for sets. For example, the set $\{1, 2, 3\}$ is **equivalent to** the set $\{2, 1, 3\}$, and both these sets are equivalent to the set $\{3, 1, 2\}$, etc. The ordering that the elements are listed in simply doesn't matter. '**Unique**' means that no element can be in a set twice. For example, $\{1, 2, 2, 3\}$ is NOT a set–although sometimes we abuse the notation and consider it to be equivalent to $\{1, 2, 3\}$.

Sometimes it's useful to define sets with additional qualifications on them. We do this by defining a set of "potential elements" with an additional qualification. The representation looks like $\{x|\text{qualification for } x\}$. That bar, $|$, is read out as "such that". As an example, $\{x|x \in \mathbb{Z}, x/2 \in \mathbb{Z}\}$ is the set of integers (the first qualification, $x \in \mathbb{Z}$) such that $x/2$ is also an integer (the second qualification, $x/2 \in \mathbb{Z}$). That's just the set of all even integers.

There's also special notation for a set which is totally empty, $\varnothing$.

A **sequence** is an ordered collection of elements. We indicate a sequence with parentheses, for example, $(1, 2, 3)$. Most of the sequences we talk about are infinite, so we typically write the first few elements and then just use $\ldots$ to indicate the remainder, for example, $(1, 2, 3, \ldots)$ represents the infinite set where we start counting each integer, starting at 1. The ordering of a sequence is important: $(1, 2, 3)$ is NOT equivalent to $(2, 1, 3)$. The elements in a sequence are also not necessarily unique: for example, $(1, 2, 3, 3, 3)$ is a perfectly valid sequence, and it isn't equivalent to $(1, 2, 3, 3)$.

We use a lot of the same notation to describe both sets and sequences, given below:

- **Membership ($\in$):** An element is a **member** of a set if it is in the set. For example, if $A = \{1, 2, 3\}$, then 1 is a member of A. The shorthand for this is $1 \in A$. We translate $\in$ as either "is a member of" or "is in".

- **For all ($\forall$):** $\forall$ just gives us convenient shorthand for listing all of the elements of a set. Let $A = \{1, 2, 3\}$. If $f(x) = 1$ $\forall x \in A$, then $f(1) = 1$, $f(2) = 1$, and $f(3) = 1$. For an example of usage, let $Cloudy(d)$ represent the event that day d is cloudy. Let W represent the set containing all the days of last week. Then for it to be truly miserable, it has to have been cloudy every day last week. Writing that in math, we would say that $Cloudy(d)\ \forall d \in W \implies Miserable$.

- **There exists ($\exists$):** Existence is kind of the inverse of $\forall$. For example, if $\exists x \in A$ such that $x \in B$, then there has to be at least one element (call it y) in set A such that y is also in set B. This is really different than the concept of $\forall$. If $\forall x \in A, x \in B$, every single element in A would have to also be in B. If $\exists x \in A, x \in B$, then all we know is that at least one element in A is also in B.

- **Subset ($\subseteq$):** A set A is a subset of a set B if every element of A is also in B. In mathematical notation, $A \subseteq B$ if $\forall x \in A, x \in B$. In other words, every element in A is also in B. The notation I used is actually called "nonstrict subset", meaning that it could be the case that $A = B$ and the two sets share exactly the same set of elements. There is also "strict subset notation", $A \subset B$. If $A \subset B$, then $A \subseteq B$, but $A \neq B$.

- **Cardinality ($|\ |$):** Cardinality is just a fancy word for the size of the set. $|A|$ is just the number of elements in A. For example, if $A = \{1, 2, 3\}$, $B = \{a, b, c, d, e, f\}$, and $C = \mathbb{Z}$, then $|A| = 3$, $|B| = 6$, and $|C| = \infty$.

