Final Assignment: Intro. Probability and Statistics
Due October 24, 2023

John McCuan

Problem 1 (asymptotics; Jeremy Mahoney's project) A function $f : \mathbb{N} \to \mathbb{R}$ is said to **approach a limit as n tends to** ∞ and we write

$$\lim_{n \nearrow \infty} f(n) = L$$

if there exists some real number L for which the following holds:

> Given any $\epsilon > 0$, there is some $N > 0$ for which
>
> $$|f(n) - L| < \epsilon \qquad \text{whenever } n > N.$$

Many functions $f : \mathbb{N} \to \infty$ satisfy

$$\lim_{n \nearrow \infty} f(n) = +\infty. \tag{1}$$

For such functions one sometimes still wishes to make quantitative statements. Some examples of statements which can sometimes be made and the associated notation are given below. Most often such statements ultimately rely on the definition of limit given above, so it is important to understand that definition well.

Definition 1 We say $f : \mathbb{N} \to \infty$ is **asymptotic to zero order** to $g : \mathbb{N} \to \infty$ **at infinity** if

$$\lim_{n \nearrow \infty} |f(n) - g(n)| = 0.$$

Definition 2 We say $f : \mathbb{N} \to \infty$ is "little-o" of $g : \mathbb{N} \to \infty$ as $n \nearrow \infty$ and write $f(n) = \circ(g(n))$ as $n \nearrow \infty$ if

$$\lim_{n \nearrow \infty} \frac{f(n)}{g(n)} = 0.$$

Definition 3 We say $f : \mathbb{N} \to \infty$ is "big-o" of $g : \mathbb{N} \to \infty$ as $n \nearrow \infty$ and write $f(n) = \bigcirc(g(n))$ as $n \nearrow \infty$ if there is some constant C and some constant N for which

$$\left| \frac{f(n)}{g(n)} \right| < C \qquad \text{whenever } n > N.$$

Definition 4 Given $p > 0$ and $g : \mathbb{N} \to \infty$ with

$$\lim_{n \nearrow \infty} g(n) = +\infty,$$

we say $f : \mathbb{N} \to \infty$ is **asymptotic of order** p to g **at infinity** as $n \nearrow \infty$ if

$$|f(n) - g(n)| = \circ(n^p).$$

(a) Notice I did not define the limit in (1). Can you give a precise definition of what this means? (Try to formulate your answer/definition based on the definition of limit to a finite value given above without looking up the definition from another source.)

(b) If $k \in \mathbb{N}$ is fixed and $f : \mathbb{N} \to \mathbb{R}$ by

$$f(n) = \frac{n!}{(n-k)!} \tag{2}$$

and $g : \mathbb{N} \to \mathbb{R}$ by

$$g(n) = n^k, \tag{3}$$

show the following:

(i)

$$\lim_{n \nearrow \infty} f(n) = +\infty.$$

(ii)

$$\lim_{n \nearrow \infty} g(n) = +\infty.$$

(iii)

$$\lim_{n \nearrow \infty} [g(n) - f(n)] = +\infty.$$

(c) If f is asymptotic of order p to g as $n \nearrow \infty$ and $q > p$, then show f is asymptotic or order q to g as $n \nearrow \infty$.

(d) With $f, g : \mathbb{N} \to \mathbb{R}$ given by (2) and (3) show the following

- **(i)** f is not asymptotic to zero order to g as $n \nearrow \infty$.
- **(ii)** f is asymptotic of order k to g as $n \nearrow \infty$ and therefore,

$$\lim_{n \nearrow \infty} \left| \frac{f(x)}{g(x)} - 1 \right| = 0.$$

(e) If $f : \mathbb{N} \to \mathbb{R}$ satisfies

$$\lim_{n \nearrow \infty} f(n) = L,$$

then is it true that f is asymptotic to zero order to L, i.e., to the constant function $g : \mathbb{N} \to \mathbb{R}$ with $g(n) \equiv L$, as $n \nearrow \infty$?

(f) If $f : \mathbb{N} \to \mathbb{R}$ is asymptotic to zero order to $g : \mathbb{N} \to \mathbb{R}$ as $n \nearrow \infty$, then is it true that

$$\lim_{n \nearrow \infty} f(n) = g(n)?$$

(g) With $f, g : \mathbb{N} \to \mathbb{R}$ given by (2) and (3) can you determine for which $m \in \mathbb{N}$ there holds

$$g(n) - f(n) = \bigcirc(n^m) \qquad \text{as } n \nearrow \infty?$$

Problem 2 (length measures cannot measure all sets; Keegan Thompson's project) Complete the steps outlined below[1] in showing it is impossible to have a translation invariant length measure on the interval $[0,1)$ with domain the collection $\wp([0,1))$ of all subsets of $[0,1)$.

