# 3: Random Variables and Distributions, Part I

PSTAT 120A: Summer 2022

Ethan P. Marzban
June 27, 2022

University of California, Santa Barbara

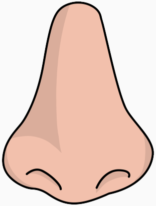
# Where We've Been

- Axioms of Probability, Probability Spaces, Counting
- Conditional Probabilities, independence, etc.

# Random Variables

# Leadup

- Let's consider again the experiment of tossing a coin twice.
- We saw previously that one possible outcome space is $\Omega = \{H, T\}^2 = \{(H, H),\ (H, T),\ (T, H),\ (T, T)\}$
- Suppose I am only interested in the *number of heads* I observed, not the actual configuration of heads and tails.
- In other words, I seek some summarizing quantity; specifically, one that takes an outcome and spits out the number of heads.
- Hm, takes in an element and spits out a number...
- Smells like a **function**!



# Random Variables

## Definition: Random Variable

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a **random variable** is a function that maps from $\Omega$ to $\mathbb{R}$. Oftentimes we use capital letters to denote random variables; for example,

$$X : \Omega \to \mathbb{R}$$

- So, in our coin tossing example, let $X$ denote the number of heads in my two coin tosses. Then:

$$X((H, H)) = 2; \quad X((H, T)) = 1; \quad X((T, H)) = 1; \quad X((T, T)) = 0$$

  - Or, equivalently, $(H, H) \mapsto 2; \quad (H, T) \mapsto 1; \quad (T, H) \mapsto 1; \quad (T, T) \mapsto 0$

# State Space

**Definition: State Space (Support)**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \mapsto \mathbb{R}$, we define the **state space** (sometimes called the **support**) of $X$ to be the image of $\Omega$. In other words, letting $S_X$ denote the state space of $X$, we have

$$S_X := X(\Omega) = \{y \in \mathbb{R} : y = X(\omega) \text{ for some } \omega \in \Omega\}$$

- So, in our coin tossing example where $X$ denotes the number of heads observed, then $S_X = \{0, 1, 2\}$.

# Classification of Random Variables

- We classify random variables based on their state space.

**Definition: Discrete/Continuous Random Variables**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$, we say $X$ is a **discrete random variable** (or just "$X$ is **discrete**") if its state space is at most countable; otherwise we say $X$ is a **continuous** random variable (or just "$X$ is **continuous**").

# Countable vs. Uncountable

- OK, so I guess it's time to finally address the "countable vs. uncountable" issue.
- Here is how I like to (intuitively) think about things. Clearly, both $\mathbb{Z}$ and $\mathbb{R}$ have an infinite number of elements.
- However, in the sense of subsets, $\mathbb{R}$ is "bigger" than $\mathbb{Z}$ (remember when we talked about comparing sets?) Therefore, it makes sense that $\mathbb{R}$ should be "bigger" than $\mathbb{Z}$ in the sense of cardinality as well.
- Additionally, between any two integers are an infinite number of real numbers!
- So, either way we cut it, it seems like the cardinality of $\mathbb{R}$ should be larger than that of $\mathbb{Z}$.
- This is why we say $\mathbb{Z}$ is **countably infinite**, whereas $\mathbb{R}$ is **uncountably infinite**.
- Intervals (closed, open, or half-open/half-closed) are also uncountably infinite.

- There is a way to make the notion of countable vs. uncountable more rigorous (and you do so in classes like MATH 8 or PSTAT 8), but we won't worry about that level of distinction for this class.

# Discrete vs. Continuous

- Returning to our coin tossing example where $X$ denotes the number of heads I observed– we saw that $S_X = \{0, 1, 2\}$ meaning $X$ is **discrete.**
- Suppose I break a stick of length 1 into two smaller pieces by picking a breakpoint at random along the length of the stick. If $L$ denotes the length of the shorter piece, then $S_L = [0, 1/2]$ which shows that $L$ is **continuous**.
- We will focus on Discrete Random Variables for now; then we'll turn our attention to continuous ones.

# Discrete Random Variables

# Leadup

- Note that our discussion thus far has been devoid of any mention of $\mathbb{P}$ (at least, beyond the notion of a probability space).
- Let's incorporate probabilities into the mix.

**Definition: Probability Mass Function (P.M.F.)**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X$, we define the **probability mass function** (or **p.m.f.**, for short) as

$$p_X(k) := \mathbb{P}(X = k)$$

for all values of $k \in \mathbb{R}$. Note that the $p_X(k)$ is nonzero only when $k \in S_X$, but $p_X(k)$ should be defined over the entire real line.

