Lecture 04-05-06

PROBABILITY RELATIONSHIPS AND MARKOV CHAINS

Probability & Statistics

Francisco Richter, Martina Boschi and Ernst Wit

Introduction

This chapter covers fundamental concepts in probability theory: independence, covariance, and characteristic functions. These are crucial for understanding the relationships between random variables, measuring how they vary together, and characterizing their distributions. We will also see how these concepts apply to practical situations and why they are important tools in statistics.

1 Independence

Independence is a foundational concept in probability, ensuring that the occurrence of one event or the realization of one random variable provides no information about another. This section introduces the mathematical definition of independence, provides examples, and shows how to determine if two random variables are independent.

Definition 1 (Independence)**.** *Two random variables X and Y are said to be independent if their joint probability mass function (or density function) can be expressed as the product of their marginal distributions for all values of x and y:*

$$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y) \tag{1}$$

This definition implies that, for independent random variables, the joint distribution does not provide any additional information beyond what is provided by the individual distributions.

Consider two independent coin tosses. Let X_1 represent the outcome of the first coin toss and X_2 represent the outcome of the second coin toss, where Heads is represented by 1 and Tails by 0. The joint probabilities are as follows:

X_1	X_2	Joint Probability $p_{X_1,X_2}(x_1,x_2)$	Product $p_{X_1}(x_1) \cdot p_{X_2}(x_2)$
1	1	0.25	0.25
1	0	0.25	0.25
0	1	0.25	0.25
0	0	0.25	0.25

Table 1: Joint probabilities of two independent coin tosses

The table confirms that the joint probabilities are equal to the product of the individual probabilities, which implies independence.
Independence between random variables can be visualized using scatter plots. When two random variables are independent, the scatter plot shows no discernible pattern, indicating that the realization of one variable does not influence the realization of the other.

Proposition 2. *If X and Y are independent random variables, then the expectation of their product is the product of their expectations:*

$$E[XY] = E[X] \cdot E[Y] \tag{2}$$

Proof. Let X and Y be independent random variables. By the definition of expectation

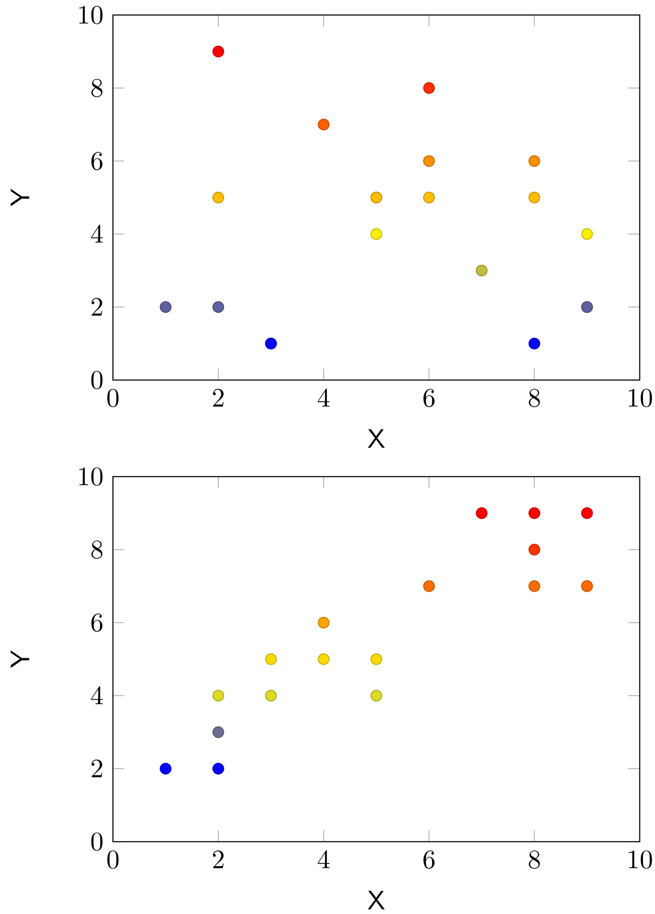

Figure 1: Scatter plots: The left plot shows independent random variables, while the right plot suggests a relationship between the variables.

and independence:

$$
\begin{aligned}
E[XY] &= \sum_x \sum_y xy \cdot p_{X,Y}(x,y) \\
&= \sum_x \sum_y xy \cdot p_X(x) p_Y(y) \quad \text{(since } X \text{ and } Y \text{ are independent)} \\
&= \left(\sum_x x \cdot p_X(x) \right) \cdot \left(\sum_y y \cdot p_Y(y) \right) \\
&= E[X] \cdot E[Y]
\end{aligned}
$$

This proves that the expectation of the product of independent random variables is the product of their expectations. □

Covariance is a measure of the joint variability of two random variables. It indicates whether the two variables tend to increase or decrease together (positive covariance) or if one increases when the other decreases (negative covariance).