- **Union ($\cup$):** The union of two sets, A and B, is just "joining" the two sets to create a new set of elements which are in either A or B. For example, if $A = \{1, 2, 3, 4\}$ and $B = \{3, 4, 5, 6\}$, then $A \cup B = \{1, 2, 3, 4, 5, 6\}$. Note that since elements of a set are unique, even though 3 and 4 appear in both. Notice how much $\cup$ looks like the logical or, $\vee$. For x to be in $A \cup B$, $x \in A \vee x \in B$: $A \cup B = \{x | x \in A \vee x \in B\}$.

- **Intersection ($\cap$):** The intersection of A and B is the set of elements in both A and B. Using our same definitions above, $A \cap B = \{3, 4\}$. Note how much $\cap$ looks like the logical and, $\wedge$. For x to be in $A \cap B$, $x \in A \wedge x \in B$: $A \cap B = \{x | x \in A \wedge x \in B\}$. Two sets are **disjoint** if their intersection is empty. For example, $C = \{1, 2, 3\}$ and $D = \{4, 5, 6\}$ are disjoint because $C \cap D = \varnothing$.

- **Cross product ($\times$):** It can sometimes get really tedious to write out elements of a set, and the cross product provides a handy notation if we want to create a set whose elements are created by taking one element out of each of a number of different sets. Let's say you're playing a game where you flip a coin (it can land on either heads or tails) and also roll a die. What is the set of possible outcomes of the game? Let $C = \{H, T\}$ be a set representing the possible outcomes to a single coin flip. Let $D = \{1, 2, 3, 4, 5, 6\}$ be a set representing all possible outcomes to the roll of a single die. Then $C \times D$ is the set of all combinations that can be created by taking one element out of C and one element out of D. In this case,

  $$C \times D = \{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}.$$

  Note that the cross-product enforces an order: $C \times D$ is not the same thing as $D \times C$.

  $$D \times C = \{1H, 2H, 3H, 4H, 5H, 6H, 1T, 2T, 3T, 4T, 5T, 6T\}.$$

  The cross product is really useful, since sample spaces blow up *exponentially* with the number of different events. For example, if we flip 10 coins in a row, we have $C \times C \times C \times C \times C \times C \times C \times C \times C \times C$ which has $2^{10} = 1024$ different elements!

- **Summation ($\sum$):** $\sum$ is just handy shorthand to represent a bunch of additions. For example, let's say we have a sequence, $(x_1, x_2, x_3, x_4, x_5)$. Then $\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5$. i is our index (current place) in the elements. The stuff under the summation sign, $i = 1$, represents where the index starts at. The stuff on the top represents where the sum ends. Sometimes we can get even lazier. Say $S = (x_1, x_2, x_3, x_4, x_5)$. Then $\sum S = \sum_{i=1}^{5} x_i$, the sum of all elements in the sequence (or set).

- **Product ($\prod$):** This is identical notation to summation, except that it represents multiplication. Using the sequence above, $\prod_{i=1}^{5} x_i = x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5$–the number obtained by multiplying all of the elements together.

3 Probability

Probability is a way of measuring uncertainty. We use the ideas behind it all the time in everyday life, but it's amazing how often human intuition about probability is just plain wrong, as we'll see during the class!

Let's start by describing the probability of particular **outcomes** to an experiment. The **sample space** of our experiment is the set of all possible outcomes. The sample space can be either **discrete** or **continuous**.

Some examples:

- The sample space describing the outcome of the roll of a die is discrete and finite: $\{1, 2, 3, 4, 5, 6\}$
- The sample space describing the flipping of two coins in a row is discrete and finite: $\{HH, HT, TH, TT\}$
- The sample space describing choosing one integer from the set of integers is discrete, but infinite: $\{\ldots -2, -1, 0, 1, 2, 3, \ldots\}$
- The sample space describing a random choice of a real number between 0 and 1 is continuous.
- The sample space describing the temperature on a particular day is also continuous: it can be any number, say, between around -50 and 150 Fahrenheit.

