PRIMES Circle Presentation

Bella, Niyathi, Iris – May 14, 2023

Quick Intro

Definitions

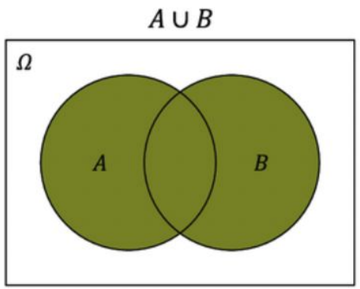
Sample Space: The set of all possible outcomes, denoted by S.

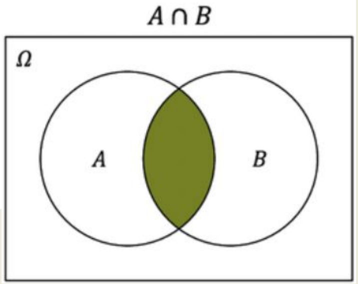
Event: A subset of the sample space. Consists of possible outcomes of the experiment, denoted by A.

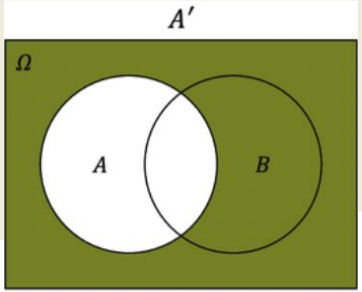
Union (AUB): outcomes that are either in A or in B or in both A and B

Intersection (AB): outcomes that are in both A and B, also denoted as A∩B

Complement: (A’): outcomes in the sample space S that are not in A







Axioms of Probability

All probability of an event has to follow the 3 Axioms of Probability.

Axiom 1: $0 \leq P(E) \leq 1$

Axiom 2: $P(S) = 1$

Axiom 3:

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

01

Conditional Probability & Independence

Conditional Probability

Given two events E and F, The conditional probability is denoted by P(E | F).

Essentially this conditional probability is the probability of both events happening with the new sample space F, so the definition and general formula for P (E | F) if P (F) > 0 is:

$$P(E|F) = \frac{P(EF)}{P(F)}$$

□ both events happening
□ event F happening

P(F) has to be greater than 0 because a probability can't be negative, and P(E | F) would be undefined if P(F) equals 0.

Multiplication Rule

Gives the probability of the intersection of two events (both events happening).

$$P(EF) = P(F)P(E|F)$$

A generalization of the above equation gives the probability of the intersection of n events.

$$P(E_1E_2E_3\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1\cdots E_{n-1})$$

Bayes' Formula

$$P(F_j|E) = \frac{P(EF_j)}{P(E)}$$

This formula is extremely useful because sometimes it is much easier to calculate the probability of an event when we know whether or not another event it depends on has occurred.

Independence

Two events E and F are independent if P(E | F) = P(E)
In other words when P(EF) = P(E) * P(F)

We can extend independence to more than two events. A set of events is independent if every subset of those events is independent.

$$P(E_{1'}, E_{2'}, \ldots, E_{r'}) = P(E_{1'})P(E_{2'})\ldots P(E_{r'})$$

Insurance Problem

- Two types of people: those who are accident prone and those who are not
- Accident-prone person will have an accident in a 1-year period with probability 0.4, and a person who is not accident prone has probability 0.2
- 30% of the population is accident prone

Suppose that a new policyholder has an accident within a year of purchasing a policy. What is the probability that they are accident prone?

Problem

Bayes' Formula

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Accident-prone person having an accident: 0.4
Non-accident-prone person having an accident: 0.2
Percent of population that is accident-prone: 0.3

A – the event that the policyholder is accident prone
B – the event that the policyholder will have an accident within a year of purchasing the policy

$$\begin{aligned} P(A|B) &= \frac{P(B|A) \cdot P(A)}{P(B)} \\ &= \frac{(0.4)(0.3)}{P(B)} \\ &= \frac{(0.4)(0.3)}{(0.4)(0.3) + (0.2)(0.7)} \\ &= \frac{0.12}{0.26} \\ &\approx 0.462 \end{aligned}$$

02

Discrete RVs

What is a Random Variable?