Recall the argument is by contradiction. It is assumed

$$\mu : \wp([0,1)) \to [0,1] \tag{4}$$

is a measure having the following two properties:

(L) If I is any interval in $[0,1)$, meaning I has one of the following forms:

$$\begin{aligned}
(a,b) &= \{x : a < x < b\} \quad \text{for some } a,b \in [0,1) \text{ with } a < b,\\
[a,b) &= \{x : a \le x < b\} \quad \text{for some } a,b \in [0,1) \text{ with } a < b,\\
(a,b] &= \{x : a < x \le b\} \quad \text{for some } a,b \in [0,1) \text{ with } a < b, \text{ or}\\
[a,b] &= \{x : a \le x \le b\} \quad \text{for some } a,b \in [0,1) \text{ with } a \le b,
\end{aligned}$$

then $\mu(I) = \text{length}(I) = b - a$.

(T) If $A \subset [0,1)$ and $t \in \mathbb{R}$ and $\{x+t : x \in A\} \subset [0,1)$, then

$$\mu(\{x+t : x \in A\}) = \mu(A).$$

A measure satisfying **(L)** is said to be a **length measure**. A measure satisfying **(T)** is said to be **translation invariant**.

(a) We have introduced (really) three kinds of measures in this course:

- **(i)** adolescent measures (including baby measures which are a special case of adolescent measures),
- **(ii)** abstract measures, and
- **(iii)** integral measures on intervals in $\mathbb{R}$.

We have focused on the special cases of **(i)** and **(iii)** which are probability measures. The abstract measures **(ii)**, which you may not have thought about too

[1] This material is from the book *Real Analysis* by Halsey Royden, who incidentally was one of my teachers in graduate school and cured me from any youthful interest in the subject of topological vector spaces. Royden's construction of a non-measurable set, however, I find to be quite inspiring.

much, are important because integral measures are constructed using integration with respect to (one dimensional) Lebesgue measure which was introduced as an abstract measure on subsets of $\mathbb{R}$. The concept of an abstract measure is also the important one for this problem:

(i) Write down carefully the definition of an (abstract) measure. Hint: This can be found in Chapter 4 on page 200 of my notes.

(ii) In the application of your definition to the assumption of the existence of the translation invariant length measure in (4) what set plays the role of the σ-algebra?

(b) An **equivalence relation** on a set S is any subset R of $S \times S$ for which the following hold

(i) $(x, x) \in R$ for all $x \in S$,

(ii) If $(x, y) \in R$, then $(y, x) \in R$, and

(iii) If $(x, y) \in R$ and $(y, z) \in R$, then $(x, z) \in R$.

Property **(i)** is called the **reflexive** property and is usually expressed by writing $x \sim x$, where the equivalence relation is informally represented by the notation "$\sim$." Similarly, an equivalence relation is said to be **symmetric** if **(ii)** holds, and this is informally expressed by writing

$$x \sim y \qquad \Longrightarrow \qquad y \sim x.$$

The third property is called the **transitive** property:

$$x \sim y \quad \text{and} \quad y \sim z \qquad \Longrightarrow \qquad x \sim z.$$

Most of the time when you use the symbol "=" in mathematics, it is denoting some equivalence relation. For example, "n=m" represents the equivalence relation of equality on the natural numbers, meaning the sets representing the natural numbers m and n have the **same number of elements**. The same symbol "=" is used to represent the relation of set equality in $\wp(S)$ where $A = B$ means the sets A and B have **the same elements**. These are two different equivalence relations with which you are familiar, and you can check that each is reflexive, symmetric, and transitive.

Show that any time one has an equivalence relation "$\sim$" on a set S, then the collection

$$\mathcal{P} = \big\{\{y \in S : y \sim x\} : x \in S\big\}$$

is a **partition** of S. Each set $A_x = \{y \in S : y \sim x\}$ is called the **equivalence class** of $x \in S$, and what you need to show is that either two equivalence classes A_x and A_w are disjoint, i.e., $A_x \cap A_w = \phi$, or identical, i.e., $A_x = A_w$. Hint: Remember that in order to show two sets are equal, you need to show each is a subset of the other.