It's useful to construct functions over events–for example, the event that, if when we roll two dice, the sum of the pips are even. These functions are called "**random variables**", which is a little odd, since they are neither random (they are functions over events) nor variables (again, they're functions!). They are typically denoted as capital letters. Random variables are always numeric. They provide a way of converting, non numeric sample spaces to something easier to work with. For example, say we are pulling socks out of a drawer, and the possible colors are white, black, and grey. The sample space is $\{whitesocks, greysocks, blacksocks\}$. We can define a random variable, say, R, which takes on the value 1 if we pull out white socks, 2 if we pull out the black socks, and 3 if we pull out the grey socks. Then we can talk about $\mathbb{P}(R \leq 2)$, for example. We then define a **probability distribution** over this sample space. We'll begin by defining a probability distribution over a discrete sample space. The probability distribution has to obey a particular set of rules. Let Ω be the sample space (this Greek letter, Omega, is the standard symbol, probably related to that expression, "the alpha and omega", meaning "everything").

- Each outcome (which I'll also call an **event**) in the sample space must have a probability defined over it, which we write as $\mathbb{P}(E)$. Probabilities can never be negative: $\forall E \in \Omega, \mathbb{P}(E) \geq 0$.
- If Y contains all of the events in X, then Y is at least as probable as X: $X \subseteq Y \implies \mathbb{P}(X) \leq \mathbb{P}(Y)$
- If X and Y are disjoint, then $\mathbb{P}(X \cup Y) = \mathbb{P}(X) + \mathbb{P}(Y)$.
- The probability of the entire sample space must be equal to 1: $\mathbb{P}(\Omega) = 1$. To put it another way, if we have a set of exclusive events S such that $\forall X, Y \in S, \mathbb{P}(X \cap Y) = 0$ but the union over all the events in S covers the entire sample space, $\bigcup_{X \in S} X = \Omega$, then the sum of all these events must equal 1: $\sum_{X \in S} \mathbb{P}(X) = 1$. You've actually seen this a lot: "percent" is just a probability multiplied by 100. This is why it makes absolutely no mathematical sense to talk about putting in 110%.

For each random variable X, we can define another random variable, called the **complement** of X, $\bar{X} = \Omega - X$. In other words, $\bar{X}$ is the probability that X does not occur. Therefore $\mathbb{P}(\bar{X}) = 1 - \mathbb{P}(X)$.

If you know some calculus, then similar statements hold for continuous distributions: just replace sums with integrals.

4 Conditional Probability

Conditioning

When we have a big sample space and somewhat complex events, then learn some small fact, it can change the probability distribution over the sample space. This is called **conditioning** on an observation.

For example, let's say that you are trying to get up in the morning and the light in your room is broken. You already put on one green sock, and let's say you're trying to find the other in your drawer. You have a bunch of mismatched socks in your drawer: two charcoal grey, four white socks, one blue sock, and one green sock. If you reach in and pull out one sock, what is the probability that you pull out the green one? Well, there are 8 different socks, so, $\Omega = \{White, Blue, Green, Charcoal\}$. Let's start by assuming that the distribution is **uniform**: you have equal probability of pulling out any of the socks in the drawer. Note that this doesn't mean the probability of pulling out any particular color of sock is uniform. Let G be a random variable which takes on the value 1 if the sock is green, B be an RV which is 1 if the sock is blue, C be the RV that is 1 if the sock is charcoal grey, and W be an RV that takes on the value 1 if the socks are white. Then $\mathbb{P}(G) = \frac{1}{8}$ because one of the 8 socks is green. The probability of pulling out a blue sock is also $\mathbb{P}(B) = \frac{1}{8}$. Since you have two grey socks, the probability of the sock being grey is $\mathbb{P}(C) = \frac{2}{8} = \frac{1}{4}$, and the probability of the sock being white is $\mathbb{P}(W) = \frac{1}{2}$. Now, your room is dim, so it's hard to actually see the color of the sock, but you can tell that it's not white. That means we now know that given our new knowledge, $\mathbb{P}(W) = 0$. We **observed** the event $\bar{W}$. There are two ways to look at this.

First, we can think of this as meaning that we have a brand new sample space, call it $\Omega_{\bar{W}}$, where $\Omega_{\bar{W}} = \{Blue, Green, Charcoal\}$.