# PMF

- Let's return to our coin tossing example: recall that

$$X((H,H)) = 2; \quad X((H,T)) = 1; \quad X((T,H)) = 1; \quad X((T,T)) = 0$$

- Now, suppose the coin were fair; then we could utilize the classical definition of probability to construct the p.m.f. of $X$:

$$p_X(k) = \begin{cases} 1/4 & \text{if } k = 0 \\ 1/2 & \text{if } k = 1 \\ 1/4 & \text{if } k = 2 \\ 0 & \text{otherwise} \end{cases}$$

# PMF

- Now, I have glossed over something which we should perhaps examine a little more closely: what does $\mathbb{P}(X = k)$ really mean?
- That is, $\mathbb{P}$ only acts on events, so what does the event $\{X = k\}$ mean?
- Well, when we write $\{X = k\}$ we really mean "the set of all outcomes $\omega \in \Omega$ that get mapped to $k$, under $X$." In other words:

$$\{X = k\} := \{\omega \in \Omega : X(\omega) = k\}$$

- In fact, we can generalize this notation even further: for a set $B \subseteq \mathbb{R}$, we write

$$\{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\}$$

  For instance, we will write

$$\{X \leq k\} := \{\omega \in \Omega : X(\omega) \leq k\}$$

- So, for example, in our coin tossing problem,

$$p_X(1) = \mathbb{P}(X = 1) = \mathbb{P}\left(\{(H, T),\ (T, H)\}\right) = \frac{|\{(H, T),\ (T, H)\}|}{4} = \frac{2}{4} = \frac{1}{2}$$

# PMF

- By the way, we can also express PMF's in tabular format:

| $\boldsymbol{k}$ | 0 | 1 | 2 |
|---|---|---|---|
| $\boldsymbol{p_X(k)}$ | 1/4 | 1/2 | 1/4 |

- Sometimes, we can get lucky and even write our p.m.f. as a (somewhat) closed-form expression:

$$p_X(k) = \begin{cases} \binom{2}{k}\left(\frac{1}{2}\right)^2 & \text{if } k = 0, 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

# PMF

- So, to reiterate: the p.m.f. represents all the possible values a random variable can take, and the probability with which the random variable attains those values.
- P.M.F's can be expressed in three possible ways: using a piecewise-defined function, using a table, or, sometimes, using a closed-form expression (with a "0 otherwise" case)

# PMF

- We have a set of tools we can use to verify whether or not a specified function is in fact the p.m.f. of a random variable.

**Theorem: Verifying that a Function is a PMF**

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and a function $p_X : \mathbb{R} \to \mathbb{R}$. If $p_X$ satisfies the following two conditions:

(1) **Nonnegativity:** $p_X(k) \geq 0$ for all $k \in \mathbb{R}$

(2) **Summing to Unity:** $\sum_k p_X(k) = 1$

then $p_X$ is the p.m.f. of a random variable.

# Example

Show that the function

$$p_X(k) = \begin{cases} \left(\frac{1}{2}\right)^k & \text{if } k = 1, 2, \cdots \\ 0 & \text{otherwise} \end{cases}$$

is a valid probability mass function.

- First note that $(1/2)^k \geq 0$ for every $k \in \{1, 2, \cdots\}$; therefore condition (1) is satisfied.
- Additionally,

  $$\sum_k p_X(k) = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = \frac{\left(\frac{1}{2}\right)}{1 - \left(\frac{1}{2}\right)} = 1 \checkmark$$

  Therefore condition (2) is satisfied.
- Thus, since both conditions are satisfied, $p_X(k)$ is the p.m.f. of a random variable.
- You need to check both conditions! It's not enough to just say "sums to unity;" you also need to check nonnegativity.

# Probabilities from PMF's

- Once we have a p.m.f. of a random variable, we can compute probabilities by summing up values of the p.m.f.:

**Theorem: Probabilities from PMF's**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X$ with p.m.f. $p_X(k)$, we have

$$\mathbb{P}(X \in B) = \sum_{\{x : x \in B\}} p_X(k)$$

- For example, in our coin tossing example, suppose we want the probability of observing at most 1 heads: then we use

$$\mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

**Definition: CMF**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X$, we define the **cumulative mass function** (or **c.m.f.**, for short) to be

$$F_X(x) := \mathbb{P}(X \leq k)$$

- So, on the previous slide, for instance, we found $F_X(1)$.

# Example (Chalkboard)

On a table, I have three boxes. I know that 2 of the 3 boxes contain a reward of \$100, but the other box will actually *cost* me \$100. Suppose I open two boxes at random (note that once a box is opened it cannot be re-opened). Letting $W$ denote my net winnings, what is the p.m.f. of $W$?

# Expectation, and Moments

# Expectation

## Definition: Expected Value

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a discrete random variable $X$, we define the **expected value** (or just **expectation**) of $X$ to be

$$\mathbb{E}[X] := \sum_k k \cdot p_X(k)$$

- So, for instance, in our coin tossing example

$$\begin{aligned}\mathbb{E}[X] &= 0 \cdot p_X(0) + 1 \cdot p_X(1) + 2 \cdot p_X(2) \\ &= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1\end{aligned}$$

- This represents the "average" number of heads.