Definition 3 (Covariance)**.** *The covariance between two random variables X and Y is defined as follows:*

$$\mathit{Cov}(X,Y) = E\left[(X - E[X])(Y - E[Y])\right] \tag{3}$$

If X and Y are independent, their covariance is zero:

$$\text{Cov}(X,Y) = 0 \tag{4}$$

However, a covariance of zero does not necessarily imply independence. It simply means that there is no linear relationship between the two variables.

Example. Consider the daily returns of two stocks, X and Y. Assume that the expected returns are $E[X] = 0.02$ and $E[Y] = 0.03$, and the expected product of the returns is $E[XY] = 0.0008$. The covariance is calculated as follows:

$$\text{Cov}(X,Y) = E[XY] - E[X]E[Y] = 0.0008 - (0.02 \times 0.03) = 0.0002 \tag{5}$$

A positive covariance indicates that the two stocks tend to move in the same direction. □

Characteristic functions provide an alternative way of describing the distribution of a random variable. They have useful properties, such as uniquely determining the distribution and simplifying the analysis of sums of independent random variables.

Definition 4 (Characteristic Function)**.** *The characteristic function of a random variable X is defined as:*

$$\phi_X(t) = E\left[e^{itX}\right] \tag{6}$$

where i is the imaginary unit.

Characteristic functions have several important properties:

- They uniquely determine the probability distribution of a random variable.
- For independent random variables X and Y, the characteristic function of their sum is the product of their individual characteristic functions:

$$\phi_{X+Y}(t) = \phi_X(t) \cdot \phi_Y(t) \tag{7}$$

Example (Characteristic Function of Independent Sum)**.** Let X and Y be independent random variables. The characteristic function of their sum can be computed as follows:

$$\begin{aligned}
\phi_{X+Y}(t) &= E\left[e^{it(X+Y)}\right] && (8)\\
&= E\left[e^{itX}e^{itY}\right] && (9)\\
&= E\left[e^{itX}\right] \cdot E\left[e^{itY}\right] && (10)\\
&= \phi_X(t) \cdot \phi_Y(t) && (11)
\end{aligned}$$

This property is particularly useful when analyzing the distribution of sums of independent random variables. □

2 Conditional Probabilities

In the previous section, we discussed how to determine the independence of random variables. But what if they are dependent? Often, information about one event can help us understand another. For example, will I wear a jacket if it rains tomorrow? Conditional probability allows us to update our knowledge based on new information, helping us make better predictions about related events.

Definition 5 (Conditional Probability)**.** *Given two random variables X and Y with $p_Y(y) > 0$, the **conditional probability** of $X = x$ given $Y = y$ is defined as:*

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

Example (Weather and Clothing)**.** Suppose we want to determine the probability that a person will wear a jacket (X) depending on whether it is raining (Y). Let $Y = 1$ if it is raining and $Y = 0$ if it is not. Similarly, let $X = 1$ if the person wears a jacket and $X = 0$ if they do not.
From historical data, we have the following information:

- Probability of rain: $P(Y = 1) = 0.3$
- Probability of wearing a jacket given rain: $P(X = 1|Y = 1) = 0.9$
- Probability of wearing a jacket given no rain: $P(X = 1|Y = 0) = 0.2$

Using the definition of conditional probability, we can find the probability that a person wears a jacket when it is raining or not. □

Conditional probability provides a mathematical framework to assess the probability of a random variable X given that another random variable Y has occurred. Understanding this is particularly essential when X and Y are not independent, as the occurrence of Y can significantly alter the probability landscape for X. In other words, when two random variables are dependent, the probability of one variable taking a certain value changes after we know that the other variable has taken a specific value.

A direct consequence of the definition of conditional probability is the Multiplication Rule. It provides a foundational link between joint and conditional probabilities, allowing for systematic computation of joint probabilities.

