Part 1

1.

What is the most rigorous mathematical definition for probability that you know or have learned?

Any of the definitions for probability (that I remember) are some variation of Kolmogorov's Axioms. Here's one from **Mathematical Statistics: An Introduction to Likelihood based Inference** by Richard J Rossi (2018).

Definition (Kolmogorov's Probability Function). *Let Ω be the sample space associated with a chance experiment, and let β be a $\sigma-$algebra of events of Ω. A set function P on β satisfying the following three properties is called a probability function or a probability measure.*

1. $P(\Omega) = 1$

2. $P(A) \geq 0$ for every event $A \in \beta$

3. If $\{A_i : i \in N\} \subset \beta$ is a collection of disjoint events, then:

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

The triple (Ω, β, P) is called a probability space.

2.

How would you describe the purpose of probability to a friend that has a limited math and stats background.

Big picture, probability is all about learning to understand random events. Randomness defines nearly every aspect of our lives, and probability gives us the tools we need to wrangle some understanding and insight into the chaos that surrounds us. With probability, events that would otherwise be incomprehensible can now be broken down and analyzed. I think a lot of people attribute probability to just dice rolls and gambling, but it's so much deeper than that. Probability can help you model the behavior of real world random events like a wildfire or hurricane; potentially saving peoples lives. The world is random, and probability is crucial to learn from the world on its own terms.

3

Give examples of the following three sample spaces.

Discrete and of infinite dimension

Number of times to use a slot machine before a jackpot.

You could get a jackpot in one go, however it could theoretically take infinitely many plays.

Finite and continuous

The precise (to some degree) mass of a 2kg bag of rice from a specific manufacturer. The bags will all be just about 2kg as advertised and they definitely can't have infinite mass, but there will be some variation to the exact mass of each bag. One bag may be 2.0001kg and another may be 2.005kg for example.

Something of interest to you

Proportions of neurodivergent employment across different corporate industries. It's a proportion, so it's between zero and 1. Finite and continuous. It's interesting to think about what industries most/least heavily employ individuals on the autism spectrum for instance. How does employment for bipolar individuals compare from healthcare to finance? Lots of interesting things to look into there I think.

4

i)

Use Bonferroni's Inequality (Example 1.2.10) to show the Bonferroni Correction method will result in a FWER of at most a 5% in the two hypothesis test scenario described above.

Let's start with the default $\alpha = 0.05$ and compute our FWER before we do the correction.

Let A, B be events where we correctly conclude the null for tests 1 and 2 respectively.

$$P(A \cap B) \geq P(A) + P(B) - 1 = .95 + .95 - 1 = 0.9$$

What we can conclude from the default α value here is that the probability we do not correctly conclude the null on at least one test is at least $1 - 0.9 = 0.1$.

Let's take a look with a bonferroni corrected $\alpha = 0.05/2 = 0.025$.

$$P(A \cap B) \geq P(A) + P(B) - 1 = .975 + .975 - 1 = 0.95$$

Looking at our new value here, our FWER is now at most $1 - 0.95 = 0.05$ as expected.

ii)

The usual Bonferroni Correction assumes that the hypothesis tests are independent of each other. Why is this important to your calculation in (i).

The issue is that if we throw dependence into the mix that the computation for $P(A \cap B)$ is now fundamentally different. If if turns out that test 2's results depend on the results of test 1, then the probability of a false positive is going to differ depending on if test 1 correctly concluded the null or not. This becomes a conditional probability problem and we can't assume that Bonferroni Inequality will hold.

Part 2

Exercise 1.4

For events A and B, find formulas for the probabilities of the following events in terms of the quantities $P(A)$, $P(B)$ and $P(A \cap B)$.

A

Either A or B or both

This is just the union of the two events.

$$P(A \cup B)$$

B

Either A or B but not both

We can use the same formula here but we need to be sure to remove the intersection of the two.

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

1.8

A

Derive the general formula for the probability of scoring i points.

$$P(\text{Score i points}) = \frac{\text{Area of ring}}{\text{Area of full circle}}$$

The denominator here is easy, it's just πr^2. For the numerator we can think of the ring as a circle with the center carved out. And that center is also a circle, just a smaller one inside of the larger one. So we just take the area of the larger circle and subtract the area of the smaller circle. The points, i, help us determine the radius of these circles. Let's look at some examples for $i = 1, 3, 5$ so we can notice some patterns and generalize.