- **Key Point:** $\mathbb{E}[X]$ may not be in the state space of $X$. As an example: consider rolling a fair six-sided die and letting $X$ denote the number that is showing. Then $\mathbb{E}[X] = 7/2$ (I leave it to you to show this), despite the fact that $S_X = \{1, 2, 3, 4, 5, 6\}$.

# Expectation of a Function

**Theorem: Law of the Unconscious Statistician (LOTUS)**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a discrete random variable $X$, we have

$$\mathbb{E}[g(X)] = \sum_k g(k) \cdot p_X(k)$$

- Note that plugging in $g(k) = k$ yields our familiar notion of expectation.
- Additionally, note that this is a *theorem*; it is not a fact, but rather something that must be proven. (We omit the proof for now).
- Also, you may ask: what does $g(X)$, a function of a random variable mean? Well, we'll discuss this in greater detail in a later lecture. For now, here is some intuition: $g : \mathbb{R} \to \mathbb{R}$ and $X : \Omega \to \mathbb{R}$ meaning $(g \circ X) : \Omega \to \mathbb{R}$; that is, $(g \circ X)$ is in fact a random variable!
  - Again, more on this in a later lecture.

# Moments

**Definition: $n^{\text{th}}$ Moment of a Random Variable**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a discrete random variable $X$, we define the $\boldsymbol{n}^{\textbf{th}}$ **moment** of $X$ to be

$$\mu_n := \mathbb{E}[X^n] = \sum_k k^n \cdot p_X(k)$$

- Note that $\mathbb{E}[X]$ is simply the first moment of $X$. For this reason, we often notate $\mathbb{E}[X]$ by $\mu$.

# Variance

- Suppose we are interested in a measure of the "spread" of a random variable.
- A sensible measure would be the "average distance from the center." This motivates our definition of variance:

**Definition: Variance**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X$, we define the **variance** of $X$ to be

$$\mathrm{Var}(X) := \mathbb{E}\left\{[X - \mathbb{E}(X)]^2\right\}$$

It turns out that

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - [\mathbb{E}(X)]^2$$

i.e. the variance of $X$ is equal to the second moment, minus the square of the first moment.

- Sometimes we are interested in the square root of variance; we call this quantity the **standard devitaion**. In other words,

$$\mathrm{SD}(X) := \sqrt{\mathrm{Var}(X)} = \sqrt{\mathbb{E}\left\{[X - \mathbb{E}(X)]^2\right\}}$$

# Summary

- That's a lot of information, and a lot of terms!
- Let me try and summarize some things.
- We start with the notion of a random variable $X$, which is a mapping from $\Omega$ to $\mathbb{R}$.
- The **state space** of $X$ is the image of $\Omega$, under $X$: $S_X := X(\Omega)$
- The **probability mass function** (p.m.f.) is a function that takes in a real number $k$ and outputs the probability that $X = k$ (i.e. the probability of the set of all outcomes that get mapped to $k$, under $X$)
  - A p.m.f. must be nonnegative, and sum to unity.
  - The **cumulative mass function** (c.m.f.) $F_X(x)$ at a point $x$ is the sum of $p_X(k)$ for which $k$ is at most $x$.
- The **expected value** (or **expectation**) of a random variable gives a measure of an "average" value of $X$.
- We can compute expectations of functions of random variables (which are, in fact, random variables) using the **Law of the Unconscious Statistician**.
- The $\boldsymbol{n}^{\text{th}}$ **moment** of a random variable $X$ is defined to be $\mu_n := \mathbb{E}[X^n]$
  - Thus, the first moment is simply the expectation
  - The **variance** of $X$, a measure of "spread," is related to the second moment of $X$.

# Comprehensive Example

Let $X$ be a random variable with p.m.f. given as below:

| $k$ | $-1$ | $1$ | $2$ |
|---|---|---|---|
| $p_X(k)$ | 2/5 | 1/4 | 7/20 |

(a) Find $\mathbb{E}[X]$

(b) Find $\mathbb{E}[X^2]$

(c) Find $F_X(x)$, the c.m.f. of $X$.

# CMF's, again

- There is something you might notice about the c.m.f. of $X$ in the previous example: it is a step function.
- This is in fact true of *all* discrete random variables: in other words:

**Fact: C.M.F.'s**

The c.m.f. of a discrete random variable $X$ is a step function, with points of discontinuity corresponding to the points in the state space of $X$ and with the magnitudes of the jump discontinuities corresponding to the values of the p.m.f. of $X$.

# Example

Suppose $X$ is a random variable with c.m.f. given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.3 & \text{if } 0 \leq x < 2 \\ 0.7 & \text{if } 2 \leq x < 4 \\ 1 & \text{if } x \geq 4 \end{cases}$$

- $p_X(0) = 0.3 - 0 = 0.3$
- $p_X(2) = 0.7 - 0.3 = 0.4$
- $p_X(4) = 1 - 0.7 = 0.3$