Lecture 10: Everything Else

This is the tenth week of the Mathematics Subject Test GRE prep course; here, we quickly review a handful of useful concepts from the fields of **probability**, **combinatorics**, and **set theory**!

As always, each of these fields are things you could spend years studying; we present here a few very small slices of each topic that are particularly key to each.

1   A Crash Course in Set Theory

Sets are, in a sense, what mathematics are founded on. For the GRE, much of the intricate parts of set theory (the axioms of set theory, the axiom of choice, etc.) aren't particularly relevant; instead, most of what we need to review is simply the **language** and **notation** for set theory.

Definition. A set S is simply any collection of elements. We denote a set S by its elements, which we enclose in a set of parentheses. For example, the set of female characters in **Measure for Measure** is

$$\{\text{Isabella, Mariana, Juliet, Francisca}\}.$$

Another way to describe a set is not by listing its elements, but by listing its properties. For example, the even integers greater than π can be described as follows:

$$\{x \mid \pi < x, x \in \mathbb{Z}, 2 \text{ divides } x\}.$$

Definition. If we have two sets A, B, we write $A \subseteq B$ if every element of A is an element of B. For example, $\mathbb{Z} \subseteq \mathbb{R}$.

Definition. A specific set that we often care about is the **empty set**, $\emptyset$ i.e. the set containing no elements. One particular quirk of the empty set is that any statement of the form $\forall x \in \emptyset \ldots$ will always be vacuously true, as it is impossible to disprove (as we disprove $\forall$ claims by using $\exists$ quantifiers!) For example, the statement "every element of the empty set is delicious" is true. Dumb, but true!

Some other frequently-occurring sets are the **open** intervals $(a, b) = \{x \mid x \in \mathbb{R}, a < x < b\}$ and **closed** intervals $[a, b] = \{x \mid x \in \mathbb{R}, a \le x \le b\}$.

Definition. Given two sets A, B, we can form several other particularly useful sets:

- The **difference** of A and B, denoted $A - B$ or $A \backslash B$. This is the set $\{x \mid x \in A, x \notin B\}$.

- The **intersection** of A and B, denoted $A \cap B$, is the set $\{x \mid x \in A \text{ and } x \in B\}$.

- The **union** of A and B, denoted $A \cup B$, is the set $\{x \mid x \in A \text{ or } x \in B, \text{ or possibly both.}\}$.

- The **cartesian product** of A and B, denoted $A \times B$, is the set of all ordered pairs of the form (a, b); that is, $\{(a, b) \mid a \in A, b \in B\}$.

- Sometimes, we will have some larger set B (like $\mathbb{R}$) out of which we will be picking some subset A (like $[0, 1]$.) In this case, we can form the **complement** of A with respect to B, namely $A^c = \{b \in B \mid b \notin A\}$.

With the concept of a set defined, we can define functions as well:

Definition. A **function** f with domain A and codomain B, formally speaking, is a collection of pairs (a, b), with $a \in A$ and $b \in B$, such that there is exactly one pair (a, b) for every $a \in A$. More informally, a function $f : A \to B$ is just a map which takes each element in A to some element of B.

Examples.

- $f : \mathbb{Z} \to \mathbb{N}$ given by $f(n) = 2|n| + 1$ is a function.

- $g : \mathbb{N} \to \mathbb{N}$ given by $f(n) = 2|n| + 1$ is also a function. It is in fact a different function than f, because it has a different domain!

- The function h depicted below by the three arrows is a function, with domain $\{1, \lambda, \varphi\}$ and codomain $\{24, \gamma, \text{Batman}\}$:
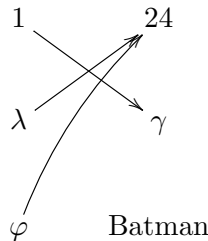

This may seem like a silly example, but it's illustrative of one key concept: functions are just **maps between sets!** Often, people fall into the trap of assuming that functions have to have some nice "closed form" like $x^3 - \sin(x)$ or something, but that's not true! Often, functions are either defined piecewise, or have special cases, or are generally fairly ugly/awful things; in these cases, the best way to think of them is just as a collection of arrows from one set to another, like we just did above.

