Math 3338: Supplemental Notes and Exercises 1, through Thursday, 9/3

Exercises are bulleted ($\bullet$) below.

A probability space is a triple $< S, \mathcal{E}, P >$ consisting of a set S called the *sample space*, a family $\mathcal{E}$ of subsets $E \subseteq S$ called the *event family*, and a real valued measure P defined for every event E from the family $\mathcal{E}$. The event family $\mathcal{E}$ is required at a minimum to contain S and $\emptyset$. More generally, $\mathcal{E}$ must satisfy:

($\mathcal{E}$.1) If E is a member of the family $\mathcal{E}$, its complement E^c must also be in the family. (Here the complement is relative to S; $E^c = \{s \in S : s \notin E\}$.)

($\mathcal{E}$.2) If $E_1, E_2, E_3 \ldots$ are (possibly an infinite number of) sets from $\mathcal{E}$, the union $\bigcup_n E_n$ must also be a member of $\mathcal{E}$.

The term σ*–algebra* is commonly used to describe such a family of sets. Finally, P must be a *probability measure*. That is, it must satisfy:

(P.1) $P(E) \in [0, 1]$ for every event E from $\mathcal{E}$.

(P.2) $P(S) = 1$ where S is the sample space.

(P.3) For any pairwise disjoint sequence of events $E_1, E_2, E_3 \ldots$ from the family $\mathcal{E}$, (this means $E_m \cap E_n = \emptyset$ for all $m \neq n$), we must have $P(\bigcup_n E_n) = \sum_n P(E_n)$.

Do the following as homework exercises.

- Suppose A, B and C are sets. (a) Prove that $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$. (b) Prove that $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$. (c) By example, show that in general $(A \cap B) \cup C \neq A \cap (B \cup C)$.

- Suppose $A_1, A_2, A_3, \ldots$ are subsets of U, and define $A^c = \{u \in U : u \notin A\}$. (a) Prove that $(\bigcup A_n)^c = \bigcap A_n^c$. (b) Prove that $(\bigcap A_n)^c = \bigcup A_n^c$. These are called *de Morgan's* laws.

- Suppose the E_1 and E_2 are sets from $\mathcal{E}$ with $E_1 \subseteq E_2$. Prove that $P(E_1) \leq P(E_2)$. Hint: Show that $E_2 = E_1 \cup (E_2 \cap E_1^c)$.

- Suppose E is from $\mathcal{E}$. Prove that $P(E^c) = 1 - P(E)$. Conclude that $P(\emptyset) = 0$.

- Let $E_1, E_2, E_3, \ldots$ be any infinite sequence of pairwise disjoint sets from $\mathcal{E}$. Conclude that $\sum_{k=1}^{\infty} P(E_k) \leq 1$. Moreover, prove that $\lim_{n\to\infty} \sum_{k=n}^{\infty} P(E_k) = 0$.

When the sample space is finite, things are pretty easy to visualize. Here's an example. Suppose we flip two coins. The sample space consists of pairs

$$S = \{(H, H), (H, T), (T, H), (T, T)\};$$

$H \sim heads$, $T \sim tails$. $\mathcal{E}$ consists of the empty set, the sample space S, as well as

$$\begin{aligned}
(\# = 1) \quad & \{(H,H)\},\ \{(H,T)\},\ \{(T,H)\},\ \{(T,T)\}, \\
(\# = 2) \quad & \{(H,H),(H,T)\},\ \{(H,H),(T,H)\},\ \{(H,H),(T,T)\}, \\
& \{(H,T),(T,H)\},\ \{(H,T),(T,T)\},\ \{(T,H),(T,T)\}, \\
(\# = 3) \quad & \{(H,H),(H,T),(T,H)\},\ \{(H,H),(H,T),(T,T)\}, \\
& \{(H,H),(T,H),(T,T)\},\ \{(H,T),(T,H),(T,T)\}.
\end{aligned}$$

(Phew, glad I didn't flip three coins.) That is, here the event family $\mathcal{E}$ consists of $\emptyset$, S, and all other subsets of S. Notice that all events from $(\# = 1)$ are pairwise disjoint, whereas not all from $(\# = 2)$ are. Can you determine two disjoint events from $(\# = 2)$? Are there any two disjoint events from $(\# = 3)$?

