Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture # 8

Probability Theory

What We've Done So Far

Axioms of probability, rules for computing probabilities of events

What Remains

Notion of event subsumed by that of *Random Variable*

Random Variable

A more abstract way of representing a random experiment: focus only on the important information, not the irrelevant details.

A New Way of Thinking about Populations

- *No longer* think of a population as a list of N objects.
- *Represent* a population by a probability model using the language of Random Variables.
- Re-express population parameters as *features* of the Random Variables that *represent* a population.
- E.g. if we say "μ is the mean of a random variable X" the idea is that μ is the mean of a *population represented* by the random variable X.

Why are we getting rid of population size N?

1. Don't know N for real-world populations.
2. For sampling N is irrelevant: all that matters are the *relative frequencies* in the population.

Key Innovation

Treat population relative frequencies as *probabilities* rather than counts divided by N.

Why Does This Make Sense?

We defined probability as long-run relative frequency. The idea of "long-run" here is repeated sampling from the population.

Treating Population Relative Frequencies as Probabilities

Discrete Data ⇒ Discrete Random Variables

It's obvious what the "right probability" is in this case. For example if 52000 people in a population of 100000 voted for Obama, we'd represent this as a probability of 0.52.

Continuous Data ⇒ Continuous Random Variables

Here things are more complicated. What proportion of people have a height of exactly 62.374827 inches? To get around this problem we'll proceed by an analogy to histograms.

The first thing to know about Random Variables is that they are neither random nor variables...

Random Variable (RV)

Fixed function from sample space S to real numbers ($X\colon S \mapsto \mathbb{R}$).

Turns *basic outcomes* of random experiment into *numbers*.

Realization

Particular value that RV takes on, i.e. the result of applying the function defined by X to the *outcome* of the random experiment.

We write $X = x$. Note that $\{X = x\}$ is an *event*.

Support Set

The set of all possible realizations of a RV.

Notation

RVs denoted by capital letters, e.g. X, Y, Z, their realizations by the corresponding lowercase letters, e.g. x, y, z.

Random Variables

$$X : S \mapsto \mathbb{R}$$

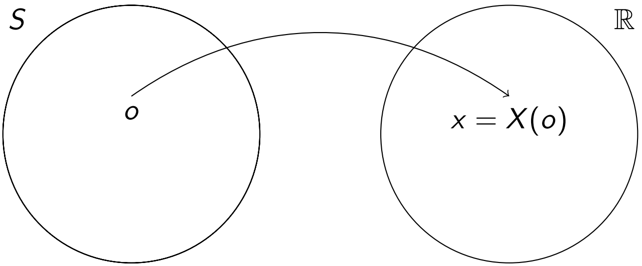


Figure : A Random Variable X is a fixed, i.e. deterministic, function that maps each basic outcome in the sample space S to a real number. Here o is a basic outcome and x is a realization, equal to $X(o)$.

Important Distinction

Although a Random Variable is a deterministic function, the *values it takes on* are random since its input, the outcome of a random experiment, is random!

Example: Coin Flip Random Variable

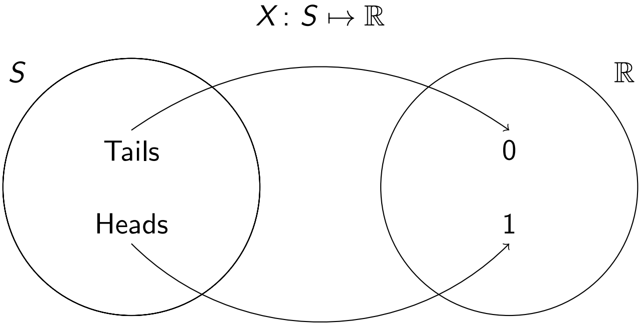


Figure : This random variable maps the outcomes of flipping a coin $\{\text{Heads}, \text{Tails}\}$ to the set $\{0, 1\} \subseteq \mathbb{R}$. Hence, its support is $\{0, 1\}$

Types of Random Variables

Discrete RVs

Takes on a finite or countably infinite number of different values, e.g. $\{-1, 0, 1\}$ or $\{0, 1, 2, 3, 4, 5, 6, \ldots ...\}$

Continuous RVs

Takes on an uncountably infinite number of different values, e.g. all real numbers in the interval $[0, 1]$ or $(-\infty, \infty)$.