- A variable whose value is unknown, represented by a function that assigns values to each of an experiment's outcomes.
- Often designated by letters
- Can be classified as **discrete** or **continuous**
- Used in probability to quantify the outcome of a random experiment

    Ex: when flipping a coin, Heads=1 & Tails=0

Definitions

Discrete RV: A discrete random variable is a variable that can take any whole number values as outcomes of a random experiment.

- In real life, a discrete random variable could be represented as the number of passengers on a train or the number of defective computers out of a group of 100

It is defined by the probability mass function (PMF), **$p(a) = P\{X = a\}$**, which is always positive over a countable number of a.

There are various types of discrete random variables, and we are going to be discussing the Binomial, Geometric, and the Poisson RVs

Expected Value & Variance

Expected Value: the predicted value of a random variable

- calculated as the sum of all possible values each multiplied by the probability of its occurrence
- essentially the weighted average of all the outcomes

$$E[X] = \sum_{x:p(x)>0} xp(x)$$

Variance: degree of spread in the random variable

- calculated as the average of squared deviations from the mean (denoted as μ), with formulas

$$\text{Var}(X) = E[(X - \mu)^2] \quad \text{OR} \quad \text{Var}(X) = E[X^2] - (E[X])^2$$

People often use expected value and variance to calculate the expected ROI for investments they make

Binomial RV

Definition: the number of predicted "successes" in an experiment consisting of N trials – the experiment can only take on 2 values (success or failure)

Parameters: n = the number of trials, P = the probability of 1 trial succeeding, K = the number of desired successes

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Expected Value: n * p

Variance: n * p * (1–p)

Binomial Example

Let's say you buy some lottery tickets- 10 of the same kind. The probability that you get a winning ticket is 0.02. What is the probability that you get 3 winning tickets? If the tickets cost you $5 each, and each winning ticket wins you $100, was it worth it to buy those tickets- or were you scammed?

Let X be the number of winning tickets you get. We can apply the formula $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$, letting n=10, k=3, and p=0.02, to find the desired probability:

$$P\{X=3\} = {}_{10}C_3\,(0.03)^3(1-0.03)^{10-3} = 0.0026.$$

Now, the expected value is simply n*p, or 10*0.03= 0.3 successes, and so your expected ROI is $30 dollars-which means you're probably scammed :D

Geometric RV

Definition: the number of predicted experiments that will be performed before reaching some number of "successes" (out of 2 options: success or failure)

Parameter: p = the probability of success for 1 trial, n (or sometimes k) = the number of the final (successful) trial.

Expected Value: 1/p – you need p * 1/p to get 100% chance of success

Variance: $(1-p)/(p^2)$

$$P(X = n) \;=\; p(1-p)^{n-1}$$

$$P(X > n) \;=\; (1-p)^{n}$$

Geometric RV Example

Example Problem: 3 out of 75 bulbs are defective, probability that the first defective light bulb is 6th?

#1: Define p, the probability that a light bulb is defective, and k, the number of trials needed to reach the "successful trial"

$$p = \frac{3}{75} = 0.04$$

$$k = 6$$

#2: Plug into formula

$$P(X = n) = p(1-p)^{n-1}$$

$$P(X = k) = p(1-p)^{k-1}$$

$$P(X = 6) = 0.04(1-0.04)^{6-1}$$

$$P(X = 6) = 0.04(0.96)^{5} = 0.0326$$

Poisson RV

Definition: the number of times an event is likely to occur over a specific interval (time, pages of a book, etc.)

- It is also a good approximation to binomial RVs for very large values of n and very small values of p

Parameter: λ = mean number of events in a given interval

Expected Value: λ

Variance: λ

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Poisson Example

Let's say that the number of typos on any given slide of this presentation follows a Poison distribution of λ=0.5. What is the probability that there is at least one error on this slide?