Theorem 6 (Multiplication Rule)**.** *Let X and Y be two random variables. The joint probability $p_{X,Y}(x,y)$ can be expressed in terms of conditional probabilities as:*

$$p_{X,Y}(x,y) = p_{X|Y}(x|y) \cdot p_Y(y) = p_{Y|X}(y|x) \cdot p_X(x)$$

Example (Medical Test)**.** Consider a medical test used to detect a certain disease. Let X represent whether a person tests positive ($X = 1$) or negative ($X = 0$), and let Y represent whether the person actually has the disease ($Y = 1$) or not ($Y = 0$).
Suppose we know the following:

- Probability that a person has the disease: $P(Y = 1) = 0.01$
- Probability of testing positive given the person has the disease: $P(X = 1|Y = 1) = 0.95$
- Probability of testing positive given the person does not have the disease (false positive rate): $P(X = 1|Y = 0) = 0.05$

Using the multiplication rule, we can find the joint probability of a person testing positive and having the disease. □

For many random variables $X_1, X_2, \dots, X_n$, the joint probability can be expressed as:

$$p_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} p_{X_i|X_1,\dots,X_{i-1}}(x_i|x_1, \dots, x_{i-1})$$

This represents the product of the conditional probabilities of each random variable occurring given the occurrence of all previous random variables.

Definition 7 (Marginal Probability)**.** *Given a joint probability distribution $p_{X,Y}(x,y)$, the marginal probability $p_X(x)$ of any outcome x for the random variable X is obtained by summing the joint probabilities over all possible outcomes y for Y. Mathematically, the marginal probability $p_X(x)$ is given by:*

$$p_X(x) = \sum_{y \in R(Y)} p_{X,Y}(x,y)$$

where the sum is over all possible outcomes of the random variable Y.

The connection between marginal and conditional probabilities can be understood through the **law of total probability**. The marginal probability $p_X(x)$ can be expressed in terms of conditional probabilities as follows:

$$p_X(x) = \sum_{y \in R(Y)} p_{X|Y}(x|y) \cdot p_Y(y)$$

This relationship demonstrates that the marginal probability of an outcome for a random variable can be obtained by considering all the ways that outcome can occur, weighted by the probability of each of those ways.

Example (Drawing Cards)**.** Consider drawing two cards from a deck without replacement. Let X represent the suit of the first card, and Y represent the suit of the second card.
Let us codify the 4 suits of cards in real numbers, i.e.

$$\{\text{Heart} \mapsto 1, \text{Spades} \mapsto 2, \text{Clubs} \mapsto 2, \text{Diamonds} \mapsto 2, \}$$

Suppose we know that the first card drawn is a heart ($X = 1$). We are interested in the probability that the second card is also a heart ($Y = 1$).
The probability of drawing a heart on the first draw is $P(X = 1) = \frac{13}{52} = \frac{1}{4}$. After drawing a heart, there are now 12 hearts left in a deck of 51 cards. Therefore, the conditional probability of drawing a heart on the second draw, given that the first card was a heart, is:

$$P(Y = 1|X = 1) = \frac{12}{51}$$

Using the Multiplication Rule, the joint probability of drawing two hearts in a row is:

$$P(X = 1, Y = 1) = P(X = 1) \cdot P(Y = 1|X = 1) = \frac{1}{4} \cdot \frac{12}{51} = \frac{12}{204} = \frac{1}{17}$$

Therefore, the probability of drawing two hearts in a row is approximately 5.88%. □

3 Bayes' Theorem

Suppose you are trying to determine if a piece of fruit picked from a bag is an apple. Your initial belief (prior probability) might be based on the overall percentage of apples in the bag. However, when you touch the fruit and feel it's round and smooth, you can update your belief based on this new evidence. Bayes' theorem provides a way to combine these sources of information.

The interplay of events in a probabilistic framework is not always straightforward. Often, we have evidence or observations and seek to update our understanding of a particular event's probability based on this new information. Bayes' theorem offers a mathematical means to achieve this. It allows us to reverse conditional probabilities, turning our perspective from the probability of observing evidence given an event to the probability of the event given the observed evidence.