Functions have several convenient properties:

Definition. We call a function f **injective** if it never hits the same point twice – i.e. for every $b \in B$, there is **at most one** $a \in A$ such that $f(a) = b$.

Example. The function h from before is not injective, as it sends both λ and φ to 24:

1   24

λ   γ

φ   Batman
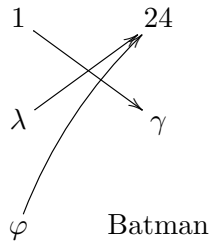
However, if we add a new element π to our codomain, and make φ map to π, our function is now injective, as no two elements in the domain are sent to the same place:

1   24

λ   γ

φ   Batman

π
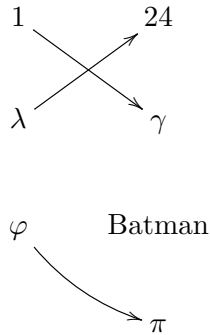
Definition. We call a function f **surjective** if it hits every single point in its codomain – i.e. if for every $b \in B$, there is **at least one** $a \in A$ such that $f(a) = b$.

Alternately: define the **image** of a function as the collection of all points that it maps to. That is, for a function $f : A \to B$, define the image of f, denoted $f(A)$, as the set $\{b \in B \mid \exists a \in A \text{ such that } f(a) = b\}$.

Then a surjective function is any map whose image is equal to its codomain: i.e. $f : A \to B$ is surjective if and only if $f(A) = B$.

Example. The function h from before is not injective, as it doesn't send anything to Batman:

1   24

λ   γ
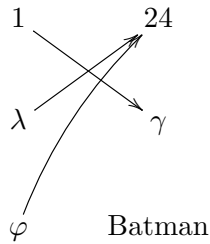
φ   Batman

However, if we add a new element ρ to our domain, and make ρ map to Batman, our function is now surjective, as it hits all of the elements in its codomain:

1   24

λ   γ

φ   Batman

ρ
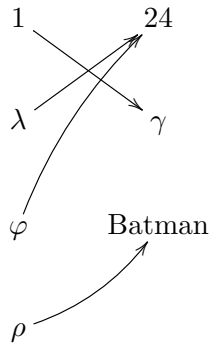
Definition. A function is called **bijective** if it is injective and surjective.

Definition. We say that two sets A, B are the same size (formally, we say that they are of the same **cardinality**,) and write $|A| = |B|$, if and only if there is a bijection $f : A \to B$.

Not all sets are the same size:

Observation. If A, B are a pair of finite sets that contain different numbers of elements, then $|A| \neq |B|$.

If A is a finite set and B is infinite, then $|A| \neq |B|$.

If A is an infinite set such that there is a bijection $A \to \mathbb{N}$, call A countable. If A is countable and B is a set that has a bijection to $\mathbb{R}$, then $|A| \neq |B|$.

We can use this notion of size to make some more definitions:

Definition. We say that $|A| \leq |B|$ if and only if there is an injection $f : A \to B$. Similarly, we say that $|A| \geq |B|$ if and only if there is a surjection $f : A \to B$.

This motivates the following theorem:

Theorem. (Cantor-Schröder-Bernstein): Suppose that A, B are two sets such that there are injective functions $f : A \to B, g : B \to A$. Then $|A| = |B|$; i.e. there is some bijection $h : A \to B$.

2 A Crash Course in Combinatorics

Combinatorics, very loosely speaking, is the art of how to **count things**. For the GRE, a handful of fairly simple techniques will come in handy:

- (**Multiplication principle.**) Suppose that you have a set A, each element $\vec{a}$ of which can be broken up into n ordered pieces $(a_1, \ldots a_n)$. Suppose furthermore that the i-th piece has k_i total possible states for each i, and that our choices for the i-th stage do not interact with our choices for any other stage. Then there are

$$k_1 \cdot k_2 \cdot \ldots \cdot k_n = \prod_{i=1}^{n} k_i$$

total elements in A.

To give an example, consider the following problem:

Problem. Suppose that we have n friends and k different kinds of postcards (with arbitrarily many postcards of each kind.) In how many ways can we mail out all of our postcards to our friends?