(c) (rational equivalence) Let $\mathbb{Q}$ denote the **rational numbers**

$$\mathbb{Q} = \left\{ \frac{m}{n} : n \in \mathbb{N} = \{1, 2, 3, \ldots\} \text{ and } m \in \mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \ldots\} \right\}.$$

Show $x \sim_{\mathbb{Q}} y$ if $x - y \in \mathbb{Q}$ defines an equivalence relation on $[0, 1)$.

As a consequence of parts **(b)** and **(c)** above, the equivalence classes

$$\big\{A_x = \{y \in [0,1) : y \sim_{\mathbb{Q}} x\} : x \in [0,1)\big\},$$

where "$\sim_{\mathbb{Q}}$" represents rational equivalence, are a partition of $[0, 1)$.

Of course, it may be the case that $A_x = A_y$ for elements $x, y \in [0, 1)$ with $x \neq y$. In the application below, however, we use a particular index set $J \subset [0, 1)$ for which

$$\big\{A_x : x \in [0,1)\big\} = \{A_x : x \in J\}$$

and $A_x = A_y$ for $x, y \in J$ implies $x = y$. That is to say, the set $J \subset [0, 1)$ contains exactly one element from each equivalence class.[2]

(d) (mod 1 addition) The function $m : [0, 1) \times [0, 1) \to [0, 1)$ given by

$$m(x,y) = \begin{cases} x + y, & \text{if } x + y < 1 \\ x + y - 1, & \text{if } x + y \geq 1 \end{cases}$$

[2]Technically, the existence of this set J follows from an application of the axiom of choice. Strange though it may seem, the existence of such a set J also implies the axiom of choice and is thus equivalent to the axiom of choice. More generally, one says the following: The existence of a non-measurable set is equivalent to the axiom of choice.

is called **mod 1 addition**. Recall that the rational numbers $\mathbb{Q}$ are countable. This means there is a bijection $r : \mathbb{N} \to \mathbb{Q}$. Let us denote the image $r(j)$ of each natural number j under this bijection is denoted by r_j so that

$$\mathbb{Q} = \{r_j\}_{j=1}^{\infty}.$$

You should now think: r_1 is the first rational number, r_2 is the second rational number, and so on.

For each $j = 1, 2, 3, \ldots$, consider the "r_j shuffle" of J defined by

$$E_j = \{m(x, r_j) : x \in J\}.$$

(i) Draw a picture of the set E_j. (You'll have to be creative about how to illustrate/draw the set J because no one knows what J actually looks like.)

(ii) Use translation invariance to show $\mu(E_i) = \mu(E_j)$ for every $i, j \in \mathbb{N}$.

(iii) Show $E_i \cap E_j = \phi$ if $i \neq j$.

(iv) Show

$$\bigcup_{j=1}^{\infty} E_j = [0, 1).$$

Hint: If $x \in [0, 1)$, there is some $x_0 \in J$ for which $A_{x_0} = A_x$. That is, $x - x_0 \in \mathbb{Q}$.

(e) As a consequence of **(d)(iii)** and **(d)(iv)** the collection $\{E_j\}_{j=1}^{\infty}$ is a countable partition of $[0, 1)$. Also, by **(d)(ii)** each set E_j has the same measure. Use the countable additivity from your definition of (abstract) measure to obtain a contradiction showing it is impossible to measure all subsets of $[0, 1)$ with a length measure.

(f) Which set in the discussion above would you identify as the "non-measurable set?"

(g) Replace the non-measurable set you identified in part **(f)** with a measurable set, say $[1/4, 3/4] \subset [0, 1)$.

(0) Construct a sequence of sets $F_1, F_2, F_3, \ldots$ corresponding to the rational shuffles of $[1/4, 3/4]$.

(i) Draw a picture of the set F_j. Now you can see the picture clearly, but you should get different cases depending on which rational r_j is involved.

(ii) Is it still true that $\mu(F_i) = \mu(F_j)$ for every $i, j \in \mathbb{N}$? If so, why? If not, does equality hold for some rationals $r_i \neq r_j$? Can you characterize the cases of failure?

(iii) Is it still true that $F_i \cap F_j = \phi$ if $i \neq j$? If so, why? If not, are the sets F_i and F_j disjoint for some rationals $r_i \neq r_j$? Can you characterize the cases of failure?

(iv) Is it still true that

$$\bigcup_{j=1}^{\infty} F_j = [0, 1) \ ?$$

(v) Why is there no contradiction in this case?