Theorem 8 (Bayes' Theorem)**.** *Given two random variables X and Y with $p_Y(y) \neq 0$, the conditional probability $p_{X|Y}(x|y)$ is:*

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x) \cdot p_X(x)}{p_Y(y)}$$

Where: $p_{X|Y}(x|y)$ is the posterior probability, $p_{Y|X}(y|x)$ is the likelihood, $p_X(x)$ is the prior probability, and $p_Y(y)$ is the evidence.

Bayes' theorem provides a way to compute a posterior probability. It relates the likelihood of observing Y given X, the prior probability of X, and the total probability of observing Y.
Starting from the Multiplication Theorem:

$$p_{X,Y}(x,y) = p_{Y|X}(y|x) \cdot p_X(x)$$

Since $p_{X,Y}(x,y) = p_{Y,X}(y,x)$, we also have:

$$p_{Y,X}(y,x) = p_{X|Y}(x|y) \cdot p_Y(y)$$

Equating the two gives:

$$p_{X|Y}(x|y) \cdot p_Y(y) = p_{Y|X}(y|x) \cdot p_X(x)$$

Rearranging, we arrive at Bayes' theorem:

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x) \cdot p_X(x)}{p_Y(y)}$$

Example. Imagine there's a rare disease, and there's a test for it. The disease affects 1% of the population, and the test is 99% accurate. If you test positive, what's the chance you actually have the disease?

Using Bayes' theorem:

Let X be the random variable representing the presence $(X = 1)$ of the disease. In case the disease is not present it assumes a value 0. Let Y be the random variable representing the test result. $Y = 1$ for a positive test, while if the test is negative it is equal to 0.

We want to find: $p_{X|Y}(1|Y = 1)$, namely we want to understand what it is the probability of the presence of the disease given a positive test.
Given:

- $p_X(1) = 0.01$ (1% of the population has the disease)
- $p_{Y|X}(Y = 1|X = 1) = 0.99$ (The test is 99% accurate)
- $p_Y(1)$ is the total positive testing probability.

$$p_{X|Y}(1|Y = 1) = \frac{p_{Y|X}(Y = 1|X = 1) \cdot p_X(1)}{p_Y(1)}$$

To find $p_Y(1)$, namely the probability of a positive test, we consider the law of total probability:

$$p_Y(1) = p_{Y|X}(1|X = 1) \cdot p_X(1) + p_{Y|X}(1|X = 0) \cdot p_X(0)$$

$$p_Y(1) = (0.99)(0.01) + (0.01)(0.99) = 0.0198$$

Plugging in the numbers:

$$p_{X|Y}(1|Y = 1) = \frac{(0.99)(0.01)}{0.0198} \approx 0.5$$

So, even with a 99% accurate test, if you test positive, there's only a 50% chance you actually have the disease! □

Bayes' theorem is foundational for the fields of Bayesian statistics and machine learning. It provides a mechanism to update our beliefs in light of new evidence, making it central to numerous applications, from medical diagnostics to recommendation systems. The theorem reminds us of the importance of prior knowledge and illustrates how, in a world filled with data, we can use this data to make more informed decisions and predictions.

Bayes' theorem forms the foundation of many NLP applications:

- **Spam Filters**: Bayesian classifiers can determine if an email is spam or not based on the frequency of certain words.
- **Sentiment Analysis**: Using Bayes' theorem, algorithms can determine the sentiment of a given text (positive, negative, neutral) by analyzing the words used.

Example (Spam Email Classification)**.** Consider an email filter that classifies emails as spam or not spam based on certain words. Let S be the random variable representing whether an email is spam ($S = 1$) or not ($S = 0$). Let W represent the presence of certain words in the email.
Suppose we have:

- $p_S(1) = 0.4$ (40% of emails are spam)
- $p_{W|S}(1|1) = 0.8$ (Probability of certain words given the email is spam)
- $p_{W|S}(1|0) = 0.1$ (Probability of certain words given the email is not spam)

We want to find the probability that an email is spam given that it contains these words, i.e., $p_{S|W}(1|1)$.
Using Bayes' theorem:

$$p_{S|W}(1|1) = \frac{p_{W|S}(1|1) \cdot p_S(1)}{p_W(1)}$$

To find $p_W(1)$:

$$p_W(1) = p_{W|S}(1|1) \cdot p_S(1) + p_{W|S}(1|0) \cdot p_S(0)$$

$$p_W(1) = (0.8)(0.4) + (0.1)(0.6) = 0.38$$

Therefore:

$$p_{S|W}(1|1) = \frac{(0.8)(0.4)}{0.38} \approx 0.842$$

So, given that the email contains certain words, there is an 84.2% chance that it is spam. □

Bayes' theorem finds applications in finance, especially in the area of risk management and investment strategies. Bayesian models can help in predicting the likelihood of certain economic conditions based on current and past data.