Assuming the coins are *fair*, assign values to $P(E)$ as follows.

$$\begin{aligned}
& P(\emptyset) = 0,\ P(S) = 1, \text{ and} \\
& P(E) = 1/4 \text{ when the event } E \text{ is from } (\# = 1) \text{ above}, \\
& P(E) = 1/2 \text{ when the event } E \text{ is from } (\# = 2), \\
& P(E) = 3/4 \text{ when the event } E \text{ is from } (\# = 3).
\end{aligned}$$

Check that P so defined is consistent with items (P.1)–(P.3) above.

Do the following as exercises.

- Suppose the given two coins are not fair. Specifically, suppose that

$$\begin{aligned}
P(\{(H,H)\}) = 3/8, \quad P(\{(H,T)\}) = 3/8, \\
P(\{(T,H)\}) = 1/8, \quad P(\{(T,T)\}) = 1/8.
\end{aligned}$$

Determine the probabilities for all other events E in $\mathcal{E}$ which are listed in $(\# = 2)$ and $(\# = 3)$ above.

- Suppose we have two other "crazy" coins and we observe that

$$P(\{(H,H)\}) = 1/8, \quad P(\{(T,T)\}) = 1/2, \quad P(\{(H,H),(H,T)\}) = 1/4.$$

Use these to calculate the probabilities of all events.

When the sample space is infinite, we may run into some technical difficulties. Here's a second example intended to demonstrate this fact. Suppose we spin a wheel with a pointer, and after it stops we record the pointer's angle clockwise from vertical. Here the sample space S is the set of all angles θ in the interval $[0, 2\pi)$. We might want to define the event family $\mathcal{E}$ as <u>all</u> subsets of $[0, 2\pi)$. We might also want to define the probability measure of a given event E by

$$P(E) = \frac{1}{2\pi} \int_E d\theta \equiv \frac{1}{2\pi} \int_0^{2\pi} \chi_E(\theta) d\theta, \text{ where } \chi_E(\theta) = \begin{cases} 1 & \text{if } \theta \in E \\ 0 & \text{otherwise.} \end{cases}$$

(χ_E is called the *characteristice function* of the set E.) All this seems pretty reasonable. For example $P(S) = 1/2\pi \int_0^{2\pi} 1\,d\theta = 1$ and $P(\emptyset) = 1/2\pi \int_0^{2\pi} 0\,d\theta = 0$. Moreover, suppose a given event E is the pointer stops between twelve noon and three o'clock. Then we get

$$\chi_E(\theta) = \begin{cases} 1 & \text{if } 0 \le \theta \le \pi/2 \\ 0 & \text{otherwise} \end{cases}$$

$$\text{which gives } P(E) = \frac{1}{2\pi}\left(\int_0^{\pi/2} 1\,d\theta + \int_{\pi/2}^{2\pi} 0\,d\theta\right) = 1/4.$$

Suppose another event E is either the pointer stops between twelve noon and three o'clock or it stops between six o'clock and nine o'clock. Here we get

$$\chi_E(\theta) = \begin{cases} 1 & \text{if } 0 \le \theta \le \pi/2 \\ 1 & \text{if } \pi \le \theta \le 3\pi/2 \\ 0 & \text{otherwise} \end{cases} \qquad \text{(Make a plot of this.)}$$

$$\text{which gives } P(E) = \frac{1}{2\pi}\left(\int_0^{\pi/2} 1\,d\theta + \int_{\pi/2}^{\pi} 0\,d\theta + \int_{\pi}^{3\pi/2} 1\,d\theta + \int_{3\pi/2}^{2\pi} 0\,d\theta\right) = 1/2.$$

But now we want to consider an event which highlights the technical difficulty mentioned above. The Riemann integral of χ_E is well defined only when E is composed of a finite union of intervals. What if we were to ask "what is the probability that the pointer stops at an angle which is a rational number?" This event corresponds to $E = \mathbb{Q} \cap [0, 2\pi)$ and is a perfectly valid subset of the sample space $[0, 2\pi)$. However, the Riemann integral of this E's characteristic function, and therefore $P(E)$, is not defined. To rectify this difficulty, we might want to just neglect such events by restricting the family of events to include only those event sets whose characteristics functions are Riemann integrable. However, by doing this, we arrive at another serious difficulty. Such a family is not a σ-algebra, and the requirement that the event family forms a σ-algebra is so critical to the theory of probability that it can not be relaxed.

Even during Riemann's lifetime, shortcomings present in his theory of the integral were well known. The modern theory of the integral was introduced by the French mathematician Henri Lebesgue around the beginning of the 20th century; see for example [1]. The centerpiece of Lebesgue's theory is the notion of the measure of a set of real numbers. I'll give a quick outline of *Lebesgue measure.* See [2] for an excellent set of detailed notes.