To solve this, we can let X be the Poisson RV, which is the number of errors on this slide.

Then, we can apply the formula and simplify to get

$$P\{X \geq 1\} = 1 - P\{X = 0\} = 1 - e^{-1/2} \approx .393$$

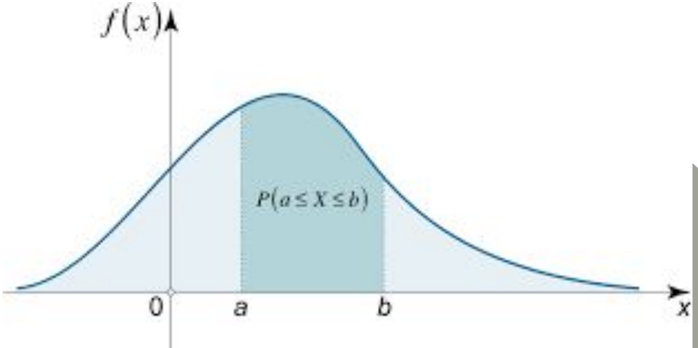
Note: $e^{-½}$ is the simplified result of $(0.5)^0(e^{-½})/0!$.

03

Continuous RVs

Definitions:

- A **continuous random variable** is a random variable that is defined over a range of values
  - In real life, they are usually measurements (height, time required to run 100 meters, etc.)
- A continuous RV is defined by a **probability density function f(x) (PDF),** a non-negative function defined over all real numbers that functions similarly to the PMF in the discrete case
- Finding the probability of a continuous RV being within an interval (a,b) requires finding the **area under its probability curve**– also known as taking an **integral** of the function between a and b.



Quick Example on Continuous RVs!

Since we talked about expected value and variance earlier, and since the continuous RV is just another random variable, let's try to find E[X] and Var(X) of a continuous RV X, defined by

$$f(x) = \begin{cases} 2x & \text{if} \quad 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Here are some formulas that we are going to use:

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx \qquad \mathrm{Var}(X) = E[X^2] - (E[X])^2$$

Quick Example on Continuous RVs!

Now let's take a look at the solution:

$$
\begin{aligned}
E[X] &= \int x f(x)\, dx \\
&= \int_0^1 2x^2\, dx \\
&= \frac{2}{3}
\end{aligned}
$$

Quick Example on Continuous RVs!

Now let's take a look at the solution:

Solution. We first compute $E[X^2]$.

$$
\begin{aligned}
E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x)\,dx \\
&= \int_0^1 2x^3\,dx \\
&= \frac{1}{2}
\end{aligned}
$$

Hence, since $E[X] = \frac{2}{3}$, we obtain

$$
\mathrm{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}
$$

04

Joint Probability

Joint distribution functions

A **joint distribution function** is simply one that describes the probability of **two or more random variables** in one function.

For any two random variables **X** and **Y,** the **joint probability distribution function** (which describes the related probabilities of X & Y) is defined as

$$F(a,b) = P\{X \leq a, Y \leq b\} \quad -\infty < a,b < \infty$$

And their **joint probability mass function** is defined as

$$p(x,y) = P(X = x, Y = y)$$

Joint RVs Example (Poisson)

Radioactive particles reach a Geiger counter (a machine that counts the number of radioactive particles) according to a Poisson process at a rate of $\mu = 0.8$ particles per second.
What is the probability that the Geiger counter detects (exactly) 1 particle in the next second and 3 or more in the next 4 seconds?

Joint Probability: $P(1 \text{ particle in } (0,1) \text{ AND } 3 \text{ or more particles in } (0,4))$

$= P(1 \text{ particle in } (0,1) \text{ AND } 2 \text{ or more particles in } (1,4)).$

- f_1 = the number of particles in $(0,1)$ follows a $\text{Poisson}(\mu = 0.8)$ distribution, and
- f_2 = the number of particles in $(1,4)$ follows a $\text{Poisson}(\mu = 2.4)$ distribution.