Discrete Random Variables I

Probability Mass Function (pmf)

Probability that a Discrete RV takes on the particular value x as a function of x:

$$p(x) = P(X = x)$$

Plug in a realization x, get out a probability $p(x)$.

Probability Mass Function for Coin Flip RV

$$X = \begin{cases} 0, \text{Tails} \\ 1, \text{Heads} \end{cases}$$

$$\begin{aligned} p(0) &= 1/2 \\ p(1) &= 1/2 \end{aligned}$$

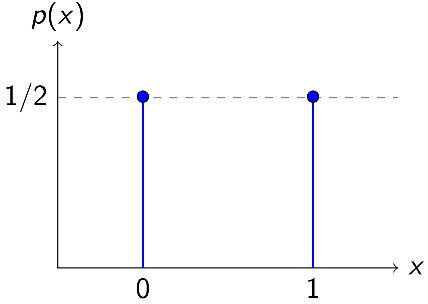
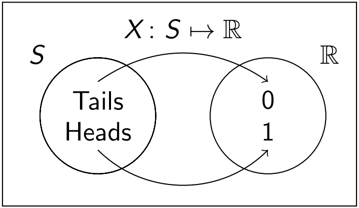


Figure : Plot of pmf for Coin Flip Random Variable

Where did this come from?



Support Set $\{0, 1\}$

The only possible realizations are $X = 0$ and $X = 1$

Events

Tracing the arrows backwards, $\{X = 0\}$ corresponds to the event "Tails" while $\{X = 1\}$ corresponds to the event "Heads."

Probabilities

Since $P(\text{Heads}) = P(\text{Tails}) = 1/2$, $P(X = 0) = P(X = 1) = 1/2$

Important Note about Support Sets

Whenever you write down the pmf of a RV, it is crucial to also write down its Support Set. Recall that this is the set of *all possible realizations for a RV*. Outside of the support set, all probabilities are zero. In other words, the pmf is only defined on the support.

Properties of Probability Mass Functions

If $p(x)$ is the pmf of a random variable X, then

(i) $0 \leq p(x) \leq 1$ for all x

(ii) $\displaystyle\sum_{\text{all } x} p(x) = 1$

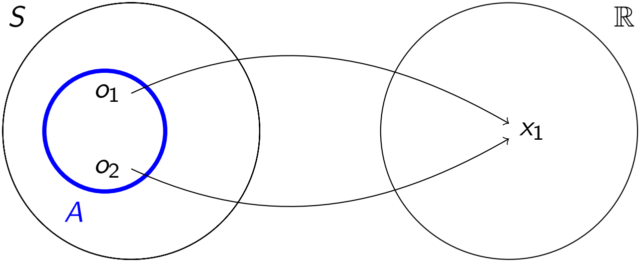
where "all x" is shorthand for "all x in the support of X."

But Where Do These Come From?

The key point is that $\{X = x\}$ is an *event*. We proceed by working backwards until we are in familiar territory.

Which Event Corresponds to $\{X = x_1\}$?

$$X : S \mapsto \mathbb{R}$$



Define $A = \{o \in S : X(o) = x_1\}$. Since it is a subset of the sample space, A is an event. Since A contains *all* the basic outcomes that map to x_1, $A = \{X = x_1\}$, hence $P(A) = P(X = x_1)$.

Link Between Axioms of Probability and Random Variables

1. We can equate $\{X = x\}$ with an event A in S for any realization x in the support of X. Thus $P(X = x) = p(x)$ is a bona fide probability and

$$0 \leq p(x) \leq 1$$

2. Since X is a *function*, the events $\{X = x\}$ are mutually exclusive and collectively exhaustive, hence:

$$\sum_{\text{all } x} p(x) = \sum_{\text{all } x} P(X = x) = P(S) = 1$$

Is This Possible? (A = Yes, B = No)

$$X : S \mapsto \mathbb{R}$$

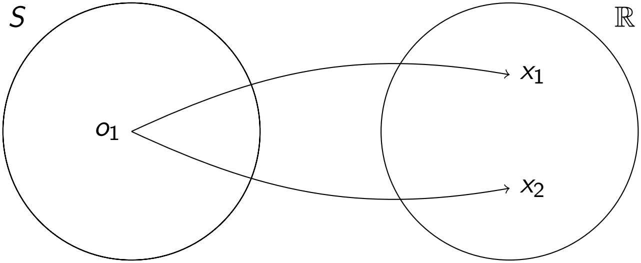
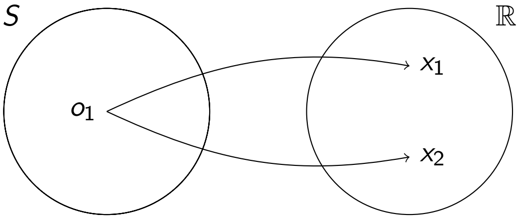


Figure : This diagram shows the random variable X mapping the basic outcome o_1 to two different realizations: x_1 and x_2. Is this allowed?