- **Portfolio Management**: Investors can update their beliefs about the expected returns of assets based on new market information.
- **Risk Assessment**: Bayesian models can evaluate the risk of investments or loans by considering both historical data and expert judgment.

Example (Market Sentiment)**.** Suppose an investor wants to determine whether the market is bullish ($M = 1$) or bearish ($M = 0$) based on recent news sentiment (N). Let:

- $p_M(1) = 0.5$ (Initial belief that the market is equally likely to be bullish or bearish)
- $p_{N|M}(1|1) = 0.7$ (Probability of positive news given a bullish market)

- $p_{N|M}(1|0) = 0.3$ (Probability of positive news given a bearish market)

If the news is positive, we want to find the updated probability that the market is bullish, i.e., $p_{M|N}(1|1)$.
Using Bayes' theorem:

$$p_{M|N}(1|1) = \frac{p_{N|M}(1|1) \cdot p_M(1)}{p_N(1)}$$

To find $p_N(1)$:

$$p_N(1) = p_{N|M}(1|1) \cdot p_M(1) + p_{N|M}(1|0) \cdot p_M(0)$$

$$p_N(1) = (0.7)(0.5) + (0.3)(0.5) = 0.5$$

Therefore:

$$p_{M|N}(1|1) = \frac{(0.7)(0.5)}{0.5} = 0.7$$

So, given positive news, there is a 70% chance that the market is bullish. □

4 A brief introduction to Markov Chains

Markov Chains offer a mathematical framework to model systems that evolve from one state to another over time. A typical representation for Markov Chains is a state diagram, as shown in the Figure below.

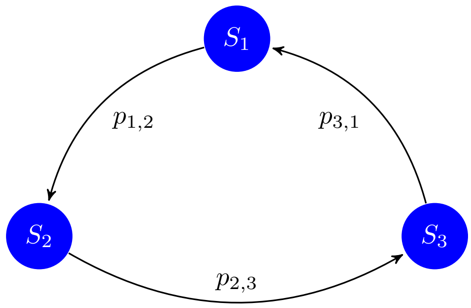

Figure 2: Example of a Markov Chain with three states and transition probabilities.

In the figure, the circles represent the states of the system, and the arrows symbolize the potential transitions between these states. The numbers adjacent to the arrows indicate the respective probabilities for these transitions. Given two states i and j, the probability of transition from state i to state j is equal to p_{ij}.
As we have traversed through various facets of probability, from the foundational concepts to the applied, it is evident that this journey has prepared us for understanding systems influenced by sequential randomness. In upcoming lectures, we will cap off our discussion by diving into simulations, offering two comprehensive examples: forest fire simulations and ant foraging simulations. These examples will serve as practical applications of the probabilistic and stochastic concepts we have learned so far.

5 Basics of Markov Chains

One of the most fundamental and widely studied stochastic processes is the Markov chain, named after the Russian mathematician Andrey Markov. This chain has a unique property that distinguishes it from other stochastic processes.

Definition 9 (Markov Chain)**.** *A sequence of random variables $\{X_0, X_1, X_2, \ldots\}$ is said to be a **Markov chain** if, for any $t \geq 0$ and any states $x_0, x_1, \ldots, x_t$, the probability of the next state x_{t+1} depends only on the current state x_t and not on the sequence of states preceding it. Mathematically, this property can be expressed as:*

$$P(X_{t+1} = x_{t+1} | X_0 = x_0, X_1 = x_1, \ldots, X_t = x_t) = P(X_{t+1} = x_{t+1} | X_t = x_t)$$

This property is often referred to as the **Markov property** or **memorylessness**. In simpler terms, the future state of a Markov chain depends only on its current state and not on its past states.

Example (Random Walk)**.** A **Random walk** is a classic example of a stochastic process and is often used to describe systems or sequences of events where the next state depends only on the current state and some random element. It can be visualized as a path taken by a particle that moves in random directions.
The simplest random walk is the one-dimensional walk. At each step, the walker takes a step either to the right $(+1)$ or to the left (-1) with equal probability.