A valid "way" to mail postcards to friends is some way to assign each friend to a postcard, so that each friend is assigned to at least at least one postcard (because we're mailing each of our friends a postcard) and no friend is assigned to two different postcards at the same time. In other words, a "way" to mail postcards is just a function from the set[1] $[n] = \{1, 2, 3, \ldots n\}$ of postcards to our set $[k] = \{1, 2, 3, \ldots k\}$ of friends!

In other words, we want to find the size of the following set:

$$A = \Big\{ \text{all of the functions that map } [n] \text{ to } [k] \Big\}.$$

We can do this! Think about how any function $f : [n] \to [k]$ is constructed. For each value in $[n] = \{1, 2, \ldots n\}$, we have to pick exactly one value from $[k]$. Doing this for each value in $[n]$ completely determines our function; furthermore, any two functions f, g are different if and only if there is some value $m \in [n]$ at which we made a different choice (i.e. where $f(m) \neq g(m)$.)

$$\underbrace{\boxed{\text{k choices}} \cdot \boxed{\text{k choices}} \cdot \ldots \cdot \boxed{\text{k choices}}}_{n \text{ total slots}}$$

Consequently, we have

$$\underbrace{k \cdot k \cdot \ldots \cdot k}_{n} = k^n$$

total ways in which we can construct distinct functions. This gives us this answer k^n to our problem!

- (**Summation principle**.) Suppose that you have a set A that you can write as the **union**[2] of several smaller disjoint[3] sets $A_1, \ldots A_n$.

  Then the number of elements in A is just the summed number of elements in the A_i sets. If we let $|S|$ denote the number of elements in a set S, then we can express this in a formula:

  $$|A| = |A_1| + |A_2| + \ldots + |A_n|.$$

  We work one simple example:

---

[1] Some useful notation: $[n]$ denotes the collection of all integers from 1 to n, i.e. $\{1, 2, \ldots n\}$.

[2] Given two sets A, B, we denote their **union**, $A \cup B$, as the set containing all of the elements in either A or B, or both. For example, $\{2\} \cup \{\text{lemur}\} = \{2, \text{lemur}\}$, while $\{1, \alpha\} \cup \{\alpha, \text{lemur}\} = \{1, \alpha, \text{lemur}\}$.

[3] Sets are called **disjoint** if they haven no elements in common. For example, $\{2\}$ and $\{\text{lemur}\}$ are disjoint, while $\{1, \alpha\}$ and $\{\alpha, \text{lemur}\}$ are not disjoint.

Question 1. *Pizzas! Specifically, suppose Pizza My Heart (a local chain/ great pizza place) has the following deal on pizzas: for 7$, you can get a pizza with any two different vegetable toppings, or any one meat topping. There are m meat choices and v vegetable choices. As well, with any pizza you can pick one of c cheese choices.*

How many different kinds of pizza are covered by this sale?

Using the summation principle, we can break our pizzas into two types: pizzas with one meat topping, or pizzas with two vegetable toppings.

For the meat pizzas, we have $m \cdot c$ possible pizzas, by the multiplication principle (we pick one of m meats and one of c cheeses.)

For the vegetable pizzas, we have $\binom{v}{2} \cdot c$ possible pizzas (we pick two different vegetables out of v vegetable choices, and the order doesn't matter in which we choose them; we also choose one of c cheeses.)

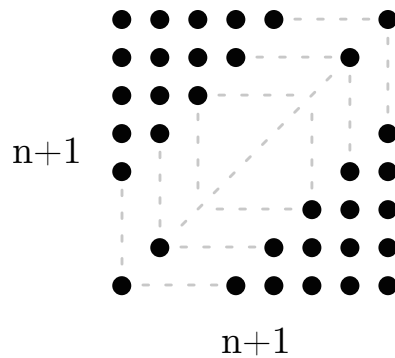Therefore, in total, we have $c \cdot \left(m + \binom{v}{2}\right)$ possible pizzas!

- (**Double-counting principle**.) Suppose that you have a set A, and two different expressions that count the number of elements in A. Then those two expressions are equal.