Another way of thinking of this as staying in the same space, but we condition on the new knowledge. We write this using conditional probability as $\mathbb{P}_{\Omega}\left[G|\bar{W}\right]$. This is said as "the probability the sock is green given that it is not white." It is the same thing as the probability that the sock is green in our new "non-white-sock" sample space: $\mathbb{P}_{\Omega}\left[G|\bar{W}\right] = \mathbb{P}_{\Omega_{\bar{W}}}[G]$. Basically, we readjust all of the probabilities given our new knowledge (in this case, it means crossing out the event of the sock being white) and then have to re-regularize the probabilities so that they sum to one. This gives us the definition of conditional probability:

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

To see this in our sock example, since we now know that the sock can't be white, there are 4 possible socks that have nonzero probability, only one of which is green. Therefore we have $\mathbb{P}\left(G|\bar{W}\right) = \frac{1}{4}$. We can get the same thing out of the definition. Since a sock can only be one color, and if the sock is green, it can't be white, we have that $\mathbb{P}\left(G \cap \bar{W}\right) = \frac{1}{8}$. The probability of a sock not being white is $\frac{1}{2}$. Therefore using our formula, $\mathbb{P}\left(G|\bar{W}\right) = \frac{\mathbb{P}\left(G\cap\bar{W}\right)}{\mathbb{P}\left(\bar{W}\right)} = \frac{1/8}{1/2} = \frac{1}{4}$.

Independence

An important property related to conditional probability is called **independence**. If two RV's A and B are independent, intuitively, it means that knowing that A occurred doesn't change the probability that B will occur. Conditioning on A doesn't change the probability of B. There is a special symbol for independence. If A is independent on B, we write $A \perp B$. (If, after conditioning on another event C, A is independent of B, we write $A \perp B|C$.)

Here are some equivalent definitions of independence:

$$A \perp B \iff \mathbb{P}(A|B) = \mathbb{P}(A) \iff \mathbb{P}(B|A) = \mathbb{P}(B) \iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$$

Bayes' Theorem

Using this definition, we can come up with a really important principle called **Bayes' Theorem**. We derive it as follows: $\mathbb{P}(B|A) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$. So Bayes' Theorem says:

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

Bayes' theorem is really important: it is often used to condition on rare events given evidence, for example, in medical tests. People make mistakes using Bayes' Theorem all the time. In a British study, they found that *about* 90% *of doctors in Britain trying to use it for diagnosis made mistakes!*

For a weird example, let's say that you think your friend might be an alien from outer space. The probability of that is really, really small. Let A be the RV that is 1 if your friend is an alien. Then $\mathbb{P}(A = 1) = .000001$. This means $\mathbb{P}(G|A) = 0.25$. Now let's say you discover that your friend's skin glows in the dark. Let's also say that $\frac{1}{4}$ of the aliens that come down to earth have glowing skin (don't ask how someone got that statistic!) But let's say there are a few different ways your skin could glow green: a weird mutation, an industrial accident, or being alien. Let's say $\mathbb{P}(G) = .000005$. So given that your friend's skin glows, what is the probability that she is an alien? $\mathbb{P}(A|G) = \frac{\mathbb{P}(A\cap G)}{\mathbb{P}(G)} = \frac{\mathbb{P}(G|A)\mathbb{P}(A)}{\mathbb{P}(G)} = \frac{.000001 \cdot .25}{.000005} = 0.05$, which means that given that your friend's skin glows green, there is a 5% chance she is an alien!

For a more realistic example that doctors honestly tend to mix up, let's say that there is a rare and highly dangerous genetic disorder which strikes .00001 percent of the population. Fortunately, there's a pretty accurate genetic test for it–it is accurate 99.9% of the time. Tests for it tend to come in batches, and since your doctor was already testing another patient, she offers to test you too. You agree–why not? You are shocked to hear the test came back positive. Your doctor now wants to perform a surgery to try to mitigate the effects disease, arguing that even if it isn't certain, it is very likely you have the disease?. Before you agree, you try to weigh the odds. What's the probability that you have the disease?