1. Start at position 0.
2. At each time step, flip a coin:
   - If heads, move $+1$ step to the right.
   - If tails, move -1 step to the left.
3. Record the position after each step.
4. Repeat for a desired number of steps.

The particle starts at position 0. At each step, it moves either one step to the right $(+1)$ or one step to the left (-1) with equal probability. The path appears quite random, and the particle can drift far from the starting position but can also return close to its starting position at various times.
A 2D random walk can be visualized on a grid or plane. At each step, the walker takes a step either up, down, left, or right with equal probability.

1. Start at position $(0, 0)$.
2. At each time step, randomly choose one of the four directions:
   - Up: $(+0, +1)$
   - Down: $(+0, -1)$
   - Left: $(-1, +0)$
   - Right: $(+1, +0)$
3. Record the position after each step.
4. Repeat for a desired number of steps.

In the 2D random walk, the particle moves in a plane, starting from the origin. It takes steps in one of four possible directions: up, down, left, or right. The resulting path is a series of connected line segments in the plane, illustrating the random journey of the particle over time. The path is visualized using the 'Greys' color map, where the starting point is black and the ending point is white, representing the time progression. □

Definition 10. *For a Markov chain with N possible states, the $N \times N$ matrix $P = [p_{ij}]$ where $p_{ij} = P(X_{t+1} = j | X_t = i)$ is called the **Transition matrix** .*

Note that the rows of a transition matrix for a Markov chain must each sum to 1.

6 Characteristics of Markov Chains

1. **Transition Probabilities**: The probabilities of moving from one state to another are called transition probabilities. They are typically represented in a matrix called the transition matrix.
2. **State Space**: The set of all possible states that the chain can be in. This could be a finite or countably infinite set. In this context, we will focus on a total number of states equal to N.
3. **Time Structure**: Transitions occur at integer time steps. Specifically, you can visualize them as the subscripts of the random variables: the generic time point is the t-th.

Example (Infection Dynamics)**.** Consider a population that can be in one of three states: Susceptible (*S*), Infected (*I*), and Recovered (*R*). Individuals transition between these states according to the following Markov chain.
The transition probabilities are defined as:

- β: The probability of a Susceptible individual becoming Infected.

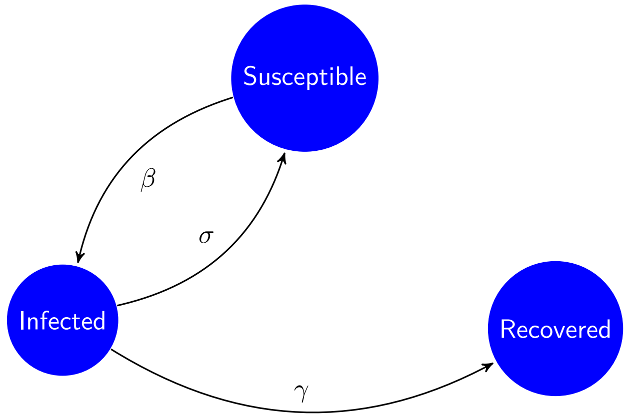

Figure 3: Example of a Markov Chain with three states and transition probabilities.

- γ: The probability of an Infected individual recovering and moving to the Recovered state.
- σ: The probability of an Infected individual transitioning back to the Susceptible state without recovering.

Given these probabilities, the transition matrix P for this Markov chain can be constructed as:

$$P = \begin{bmatrix} 1-\beta & \beta & 0 \\ \sigma & 1-\sigma-\gamma & \gamma \\ 0 & 0 & 1 \end{bmatrix}$$

Where the rows represent the current state and the columns represent the next state. This Markov chain captures the essential dynamics of many infectious diseases. The future state of each individual depends only on their current state, satisfying the Markov property. For example, if a person is currently Infected, the probability that they will be Recovered in the next time step is γ, irrespective of their past states. □

7 Chapman-Kolmogorov Theorem

Let $P_{ij}^{(n)}$ be the probability that the system transitions from state i to state j in n steps. Then:

$$P_{ij}^{(n)} = P(X_{t+n} = j | X_t = i) = (P^n)_{ij}$$

Notes:

- $P_{ij}^{(n)}$ can be found using the (i, j)th element of the matrix P^n.
- The potential paths from i to j in n steps are up to N^{n-1} (N being the number of states).
- Matrix multiplication of P by itself n times accumulates all transition probabilities.