Note that $P(i = 1)$ is just shorthand for $P(\text{Score } i \text{ points})$.

$$P(i = 1) = \frac{\pi r^2 - \pi(\frac{4}{5}r)^2}{\pi r^2}$$
$$P(i = 3) = \frac{\pi(\frac{3}{5}r)^2 - \pi(\frac{2}{5}r)^2}{\pi r^2}$$
$$P(i = 5) = \frac{\pi(\frac{1}{5}r)^2 - \pi(\frac{0}{5}r)^2}{\pi r^2}$$

Alright, from here we can begin the derivation. What we see is the radius of both circles are scaled down as the points increase. $i = 1$ uses the entire outside circle, but the 4/5 comes in because that inner circle is one region deeper in. In other words, we go from $(5/5)^2$ scaling r to $((5 - i)/5)^2$ scaling r for the inner circle. That $5 - i$ allows us to creep further into the circle and we want to use that for both circles. For the outside ring, since it's one region out, I use $5-i+1$ as that's how I think of it intuitvely, but will simplify that to $6 - i$ instead.

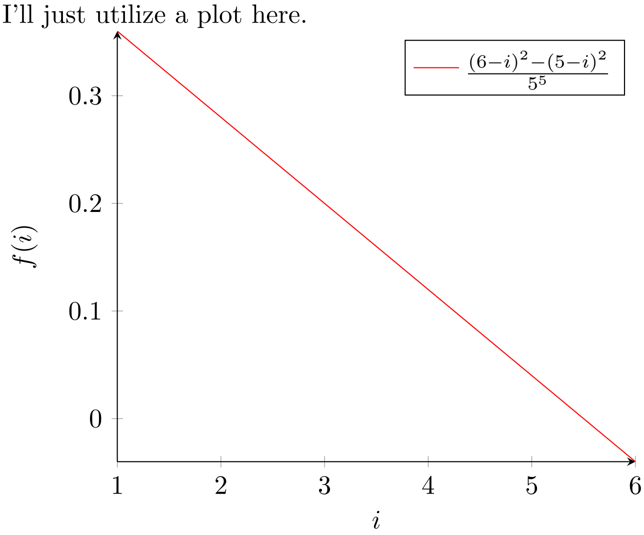
Derivation on the following page.

$$\begin{aligned}
P(\text{Score } i \text{ points}) &= \frac{\pi\left(\frac{5-i+1}{5}r\right)^2 - \pi\left(\frac{5-i}{5}r\right)^2}{\pi r^2} \\
&= \frac{\pi\left(\frac{6-i}{5}r\right)^2 - \pi\left(\frac{5-i}{5}r\right)^2}{\pi r^2} \\
&= \frac{\pi\left(\left(\frac{6-i}{5}\right)^2 r^2 - \left(\frac{5-i}{5}\right)^2 r^2\right)}{\pi r^2} \\
&= \frac{r^2\left(\left(\frac{6-i}{5}\right)^2 - \left(\frac{5-i}{5}\right)^2\right)}{r^2} \\
&= \left(\frac{6-i}{5}\right)^2 - \left(\frac{5-i}{5}\right)^2 \\
&= \frac{(6-i)^2 - (5-i)^2}{5^5}
\end{aligned}$$

B

Show that the function for scoring i points is a decreasing function of i.

I'll just utilize a plot here.



Of note that this plot isn't really perfect for this problem, as this plot is continuous and this problem is discrete. But I'm just looking at the function itself without accounting for i only being integer values because I do not want to learn how to do that in LaTeX.

C

Show that this function is a probability function according to the Kolmogorov Axions.

1. $P(A) \geq 0 \forall A \in \beta$

We can check the plot for this one. All possible values are greater than zero. That line extends beyond 5, but that leaves our domain and so it does not count. For $i = 1, 2, 3, 4, 5$ this checks out.

2. $P(S) = 1$

We can check this with some manual computation.

i	$P(i)$
1	0.36
2	0.28
3	0.20
4	0.12
5	0.04

This sums to 1. We're good!

3. If $A_1, A_2, \cdots \in \beta$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty}) = P(\sum_{i=1}^{\infty} P(A_i))$.

I'll primarily use language for this one. We know these are pairwise disjoint as you can't, for example, score 1 and 2 points at the same point. They are all mutually exclusive from one another. Due to this we can just return to the table from the second axiom. The probability of all the unions is the exact same as summing up all the individual probabilities.

1.11

Show that the intersection of two sigma algebras is a sigma algebra.