(h) Can you think of a different choice for the measurable set in part **(g)** above making the discussion of part **(g)** more interesting? How about a Cantor set?

Problem 3 (Chebyshev inequality; inspired by Ansel Erol's probability project) Let $\delta : \mathbb{R} \to [0, \infty)$ be a probability density (MDF) for which the mean

$$\mu = \int_{\omega \in \mathbb{R}} \omega \, \delta(\omega)$$

and the variance

$$\sigma^2 = \int_{\omega \in \mathbb{R}} (\omega - \mu)^2 \, \delta(\omega) \tag{5}$$

are well-defined. Let $\sigma = \sqrt{\sigma^2}$ denote the standard deviation of the measure determined by δ and let $k > 0$ be a fixed (proportionality) constant. Chebyshev's inequality states that under these assumptions

$$\int_{\{\tau \in \mathbb{R} : |\tau - \mu| < k\sigma\}} \delta \geq 1 - \frac{1}{k^2}. \tag{6}$$

The objective of this problem is to walk you through a proof of this inequality. I will begin with an auxiliary inequality which is a special case of the inequality in Lemma 1 (the McMarkov lemma) of Problem 10 in Assignment 9. I will give the derivation of this inequality which we can call the Chevy inequality:

$$\int_{\omega \in \mathbb{R}} (\omega - \mu)^2 \, \delta(\omega) \geq k^2 \sigma^2 \int_{\{\tau \in \mathbb{R} : |\tau - \mu| \geq k\sigma\}} \delta. \tag{7}$$

Here is the derivation:

$$\int_{\omega \in \mathbb{R}} (\omega - \mu)^2 \, \delta(\omega) \geq \int_{\omega \in \{\tau \in \mathbb{R} : (\tau - \mu)^2 \geq k^2\sigma^2\}} (\omega - \mu)^2 \, \delta(\omega) \tag{8}$$

$$\geq \int_{\omega \in \{\tau \in \mathbb{R} : (\tau - \mu)^2 \geq k^2\sigma^2\}} k^2 \sigma^2 \delta(\omega) \tag{9}$$

$$= k^2 \sigma^2 \int_{\omega \in \{\tau \in \mathbb{R} : (\tau - \mu)^2 \geq k^2\sigma^2\}} \delta(\omega) \tag{10}$$

$$= k^2 \sigma^2 \int_{\omega \in \{\tau \in \mathbb{R} : |\tau - \mu| \geq k\sigma\}} \delta(\omega) \tag{11}$$

$$= k^2 \sigma^2 \int_{\{\tau \in \mathbb{R} : |\tau - \mu| \geq k\sigma\}} \delta. \tag{12}$$

The reasoning behind each step is as follows:

1. (8) follows because the integrand is everywhere non-negative and the integral on the right is over a potentially smaller set.

2. (9) follows because $(\omega - \mu)^2 \geq k^2\sigma^2$ for ω in the (smaller) set
$$\{\tau \in \mathbb{R} : (\tau - \mu)^2 \geq k^2\sigma^2\}.$$

3. (10) is the linearity of the integral.

4. (11) is a consequence of the fact that
$$\{\tau \in \mathbb{R} : (\tau - \mu)^2 \geq k^2\sigma^2\} \qquad \text{and} \qquad \{\tau \in \mathbb{R} : |\tau - \mu| \geq k\sigma\}$$
are the same set.

5. (12) is just a change of notation because the variable of integration ω need no longer appear in the integrand.

Complete steps **(a)**-**(d)** to derive Chebyshev's inequality (6):

(a) Divide both sides of the Chevy inequality (7) by $k^2\sigma^2$ and reverse the order to get an estimate from above on the integral
$$\int_{\{\tau\in\mathbb{R}:|\tau-\mu|\geq k\sigma\}} \delta. \tag{13}$$

(b) Replace the integral (13) appearing in your inequality from part **(a)** using the value of
$$\int_{\{\tau\in\mathbb{R}:|\tau-\mu|<k\sigma\}} \delta + \int_{\{\tau\in\mathbb{R}:|\tau-\mu|\geq k\sigma\}} \delta.$$

(c) Replace the other integral appearing in your inequality from part **(a)** with its value from (5).

(d) Algebraically simplify and rearrange what you have to complete the derivation.

(e) If one believes in "probability" and "random variables," which probably no one should, then Chebyshev's inequality (6) may be rephrased as an inequality involving such things:

- How would a random variable X be introduced with values probabilistically related to δ?
- Within this framework, how would Chebychev's inequality (6) be expressed?