Another representation of this concept, considering any times $s < t < u$, is given by:

$$P_{ij}(s, u) = \sum_{k=1}^{N} P_{ik}(s, t) \cdot P_{kj}(t, u) \tag{12}$$

Example (Financial Market Dynamics with Chapman-Kolmogorov Equations)**.** Consider the previously discussed financial market model.
The possible market states are $N = 3$ ($k = 1$ refers to a Bull market, $k = 2$ for a Bear market, and $k = 3$ for Stagnant market).
If we're interested in the probability of transitioning from a Bull market to a Bear market over five steps, P^5, we can employ the Chapman-Kolmogorov equation. Using our already computed matrices P^2 and P^3, the equation becomes:

$$(P^5)_{12} = \sum_{k=1}^{3} (P^2)_{1k} \cdot (P^3)_{k3} \tag{13}$$

Using the above equation, we find that:

$$(P^5)_{12} \approx 0.3526$$

This methodology not only simplifies computations but also offers insights into multi-step transitions in financial markets, enabling better predictive models. □

8 Limit Distribution: Irreducibility, Aperiodicity, and Ergodicity

Markov Chains, with their inherent ability to model complex stochastic systems, have significant applications across various domains, from finance and meteorology to social sciences. One of the most pivotal inquiries in the realm of Markov Chains pertains to their long-term behavior. Specifically, will the system stabilize into a steady state or equilibrium? This section delves into the notions that underpin this behavior, including the limit distribution and the properties of irreducibility, aperiodicity, and ergodicity.

Definition 11 (Limit Distribution)**.** *A distribution π is called a **limit distribution** for a Markov Chain if*

$$\lim_{n\to\infty} P_{ij}(0, n) = \pi_j$$

for every state i. The limit distribution provides insights into the enduring behavior of the system and is inherently linked to the properties discussed below.

The distribution π encapsulates the stable or steady-state probabilities associated with each state as the number of transitions grows indefinitely large. For a finite Markov Chain, the cumulative sum of all elements of π equals 1, emphasizing that π is a probability distribution.

Example (Convergence of P to π)**.** For the given transition matrix P, let's inspect its powers:
For P^5:

$$\begin{bmatrix} 0.34296 & 0.35264 & 0.3044 \\ 0.34664 & 0.33984 & 0.31352 \\ 0.33752 & 0.3648 & 0.29768 \end{bmatrix}$$

For P^{10}:

$$\begin{bmatrix} 0.34260 & 0.35183 & 0.30557 \\ 0.34251 & 0.35210 & 0.30539 \\ 0.34268 & 0.35159 & 0.30573 \end{bmatrix}$$

For P^{20}:

$$\begin{bmatrix} 0.34259 & 0.35185 & 0.30556 \\ 0.34259 & 0.35185 & 0.30556 \\ 0.34259 & 0.35185 & 0.30556 \end{bmatrix}$$

By the time we examine P^{50} and P^{100}, the matrix has stabilized to:

$$\begin{bmatrix} 0.34259 & 0.35185 & 0.30556 \\ 0.34259 & 0.35185 & 0.30556 \\ 0.34259 & 0.35185 & 0.30556 \end{bmatrix}$$

From the matrices above, we discern a clear trend: as we raise P to higher powers, the rows of the matrix are converging to the limit distribution π. This showcases the theoretical underpinning that, given certain conditions, the Markov Chain will stabilize to a unique long-term distribution. □

Definition 12. *A Markov Chain is **irreducible** if it is possible to traverse from any state to any other state within a finite number of steps. Formally, for any states $i, j \in \{1, ..., N\}$, there exists $n \geq 1$ such that $P_{ij}^{(n)} > 0$.*

Irreducibility plays a paramount role in systems like social networks, ensuring the flow of information across the entire network.

Definition 13. *A state of a Markov Chain exhibits **aperiodicity** if it doesn't revisit itself in a fixed pattern. Formally, a state i is aperiodic if the greatest common divisor of the set of steps n at which it returns to itself is one:* $\gcd\{n : P_{ii}^{(n)} > 0\} = 1$.

Aperiodicity is crucial in financial models to avoid deterministic cyclical behaviors, ensuring the model captures the nuances of real-world dynamics.