This is NOT POSSIBLE

$$X : S \mapsto \mathbb{R}$$



- A function cannot map a single input to two different outputs! You may know this as the “vertical line test.”
- We have to rule this situation out since it would mean that $\{X = x_1\}$ and $\{X = x_2\}$ are *not mututally exclusive*.

Is This Possible? (A = Yes, B = No)

$$X : S \mapsto \mathbb{R}$$

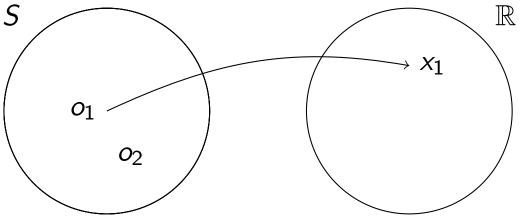
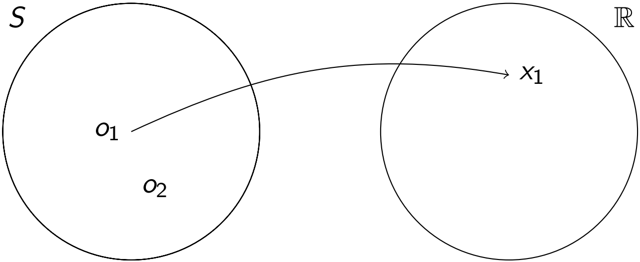


Figure : This diagram shows the random variable X mapping the basic outcome o_1 to the realization x_1. The basic outcome o_2 doesn't get mapped anywhere. Is this allowed?

This is NOT POSSIBLE

$$X : S \mapsto \mathbb{R}$$



- A function associates *every* value in its domain with a value in its range: o_2 has to map to something!
- We have to rule this situation out since it would mean that the events $\{X = x\}$ are *not collectively exhaustive*.

Cumulative Distribution Function (CDF)

This Def. is the same for continuous RVs.

The CDF gives the probability that a RV X does not exceed a specified threshold x_0, as a function of x_0

$$F(x_0) = P(X \leq x_0)$$

Important!

The threshold x_0 is allowed to be *any real number*. In particular, it doesn't have to be in the support of X!

CDF of Coin Flip Random Variable

$$X = \begin{cases} 0, \text{Tails} \\ 1, \text{Heads} \end{cases}$$

$$F(x_0) = \begin{cases} 0, & x_0 < 0 \\ \frac{1}{2}, & 0 \le x_0 < 1 \\ 1, & x_0 \ge 1 \end{cases}$$

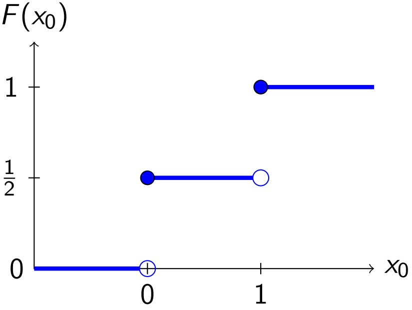


Figure : CDF for Coin Flip Random Variable

Where did we get this from?

Discrete RVs: Sum the pmf to get the CDF

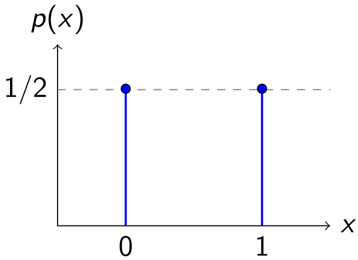
$$\boxed{F(x_0) = \sum_{x \leq x_0} p(x)}$$

Proof

Use the fact that the events $\{X = x\}$ are mutually exclusive:

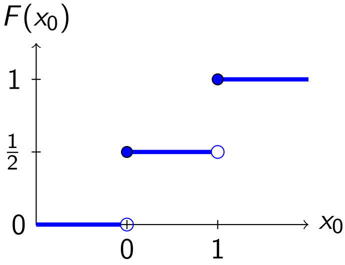
$$F(x_0) = P(X \leq x_0) = P\left(\bigcup_{x \leq x_0} \{X = x\}\right) = \sum_{x \leq x_0} P(X = x) = \sum_{x \leq x_0} p(x)$$

Probability Mass Function



$$
\begin{aligned}
p(0) &= 1/2 \\
p(1) &= 1/2
\end{aligned}
$$

Cumulative Dist. Function



$$
F(x_0) = \begin{cases} 0, & x_0 < 0 \\ \frac{1}{2}, & 0 \leq x_0 < 1 \\ 1, & x_0 \geq 1 \end{cases}
$$

Properties of CDFs

These are also true for continuous RVs.

1. $\lim_{x_0 \to \infty} F(x_0) = 1$
2. $\lim_{x_0 \to -\infty} F(x_0) = 0$
3. Non-decreasing: $x_0 < x_1 \Rightarrow F(x_0) \leq F(x_1)$
4. Right-continuous ("open" versus "closed" on prev. slide)

Since $F(x_0) = P(X \leq x_0)$, we have $0 \leq F(x_0) \leq 1$ for all x_0

Bernoulli Random Variable – Generalization of Coin Flip

Support Set

$\{0, 1\}$ – 1 traditionally called "success," 0 "failure"

Probability Mass Function

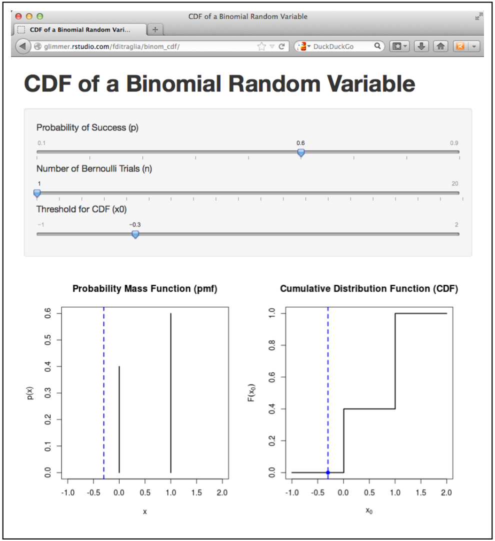
$$
\begin{aligned}
p(0) &= 1 - p \\
p(1) &= p
\end{aligned}
$$

Cumulative Distribution Function

$$
F(x_0) = \begin{cases} 0, & x_0 < 0 \\ 1 - p, & 0 \leq x_0 < 1 \\ 1, & x_0 \geq 1 \end{cases}
$$

http://glimmer.rstudio.com/fditraglia/binom_cdf/

Set the second slider to 1 and play around with the others.



Average Winnings Per Trial

If the realizations of the coin-flip RV were payoffs, how much would you expect to win per play *on average* in a long sequence of plays?

$$X = \begin{cases} \$0, \text{Tails} \\ \$1, \text{Heads} \end{cases}$$

Fair Price

If I were *bankrolling* a long sequence of trials of this game, how much should I charge per play so that I break even in the long run?

$$X = \begin{cases} \$0, \text{Tails} \\ \$1, \text{Heads} \end{cases}$$

Expected Value:

Probability-Weighted Average

of Realizations

Expected Value (aka Expectation)

The expected value of a discrete RV X is given by

$$E[X] = \sum_{\text{all } x} x \cdot p(x)$$

Treating the random variable X as a probability model for a population, we equate $E[X]$ with the population mean μ.

Expected Value of Bernoulli Random Variable

Suppose $p = 1/4$ since you can only enter numeric results on the clicker.

$$X = \begin{cases} 0, \text{Failure: } 1 - p \\ 1, \text{Success: } p \end{cases}$$

Expected Value of Bernoulli Random Variable

Suppose $p = 1/4$ since you can only enter numeric results on the clicker.

$$X = \begin{cases} 0, \text{Failure: } 1-p \\ 1, \text{Success: } p \end{cases}$$

$$\sum_{\text{all } x} x \cdot p(x) = 0 \cdot (1-p) + 1 \cdot p = p$$

Next Time: Midterm I