The measure of a bounded interval I, denoted by $m(I)$, with endpoints $a < b$ is given by $b - a$. That is, $m(I)$ is just the length of I. The measure of the empty set, $m(\emptyset)$, is defined to be zero. One can prove that any bounded and nonempty open subset of real numbers, say O, is composed of a unique countable (or possibly finite) union of disjoint nonempty open intervals, $O = \bigcup_n (a_n, b_n)$. The measure of O is given by

[1] `http://en.wikipedia.org/wiki/Lebesgue`
[2] `http://web.media.mit.edu/~lifton/snippets/measure_theory.pdf`

$m(O) = \sum_n (b_n - a_n)$ and is a well defined real number. Now, suppose E denotes a subset of S where S is a nonempty bounded interval. The *outer measure* of E is denoted by $m_o(E)$ and is given by the well defined number

$$m_o(E) \equiv \inf\{m(O) : E \subseteq O \cap S,\ O \text{ is open relative to } S\} \leq m(S).$$

(While not technically correct, you may think of inf as meaning: *the minimum of.*) The *inner measure* of E is denoted by $m_i(E)$ and is given by

$$m_i(E) \equiv m(S) - m_o(E^c),$$

where $E^c = \{s \in S : s \notin E\}$. For any subset $E \subseteq S$ we always have $0 \leq m_i(E)$ as well as $m_i(E) \leq m_o(E)$. In the case when $m_i(E) = m_o(E)$ we say E is *Lebesgue measurable*, and we denote its measure $m(E)$ as this common value. It is very difficult to even imagine a bounded set of real numbers which is not Lebesgue measurable, yet in a weird sense such sets do exist; see [3] for example. Some important consequences of Lebesgue's theory include:

(L.1) The family $\mathcal{E}$ of all Lebesgue measurable subsets of a bounded interval is a σ-algebra.

(L.2) If $E_1, E_2, E_3, \ldots$ are disjoint members of $\mathcal{E}$, then $m(\bigcup_n E_n) = \sum_n m(E_n)$.

Do the following as homework exercises.

• Consider $E = \mathbb{Q} \cap [0, 2\pi)$ discussed above in the wheel example. Show that the set E is Lebesgue measurable and $m(E) = 0$. Hint: E is countable and may be written as $E = \bigcup_{n=1}^{\infty} \{r_n\}$. Moreover $m(\{r_n\}) = 0$ for each n.

• Again from the wheel example, show that $< [0, 2\pi), \mathcal{E}, m(E)/2\pi >$, where $\mathcal{E}$ denotes the family of Lebesgue measurable subsets of $[0, 2\pi)$, defines a probability space.

• Suppose f is a continuous and nonnegative real valued function defined on the interval $[0, 1]$. Define a set $F_y = \{x \in [0, 1] : f(x) > y\}$, and let $\Lambda_f(y)$ denote the measure of F_y. (a) Visualize what the function $\Lambda_f(y)$ represents for $y > 0$. (b) Try to visualize what $\int_0^\infty \Lambda_f(y)\, dy$ represents. Hint: Draw a plot of a typical f and then superimpose a horizontal line on your plot. Sketch the set of x's where $f(x)$ lies above the line. (c) Is $\Lambda_f(y)$ a nonincreasing function of y?

We return now to an abstract probability space $< S, \mathcal{E}, P >$. A *random variable* is a real valued function, say X, defined on the sample space S, i.e. $X : S \to \mathbb{R}$, such that for every $y \in \mathbb{R}$ the set $\{s \in S : X(s) \leq y\}$ is a member of the event family $\mathcal{E}$.

Let's return for the moment to the two coin example discussed earlier. The sample space there was given by $S = \{(H, H), (H, T), (T, H), (T, T)\}$. Consider two payoff

[3] http://en.wikipedia.org/wiki/Vitali_set

functions:

$$X(s) = \begin{cases} +1.0 & \text{if } s = (H,H) \\ 0.0 & \text{if } s = (H,T) \\ 0.0 & \text{if } s = (T,H) \\ -1.0 & \text{if } s = (T,T), \end{cases} \qquad Y(s) = \begin{cases} +2.0 & \text{if } s = (H,H) \\ 1.0 & \text{if } s = (H,T) \\ 0.0 & \text{if } s = (T,H) \\ -1.0 & \text{if } s = (T,T). \end{cases}$$

Both of these functions define random variables for this example.