First, let D be the RV representing that you have the disease, and let T be the RV representing that the test is positive. Then $\mathbb{P}(D) = .00001$. Since we know that the test makes mistakes–and apparently a "false positive" is as likely as a "false negative", we have that $\mathbb{P}(T|D) = 0.999$ (the probability the test was positive given that you have the disease) and $\mathbb{P}(\bar{T}|\bar{D}) = .999$ (the probability that the test is negative given that you don't have the disease). But what we care about is $\mathbb{P}(D|T)$–the probability you have the disease given the test is positive. We can use Bayes' Theorem:

$$\mathbb{P}(D|T) = \frac{\mathbb{P}(T|D)\,\mathbb{P}(D)}{\mathbb{P}(T)} = \frac{\mathbb{P}(T|D)\,\mathbb{P}(D)}{\mathbb{P}(D,T) + \mathbb{P}(\bar{D},T)} = \frac{\mathbb{P}(T|D)\,\mathbb{P}(D)}{\mathbb{P}(T|D)\,\mathbb{P}(D) + \mathbb{P}(T|\bar{D})\,\mathbb{P}(\bar{D})} = \frac{0.999 \cdot .00001}{0.999 \cdot .00001 + (.001)(.99999)} \approx 0.00989$$

So stay away from the surgery, it's over 99% probable you don't have the disease! What happened? The disease is so rare that even though the test makes very few mistakes, it has a lot of opportunity to make mistakes that find the disease where there isn't one.

5 Expectation

This is the most important probability concept we use. Expectation is essentially the "weighted average" of a random variable over a probability distribution. Let X be a discrete random variable. Then the expectation of X is defined as

$$\mathbb{E}[X] = \sum_x x\mathbb{P}(x)$$

Although it may look a little nasty, it has a nice intuitive interpretation. Let's say you're playing a dice game where you roll a single die. If the die lands on an odd number, you have to give 1 dollar to your opponent. If the die lands on a 6, you get 6 dollars from your opponent. Should you play the game or not? Let's look at the probabilities. Let X be a random variable representing the result of the die roll, and let Y be the RV representing your earnings. Then $\mathbb{P}(X = 6) = \frac{1}{6}$, so $\mathbb{P}(Y = 6) = \frac{1}{6}$ –you'll earn 6 dollars. $\mathbb{P}(X \in \{1,3,5\}) = \frac{1}{2}$, so $\mathbb{P}(Y = -1) = \frac{1}{2}$ –you'll lose a dollar. $\mathbb{P}(X \in \{2,4\}) = \frac{1}{3}$, so $\mathbb{P}(Y = 0) = \frac{1}{3}$. How much do you expect to earn? Well, we'd expect $6\mathbb{P}(Y = 6) + (-1)\mathbb{P}(Y = -1) = 6(\frac{1}{6}) + (-1)(\frac{1}{2}) = 1 - \frac{1}{2} = \frac{1}{2}$–in expectation, you'll expect to win money out of the game, so it's definitely worth playing!

Simple expectation doesn't capture everything because humans' value for money isn't linear. For example, let's say you enter a lottery where one person wins all of the money paid in. With probability $\frac{1}{1000000}$, say, you win the jackpot; otherwise you pay 1 dollar. Then even though the expectation of your winnings here is 0, you may very well not want to play. Your value for the millionth dollar in the jackpot isn't the same as the value for the dollar you're about to lose; when you have a lot of money, you don't notice a dollar here and there, but when you don't have much money, you value what you have more. So even though the monetary expectation is 0, meaning you break even in expectation, you'd probably do better not to play. This touches on the concept of **utility functions**, which we'll be discussing during the class!

6 More Resources

If you want another probability resource, there are lots of free online textbooks. here are a few:

- `http://www.math.dartmouth.edu/~prob/prob/prob.pdf`
- `http://cs.wellesley.edu/~cs249B/math/Probability/CS249BProbabilityPrimer.pdf`
- `http://aleph0.clarku.edu/~djoyce/ma217/book-5-17-03.pdf`