Plug in values!

$$f_1(1) \cdot (1 - f_2(0) - f_2(1)) = e^{-0.8}\frac{0.8^1}{1!} \cdot \left(1 - e^{-2.4}\frac{2.4^0}{0!} - e^{-2.4}\frac{2.4^1}{1!}\right) \approx .2486.$$

Joint Distribution (Independent)

When computing a joint distribution, the two variables are independent of each other when

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$

In simple words, X and Y are independent if knowing the value of one does not change the value of the other.

If the variables are independent, the desired probabilities are multiplied, in accordance to the laws of conditional probability.

Joint Distribution: Discrete

- Suppose X and Y are two discrete random variables and that X takes values from $\{x_1, x_2, \ldots, x_n\}$ and and Y takes values $\{y_1, y_2, \ldots, y_m\}$
- The ordered pair (X, Y) take values in the product $\{(x_1, y_1),(x_2, y_2), \ldots (x_n, y_m)\}$
- The **joint probability mass function** (joint pmf) of X and Y is the function $p(x_i, y_j)$ giving the probability of the joint outcome $X = x_i$, $Y = y_j$.
- This is organized in a joint probability table
- The total probability must equal 1, represented by a double sum

$$\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) = 1$$

$X\backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_m
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$\cdots$	$p(x_1, y_j)$	$\cdots$	$p(x_1, y_m)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$\cdots$	$p(x_2, y_j)$	$\cdots$	$p(x_2, y_m)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_i	$p(x_i, y_1)$	$p(x_i, y_2)$	$\cdots$	$p(x_i, y_j)$	$\cdots$	$p(x_i, y_m)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	
x_n	$p(x_n, y_1)$	$p(x_n, y_2)$	$\cdots$	$p(x_n, y_j)$	$\cdots$	$p(x_n, y_m)$

Joint Distribution: Discrete Example

- Suppose you roll two dice. Let X be the value on the first die and let T be the total on both dice. This would be represented in the joint probability table as follows

$X \backslash T$	2	3	4	5	6	7	8	9	10	11	12
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36

Joint Distribution: Continuous

- Continuous case: replace discrete sets of values by continuous intervals, and use the sums by integrals to compute

- If X takes values in [a, b] and Y takes values in [c, d] then the pair (X,Y) takes values in the product [a, b] × [c, d]

- The total probability must equal 1, expressed as a double integral:

$$\int_c^d \int_a^b f(x, y)\, dx\, dy = 1$$

- We will not be providing a worked example for this, as it involves a lot of calculus

05

Markov Chains

Markov Chain Sequence

Consider an event with different **states** it could be in. A Markov chain is a model describing a **sequence** of events when the probability of each event depends **only** on the probability of the previous event.

A sequence of random variables X_0, X_1, ..., can be in states {0, 1, ..., M}. X_n is the state at time n, and the system is in state i at time n if $X_n = i$.

Example:
Any day in the town Lexington can be sunny, cloudy, rainy, or windy.
The sequence of days are X_0, X_1, ...; the states are sunny, cloudy, rainy, or windy; the date would be time n.

Transitional Probabilities

The sequence of random variables X_0, X_1, ..., is said to form a Markov chain if each time the system is in state i, there is a fixed probability, P_{ij}, that the system will next be in state *j*.

The values P_{ij}, $0 \leq i \leq M$, $0 \leq j \leq N$ are called *transition probabilities* of a markov chain if they satisfy:

1. $P_{ij} \geq 0$ ← The probability of being in state j if currently in state i can not be negative
2. $\sum_{j=0}^{M} P_{ij} = 1$ ← An event has 100% probability of transitioning into the next event
3. $i = 0, 1, ..., M$ ← An event could be in states 0, 1, ..., *M*

Transitional Probabilities continued

The transition probabilities can be represented as

$P\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0\} = P_{ij}$ for all $i_0, \ldots, i_{n-1}, i, j$

Example:
In Lexington, if today is sunny, tomorrow has 0.5 chance of being sunny, 0.15 chance of being cloudy, 0.1 chance of being rainy, and 0.25 chance of being windy.