Again, we work a simple example:

Question 2. *Without using induction, prove the following equality:*
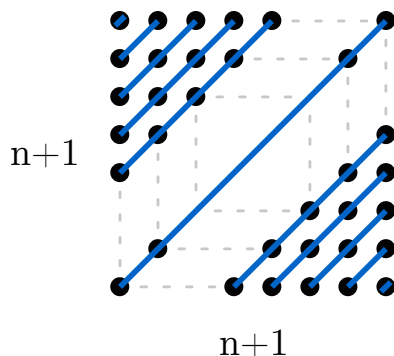
$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

First, make a $(n+1) \times (n+1)$ grid of dots:

n+1

n+1

How many dots are in this grid? On one hand, the answer is easy to calculate: it's $(n+1) \cdot (n+1) = n^2 + 2n + 1$.

On the other hand, suppose that we group dots by the following diagonal lines:

n+1

n+1

The number of dots in the top-left line is just one; the number in the line directly beneath that line is two, the number directly beneath that line is three, and so on/so forth until we get to the line containing the bottom-left and top-right corners, which contains $n+1$ dots. From there, as we keep moving right, our lines go down by one in size each time until we get to the line containing only the bottom-right corner, which again has just one point.

So, if we use the summation principle, we have that there are

$$1 + 2 + 3 + \ldots + (n-1) + n + (n+1) + n + (n-1) + \ldots + 3 + 2 + 1$$

points in total.

Therefore, by our double-counting principle, we have just shown that

$$n^2 + 2n + 1 = 1 + 2 + 3 + \ldots + (n-1) + n + (n+1) + n + (n-1) + \ldots + 3 + 2 + 1.$$

Rearranging the right-hand side using summation notation lets us express this as

$$n^2 + 2n + 1 = (n+1) + 2\sum_{i=1}^{n} i;$$

subtracting $n+1$ from both sides and dividing by 2 gives us finally

$$\frac{n^2 + n}{2} = \sum_{i=1}^{n} i,$$

which is our claim!

- (Pigeonhole principle, simple version): Suppose that $kn + 1$ pigeons are placed into n pigeonholes. Then some hole has at least $k + 1$ pigeons in it. (In general, replace "pigeons" and "pigeonholes" with any collection of objects that you're placing in various buckets.)

  We look at an example:

  Question 3. *Suppose that "friendship" is[4] a symmetric relation: i.e. that whenever a person A is friends with a person B, B is also friends with A. Also, suppose that you are never friends with yourself[5] (i.e. that friendship is antireflexive.)*

---

[4]Magic!

[5]Just for this problem. Be friends with yourself in real life.

Then, in any set S of greater than two people, there are at least two people with the same number of friends in S.

Let $|S| = n$. Then every person in S has between 0 and $n-1$ friends in S. Also notice that we can never simultaneously have one person with 0 friends and one person with $n - 1$ friends at the same time, because if someone has $n - 1$ friends in S, they must be friends with everyone besides themselves.

Therefore, each person has at most $n - 1$ possible numbers of friends, and there are n people total: by the pigeonhole principle, if we think of people as the "pigeons" and group them by their numbers of friends (i.e. the "pigeonholes" are this grouping by numbers of friends,) there must be some pair of people whose friendship numbers are equal.

Some sorts of sets are very frequently counted:

- If we have a set of n objects, there are $n! = n \cdot (n-1) \cdot \ldots \cdot 1$ many ways to **order** this set. For example, the set $\{a, b, c\}$ has $3! = 6$ orderings:

$$abc, acb, bac, bca, cab, cba.$$

- Suppose that we have a set of n objects, and we want to pick k of them without repetition in order. Then there are $n \cdot (n-1) \cdot \ldots \cdot (n - (k+1))$ many ways to choose them: we have n choices for the first, $n-1$ for the second, and so on/so forth until our k-th choice (for which we have $n - (k+1)$ choices.) We can alternately express this as $\frac{n!}{(n-k)!}$; you can see this algebraically by dividing $n!$ by $k!$, or conceptually by thinking our our choice process as actually ordering all n elements (the $n!$ in our fraction) and then forgetting about the ordering on all of the elements after the first k, as we didn't pick them (this divides by $(n-k)!$.)