- For every real number y, determine: (a) $\{s \in S : X(s) \leq y\}$. (b) $\{s \in S : Y(s) \leq y\}$.

Let's look at the wheel example again. The sample space there was given by $S = [0, 2\pi)$, and we take $\mathcal{E}$ to be the family of measurable subsets of S. A function X from a bounded interval $I \subset \mathbb{R}$ into the reals is said to be a *measurable function* if for every $y \in \mathbb{R}$ the set $\{x \in I : X(x) \leq y\}$ is a measurable subset of I. Consider two payoff functions:

$$X(s) = s, \quad Y(s) = \begin{cases} 1/s & \text{if } s \neq 0 \\ 0 & \text{if } s = 0. \end{cases}$$

Both of these functions are measurable on $[0, 2\pi)$ and therefore for this example define random variables.

- For every real number y, determine: (a) $\{s \in S : X(s) \leq y\}$. (b) $\{s \in S : Y(s) \leq y\}$. (c) Conclude that both functions are measurable. (As with measurable sets, it's hard to imagine a function which is not measurable.)

The concept of a random variable gives rise to what is called the random variable's *distribution function.* This function is also commonly called the *cumulative distribution function* or simply the CDF. Let $< S, \mathcal{E}, P >$ be a probability space and let X be an associated random variable. Then, the CDF for X, denoted here and in the text by F_X, is defined by

$$F_X(x) = P(\{s \in S : X(s) \leq x\}).$$

In the literature, it is standard convention to use the convenient but somewhat inconsistent shorthand notation

$$P(X \leq x) \text{ in place of } P(\{s \in S : X(s) \leq x\}).$$

Notice that $F_X : \mathbb{R} \to [0, 1]$ no matter what the underlying probability space is. The following is important to see however: Even when two random variable, say X and Y, share the same underlying probability space, in general their CDFs, F_X and F_Y, are different.

- Determine the CDFs for X and Y given in the two coins example above when the probability of an event E having n outcomes is given by $P(E) = n/4$; e.g.

$$P(\{(H,H)\}) = 1/4, \ \ldots, \ P(\{(T,T)\}) = 1/4, \ P(\{(H,H),(H,T)\}) = 2/4, \ \textit{etc.}$$

• Determine the CDFs for X and Y given in the wheel example above when the probability of a measurable event E occurring is given by $P(E) = m(E)/2\pi$.

• Suppose a random variable X has as its distribution function

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1. \end{cases}$$

For the random variable $Y = 2X + 1$, determine its distribution function $F_Y(x)$.

• This is an interesting and somewhat challenging exercise. For any CDF, prove:

$$\text{(a)} \lim_{x \to -\infty} F_X(x) = 0. \qquad \text{(b)} \lim_{x \to +\infty} F_X(x) = 1.$$

It's also true that $F_X(x)$ is a nondecreasing function of x and is right continuous in the sense that $F_X(x) = \lim_{y \downarrow x} F_X(y)$ for every $x \in \mathbb{R}$. I'll prove these last two facts in class.

When a random variable's distribution function $F_X(x)$ is a differentiable function of x, its *probability density function*, or simply its PDF, will be denoted by $f_X(x)$ (as done in the text) and given by the formula

$$f_X(x) = \frac{d}{dx} F_X(x).$$

Note that knowing a random variable's PDF allows us to recover its CDF via the fundamental theorem of calculus

$$F_X(x) = F_X(x) - F_X(-\infty) = \int_{-\infty}^{x} f_X(y)\, dy.$$

Here's an example. Suppose $S = [0,1]$, $\mathcal{E}$ is the family of measurable subsets E of S, and $P(E) = m(E)$, $(= \int_0^1 \chi_E(x)\, dx)$. Let a random variable X be given by $X(s) = 2s+1$. Then calculate that

$$F_X(x) = P(X \le x) = m(\{s \in [0,1] : 2s + 1 \le x\}) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{2}(x-1) & \text{if } 1 \le x \le 3 \\ 1 & \text{if } x > 3, \end{cases}$$

$\Rightarrow$

$$f_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1/2 & \text{if } 1 < x < 3 \\ 0 & \text{if } x > 3. \end{cases} \quad \text{(I'll discuss differentiability issues in class.)}$$

• Use the probability space from the previous example to determine the CDF and PDF for the random variables: (a) $X(s) = 2s - 1$. (b) $X(s) = s(1-s)$. (Don't worry about points where the CDF is not differentiable.)