- If today is cloudy, tomorrow has 0.2 chance of being sunny, 0.15 chance of being cloudy, 0.5 chance of being rainy, and 0.15 chance of being windy.
- If today is rainy, tomorrow has 0.25 chance of being sunny, 0.5 chance of being cloudy, 0 chance of being rainy, and 0.25 chance of being windy.
- If today is windy, tomorrow has 0.6 chance of being sunny, 0.2 chance of being cloudy, 0 chance of being rainy, and 0.2 chance of being windy.

Transition Matrix

However, it is more convenient to arrange the transition probabilities P_{ij} in a matrix, and this matrix is called the transition probability matrix.

$$\begin{bmatrix} P_{00} & P_{01} & \cdots & P_{0M} \\ P_{10} & P_{11} & \cdots & P_{1M} \\ \vdots & & & \\ P_{M0} & P_{M1} & \cdots & P_{MM} \end{bmatrix}$$

For example, the transition matrix of weather in Lexington is:

Weather of next day

Weather of current day

$$\begin{array}{r c} & \begin{array}{cccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} & \text{Windy} \end{array} \\ \begin{array}{r} \text{Sunny} \\ \text{Cloudy} \\ \text{Rainy} \\ \text{Windy} \end{array} & \begin{bmatrix} 0.5 & 0.15 & 0.1 & 0.25 \\ 0.2 & 0.15 & 0.5 & 0.15 \\ 0.25 & 0.5 & 0 & 0.25 \\ 0.6 & 0.2 & 0 & 0.2 \end{bmatrix} \end{array}$$

*Notice how each row adds up to 1

Transformation Matrix

The transition matrix can be used as a transformation matrix. The transformation matrix can be used in many different ways.

For example, the weather in Lexington can be represented as

$$T = \begin{bmatrix} 0.5 & 0.15 & 0.1 & 0.25 \\ 0.2 & 0.15 & 0.5 & 0.15 \\ 0.25 & 0.5 & 0 & 0.25 \\ 0.6 & 0.2 & 0 & 0.2 \end{bmatrix}$$

The probability of weather in Lexington in *n* days is:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} T^n$$

Transformation Matrix continued

Continuing with our example, the probability of each weather in 2 days (if day 1 is sunny) is:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \left(\begin{bmatrix} 0.5 & 0.15 & 0.1 & 0.25 \\ 0.2 & 0.15 & 0.5 & 0.15 \\ 0.25 & 0.5 & 0 & 0.25 \\ 0.6 & 0.2 & 0 & 0.2 \end{bmatrix} \cdot \begin{bmatrix} 0.5 & 0.15 & 0.1 & 0.25 \\ 0.2 & 0.15 & 0.5 & 0.15 \\ 0.25 & 0.5 & 0 & 0.25 \\ 0.6 & 0.2 & 0 & 0.2 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.455 & 0.1975 & 0.125 & 0.2225 \\ 0.345 & 0.3325 & 0.095 & 0.2275 \\ 0.375 & 0.1625 & 0.275 & 0.1875 \\ 0.46 & 0.16 & 0.16 & 0.22 \end{bmatrix}$$

$$= \begin{bmatrix} 0.455 & 0.1975 & 0.125 & 0.2225 \end{bmatrix}$$

So there's 45.5% chance it will be sunny again, 19.75% it will be cloudy, 12.5% it will be rainy, and 22.25% it will be windy in 2 days.