- Suppose that we have a set of n objects, and we want to pick k of them without repetition and without caring about the order in which we pick these k elements. Then there are $\frac{n!}{k!(n-k)!}$ many ways for this to happen. We denote this number as the **binomial coefficient** $\binom{n}{k}$.

- Finally, suppose that we have a set of n objects, and we want to pick k of them, where we can pick an element multiple times (i.e. with repetition.) Then there are n^k many ways to do this, by our multiplication principle from before.

3   A Crash Course in Probability

We give the basics of probability here.

Definition. A **finite probability space** consists of two things:

- A finite set Ω.

- A **measure** Pr on Ω, such that $Pr(\Omega) = 1$. In case you haven't seen this before, saying that Pr is a measure is a way of saying that Pr is a function $\mathcal{P}(\Omega) \to \mathbb{R}^+$, such that the following properties are satisfied:

  - $Pr(\emptyset) = 0$.

  - For any collection $\{X_i\}_{i=1}^{\infty}$ of subsets of Ω, $Pr\left(\bigcup_{i=1}^{\infty} X_i\right) \leq \sum_{i=1}^{n} Pr(X_i)$.

  - For any collection $\{X_i\}_{i=1}^{\infty}$ of **disjoint** subsets of Ω, $Pr\left(\bigcup_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{n} Pr(X_i)$.

For a **general** probability space, i.e. one that may not be finite, the definition is almost completely the same: the only difference is that Ω is not restricted to be finite, while Pr becomes a function defined only on the "measurable" subsets of Ω. (For the GRE, you can probably assume that any set you run into is "measurable." There are some pathological constructions in set theory that can be nonmeasurable; talk to me to learn more about these!)

For example, one probability distribution on $\Omega = \{1, 2, 3, 4, 5, 6\}$ could be the distribution that believes that $Pr(\{i\}) = 1/6$ for each individual i, and more generally that $Pr(S) = |S|/6$ for any subset S of Ω. In this sense, this probability distribution is capturing the idea of rolling a fair six-sided die, and seeing what comes up.

This sort of "fair" distribution has a name: namely, the **uniform** distribution!

Definition. The **uniform** distribution on a finite space Ω is the probability space that assigns the measure $|S|/|\Omega|$ to every subset S of Ω. In a sense, this measure thinks that any two elements in Ω are "equally likely;" think about why this is true!

We have some useful notation and language for working with probability spaces:

Definition. An **event** S is just any subset of a probability space. For example, in the six-sided die probability distribution discussed earlier, the set $\{2, 4, 6\}$ is an event; you can think of this as the event where our die comes up as an even number. The probability of an event S occurring is just $Pr(S)$; i.e. the probability that our die when rolled is even is just $Pr(\{2, 4, 6\}) = 3/6 = 1/2$, as expected.

Notice that by definition, as Pr is a measure, for any two events A, B, we always have $Pr(A \cup B) \leq Pr(A) + Pr(B)$. In other words, given two events A, B, the probability of either A or B happening (or both!) is at most the probability that A happens, plus the probability that B happens.

Definition. A real-valued **random variable** X on a probability space Ω is simply any function $\Omega \to \mathbb{R}$.

Given any random variable X¡ we can talk about the **expected value** of X; that is, the "average value" of X on Ω, where we use Pr to give ourselves a good notion of what "average" should mean. Formally, we define this as the following sum:

$$\sum_{\omega in \Omega} Pr(\omega) \cdot X(\omega).$$

For example, consider our six-sided die probability space again, and the random variable X defined by $X(i) = i$ (in other words, X is the random variable that outputs the top face of the die when we roll it.)

The expected value of X would be

$$\sum_{\omega in \Omega} Pr(\omega) \cdot X(\omega) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{21}{6} = \frac{7}{2}.$$

In other words, rolling a fair six-sided die once yields an average face value of 3.5.

Definition. Given a random variable X, if $\mu = \mathbb{E}(X)$, the **variance** $\sigma^2(X)$ of X is just $\mathbb{E}((X - \mu)^2)$. This can also be expressed as $\mathbb{E}(X^2) - (\mathbb{E}(X))^2$, via some simple algebraic manipulations.

The **standard deviation** of X, $\sigma(X)$, is just the square root of the variance.