For a set to be a sigma algebra, represented by β, it must meet the following requirements:
1. $\emptyset \in \beta$
2. $A \in \beta \implies A^c \in \beta$
3. $\{A_1, A_2, \ldots\} \in \beta \implies \cup_{i=1}^{\infty} A_i \in \beta$

Let X, Y both be sigma algebras of S.
1.
Because X is a sigma algebra, it follows that $\emptyset \in X$.
Because Y is a sigma algebra, it follows that $\emptyset \in Y$.
Therefore, $\emptyset \in X \cap Y$
2.
Let $A \in X \cap Y$
$A \in X$, so $A^c \in X$
$A \in Y$, so $A^c \in Y$
Therefore, $A^c \in X \cap Y$
3.
Let $\{A_1, A_2, \cdots\} \in X \cap Y$
$\{A_1, A_2, \cdots\} \in X$, so $\cup_{i=1}^{\infty} A_i \in X$
$\{A_1, A_2, \cdots\} \in Y$, so $\cup_{i=1}^{\infty} A_i \in Y$
Therefore, $\{A_1, A_2, \cdots\} \in X \cap Y$
Conclusion:
$X \cap Y$ is a sigma algebra as it fulfills all of the necessary requirements.

1.18

If n balls are placed at random into n cells, find the probability that exactly one cell remains empty.

Answer provided by the book: $\binom{n}{2}n!/n^n$

Let's work backwards from the answer and explain each component:

Denominator: n^n.

This bit I think depends on how you interpret the question. My hunch here is that $n!$ makes more sense as I assumed a ball would go away every time one was assigned a cell. However, if we assume we're sampling with replacement for the denominator specifically then we do end up with n^n.

Numerator: $\binom{n}{2}$

This part of the numerator here is interesting. The purpose of it is that for one cell to be empty, one cell must contain 2 balls. If we have n balls and must choose 2 to go into a cell, we have $\binom{n}{2}$ ways to do that.

Numerator: $n!$

This chunk is actually two parts simplified together. So, first off, we need to select a cell to be empty. We have $\binom{n}{1} = n$ ways of doing that. Then, for the rest of the cells, we have $n-1, n-2, \cdots, 1$ ways of assigning the remaining balls. So we get $n \cdot (n-1)! = n!$.

1.31

A

Prove that, in general, sampling with replacement from the set $\{x_1, x_2, \cdots, x_n\}$, the outcome with average $(x_1 + x_2 + \cdots + x_n)/n$ is the most likely, having probability $\frac{n!}{n^n}$.

Before we start with a proof, some intuition. My brain jumps to dice here, so let's say we roll a 4-sided dice twice and take the average out the results. We're sampling with replacement here because we can roll the same value multiple times.

Here's a table of the possible outcomes.

	x=1	x=2	x=3	x=4
y=1	1	1.5	2	2.5
y=2	1.5	2	2.5	3
y=3	2	2.5	3	3.5
y=4	2.5	3	3.5	4

Note how 2.5 has the most cells in the table?

2.5 is also the expected value of a d4 dice.

$$\begin{aligned} E[X] &= \sum_{i=1}^{4} \frac{1}{4} x_i \\ &= \frac{1}{4} \cdot (1 + 2 + 3 + 4) \\ &= 2.5 \end{aligned}$$

Really all this problem is asking is if this trend generalizes to an n-sided dice being rolled n times and averaged.

To be honest, I'm late on this assignment and unsure how I would tackle this proof. But, with this d4 example we would see what's called a trianglular distribution, where the average is the most common value. I'm sure this would continue on as you scale up the experiment.

B

Use Stirling's Formula to show that $n!/n^n \approx \sqrt{2n\pi}/e^n$

Stirling's Formula: $n! \approx \sqrt{2\pi} n^{n+1/2} e^{-n}$

So we just divide by n^n and simplify.

$$\begin{aligned}\frac{n!}{n^n} &\approx \frac{\sqrt{2\pi}n^{n+1/2}e^{-n}}{n^n} \\ &\approx \sqrt{2\pi}\cdot n^{1/2}\cdot\frac{1}{e^n} \\ &\approx \frac{\sqrt{2\pi n}}{e^n}\end{aligned}$$

C

Show that the probability that a particular x_i is missing from an outcome is:

$$\left(1-\frac{1}{2}\right)^n \to e^{-1} \text{ as } n\to\infty$$

Let's first understand the leftmost side here. Let's assume every x_i is equally likely because I think we have to assume that here.

Then the probability that you don't see a specific outcome is simply $1-\frac{1}{n}$. Since we're checking every single x_i and we have n of them, we get $\left(1-\frac{1}{2}\right)^n$.

As for the convergence, quick sanity check. $e^{-1}\approx 0.36787$

$(1-\frac{1}{1000})\approx 0.36769$.

Okay, I'm convinced it does then.

I'm going to be honest, I do not remember how to evaluate limits like these. So I'm just going to plot this!