Calculations and Observations

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} T^3 = \begin{bmatrix} 0.432 & 0.205 & 0.144 & 0.219 \\ 0.399 & 0.195 & 0.201 & 0.205 \\ 0.401 & 0.256 & 0.119 & 0.224 \\ 0.434 & 0.217 & 0.126 & 0.223 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} T^5 = \begin{bmatrix} 0.422 & 0.212 & 0.148 & 0.218 \\ 0.419 & 0.209 & 0.156 & 0.216 \\ 0.419 & 0.219 & 0.143 & 0.219 \\ 0.422 & 0.214 & 0.145 & 0.219 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} T^{15} = \begin{bmatrix} 0.421 & 0.213 & 0.149 & 0.218 \\ 0.421 & 0.213 & 0.149 & 0.218 \\ 0.421 & 0.213 & 0.149 & 0.218 \\ 0.421 & 0.213 & 0.149 & 0.218 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} T^{30} = \begin{bmatrix} 0.421 & 0.213 & 0.149 & 0.218 \\ 0.421 & 0.213 & 0.149 & 0.218 \\ 0.421 & 0.213 & 0.149 & 0.218 \\ 0.421 & 0.213 & 0.149 & 0.218 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} T^{100} = \begin{bmatrix} 0.421 & 0.213 & 0.149 & 0.218 \\ 0.421 & 0.213 & 0.149 & 0.218 \\ 0.421 & 0.213 & 0.149 & 0.218 \\ 0.421 & 0.213 & 0.149 & 0.218 \end{bmatrix}$$

- calculated using calculator (with rounding)
- notice how each row eventually becomes similar
- notice how the values doesn’t really change after a large number of transformations

Steady State

As seen in the previous slide, the probabilities doesn’t really change after a large number of transformations. Eventually the system reaches a state where further transformations wouldn’t change the probabilities. This is called a **steady state,** and it gives the long-term probability of a system.

The steady state can also be calculated with precise math.

Steady State continued

Let S be the steady state and T be the transition matrix.

ST = S □ further transformations does not change the steady state

$$\begin{array}{cccc} \text{Sunny} & \text{Cloudy} & \text{Rainy} & \text{Windy} \\ \end{array}$$
$$\begin{bmatrix} a & b & c & (1-a-b-c) \end{bmatrix} \begin{bmatrix} 0.5 & 0.15 & 0.1 & 0.25 \\ 0.2 & 0.15 & 0.5 & 0.15 \\ 0.25 & 0.5 & 0 & 0.25 \\ 0.6 & 0.2 & 0 & 0.2 \end{bmatrix} = \begin{bmatrix} a & b & c & (1-a-b-c) \end{bmatrix}$$

$$\downarrow$$

$$\left\{ \begin{array}{c} 0.5a + 0.2b + 0.25c + 0.6(1-a-b-c) = a \\ 0.15a + 0.15b + 0.5c + 0.2(1-a-b-c) = b \\ 0.1a + 0.5b + 0c + 0(1-a-b-c) = c \\ 0.25a + 0.15b + 0.25c + 0.2(1-a-b-c) = 1-a-b-c \end{array} \right\}$$

System of equations

$$\downarrow$$

$$\left\{ \begin{array}{c} 1.1a + 0.4b + 0.35c = 0.6 \\ 0.05a + 1.05b - 0.3c = 0.2 \\ 0.1a + 0.5b - c = 0 \\ 1.05a + 0.95b + 1.05c = 0.8 \end{array} \right\}$$

Steady State continued

Put into matrix form:

$$\begin{bmatrix} 1.1 & 0.4 & 0.35 & 0.6 \\ 0.05 & 1.05 & -0.3 & 0.2 \\ 0.1 & 0.5 & -1 & 0 \\ 1.05 & 0.95 & 1.05 & 0.8 \end{bmatrix}$$

Using RREF, a method used for matrix calculations:

$$\begin{bmatrix} 1.1 & 0.4 & 0.35 & 0.6 \\ 0.05 & 1.05 & -0.3 & 0.2 \\ 0.1 & 0.5 & -1 & 0 \\ 1.05 & 0.95 & 1.05 & 0.8 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0.4208 \\ 0 & 1 & 0 & 0.2129 \\ 0 & 0 & 1 & 0.1485 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$S = \begin{bmatrix} 0.4208 & 0.2129 & 0.1485 & 0.2178 \end{bmatrix}$$

On any given day, Lexington has 42.08% chance of being sunny, 21.29% of being cloudy, 14.85% of being rainy, and 21.78% of being windy.

Thank you!