Definition. Given a random variable $X : \Omega \to \mathbb{R}$ on a probability space Ω, we can define the **density function** for X, denoted $F_X(t)$, as

$$F_X(t) = Pr(X(\omega) \leq t).$$

Given this function, we can define the **probability density function** $f_X(t)$ as $\frac{d}{dt} F_x(t)$. Notice that for any a, b, we have

$$Pr(a < X(\omega) \leq b) = \int_a^b f_X(t) dt.$$

A random variable has a **uniform distribution** if its probability density function is a constant; this expresses the idea that uniformly-distributed things "don't care" about the differences between elements of our probability space.

A random variable X with standard deviation σ, expectation μ has a **normal distribution** if its probability density function has the form

$$f_X(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-(t-\mu)^2/2\sigma^2}.$$

This generates the standard "bell-curve" picture that you've seen in tons of different settings. One useful observation about normally-distributed events is that about 68% of the events occur within one standard deviation of the mean (i.e. $Pr(\mu - \sigma < X \leq \mu + \sigma) \approx .68$), about 95% of events occur within two standard deviations of the mean, and about 99.7% of events occur within three standard deviations of the mean.

Definition. For any two events A, B that occur with nonzero probability, define $Pr(A$ given $B)$, denoted $Pr(A|B)$, as the likelihood that A happens given that B happens as well. Mathematically, we define this as follows:

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

In other words, we are taking as our probability space all of the events for which B happens, and measuring how many of them also have A happen.

Definition. Take any two events A, B that occur with nonzero probability. We say that A and B are **independent** if knowledge about A is useless in determining knowledge about B. Mathematically, we can express this as follows:

$$Pr(A) = Pr(A|B).$$

Notice that this is equivalent to asking that

$$Pr(A) \cdot Pr(B) = Pr(A \cap B).$$

Definition. Take any n events $A_1, A_2, \ldots A_n$ that each occur with nonzero probability. We say that these n events are are **mutually independent** if knowledge about any of these A_i events is useless in determining knowledge about any other A_j. Mathematically, we can express this as follows: for any $i_1, \ldots i_k$ and $j \neq i_1, \ldots i_k$, we have

$$Pr(A_j) = Pr(A_j | A_{i_1} \cap \ldots \cap A_{i_k}).$$

It is not hard to prove the following results:

Theorem. A collection of n events $A_1, A_2, \ldots A_n$ are mutually independent if and only if for any distinct $i_1, \ldots i_k \subset \{1, \ldots n\}$, we have

$$Pr(A_{i_1} \cap \ldots \cap A_{i_k}) = \prod_{j=1}^{k} A_{i_j}.$$

Theorem. Given any event A in a probability space Ω, let $A^c = \{\omega \in \Omega \mid \omega \notin A\}$ denote the **complement** of A.

A collection of n events $A_1, A_2, \ldots A_n$ are mutually independent if and only if their complements $\{A_1^c, \ldots A_n^c\}$ are mutually independent.

It is also useful to note the following non-result:

Not-theorem. Pairwise independence does not imply independence! In other words, it is possible for a collection of events $A_1, \ldots A_n$ to all be **pairwise independent** (i.e. $Pr(A_i \cap A_j) = Pr(A_i)Pr(A_j)$ for any i, j) but not mutually independent!

Example. There are many, many examples. One of the simplest is the following: consider the probability space generated by rolling two fair six-sided dice, where any pair (i, j) of faces comes up with probability $1/6$.

Consider the following three events:

- A, the event that the first die comes up even.

- B, the event that the second die comes up even.

- C, the event that the sum of the two dice is odd.

Each of these events clearly has probability 1/2. Moreover, the probability of $A \cap B, A \cap C$ and $B \cap C$ are all clearly 1/4; in the first case we are asking that both dice come up even, in the second we are asking for (even, odd) and in the third asking for (odd, even), all of which happen 1/4 of the time. So these events are pairwise independent, as the probability that any two happen is just the products of their individual probabilities.

However, $A \cap B \cap C$ is impossible, as $A \cap B$ holds iff the sum of our two dice is even! So $Pr(A \cap B \cap C) = 0 \neq Pr(A)Pr(B)Pr(C) = 1/8$, and therefore we are not mutually independent.