Definition 14. *A Markov Chain is termed* ***ergodic*** *if it embodies both irreducibility and aperiodicity. Ergodicity ensures the existence of a unique limit distribution π.*

Example (Board Game Dynamics)**.** Consider a simplified board game where players move across a linear track of 10 spaces based on dice rolls. The goal is to reach the 10th space, but there's a catch: the 4th and 7th spaces contain portals. Landing on the 4th space sends the player back to the 1st space, while the 7th space propels the player directly to the 10th space.
In this scenario, we can model the game as a Markov Chain, where each space on the board is a state. Transition probabilities depend on dice roll outcomes and the portal mechanics. For instance, if a player is on the 3rd space, there's a $\frac{1}{6}$ chance they'll land on the 4th space (and be sent back to the 1st space) in the next move.
The game's dynamics are both irreducible and aperiodic. It's irreducible because a player can move from any space to any other space (either directly or indirectly through dice rolls and portals). It's aperiodic because, while a player may revisit certain spaces multiple times due to the portals, there's no fixed pattern or cycle length for revisiting a specific space.
As players play this game repeatedly, we might be interested in the long-term probabilities or limit distribution of a player occupying each space. This distribution will reveal insights like the likelihood of players getting caught in the portal trap at the 4th space or the average number of turns to reach the 10th space. Ergodicity assures us that this distribution will stabilize over time, no matter the starting position. □

In the absence of ergodicity, a Markov Chain might produce vastly divergent outcomes over different simulation runs, complicating predictions and risk assessments. Thus, understanding these attributes is not merely academic; they underpin practical applications across various sectors, playing a crucial role in our understanding of complex systems.

9 Modeling Real-World Processes

9.1 Forest Fire Simulation

The dynamics of forest fires play a pivotal role in ecosystem conservation and the safety of communities adjacent to wooded regions. These fires, influenced by various environmental and situational factors, exhibit intricate patterns of spread, making predictions daunting. However, using mathematical models, especially Markov Chains, we can unravel the stochastic behavior underlying forest fires.
In the Forest Fire Simulation Model, the forest is depicted as a two-dimensional grid, where each cell represents a specific section of the forest. These cells can adopt one of four states: Grass ('G'), Tree ('T'), Burning ('B'), or Empty ('E') which indicates a previously burned area. As the grid evolves in discrete time intervals, the system's behavior aligns with a Markov Chain model, where the subsequent state of each cell hinges exclusively on its present state and the states of its immediate neighbors.
Mathematically, if G denotes the grid and S_C represents the state of a specific cell C, then:

$$S_C \in \{\text{Grass}, \text{Tree}, \text{Burning}, \text{Empty}\}$$

The transitions between these states adhere to probabilistic rules:

- A Grass cell has a probability p_{growth} to metamorphose into a Tree.
- A Tree cell ignites, becoming a Burning cell, if neighboring cells are ablaze, governed by a probability p_{ignite}.
- Occasionally, a Grass cell might spontaneously combust due to rare events like lightning, a phenomenon represented by $p_{\text{spontaneous}}$.
- Post-combustion, a Burning cell invariably transforms into an Empty cell, symbolizing the aftermath of the fire.

These rules can be encapsulated within a transition function T, which outlines the state of a cell at time $t+1$ contingent upon its and its neighbors' states at time t. This adherence to the Markov property underscores the essence of our model, wherein transitions are strictly influenced by the present state of affairs.

Experiment 1. *Our simulation experiment aims to emulate the progression and eventual aftermath of a wildfire in a forest. The grid is initialized with a certain distribution of Grass and Trees, with a few cells set to the Burning state to simulate the fire's* *ignition points. As the simulation progresses, the fire spreads according to the aforementioned probabilistic rules, affecting neighboring cells and altering the landscape. Over time, the fire exhausts available fuel (Grass and Trees) and transforms cells into the Empty state, representing burned regions. The simulation provides valuable insights into how fires spread, the efficacy of natural barriers, and potential strategies to contain or mitigate wildfires.*

A salient advantage of construing this simulation within the Markov Chain paradigm is the capability to prognosticate the forest's evolution over time. By discerning the steady-state probabilities, we can extrapolate vital data, such as the perennial likelihood of a particular cell being in any given state. Such probabilistic insights are instrumental for forest management strategies, from determining reforestation zones to instituting firebreaks and marshaling firefighting assets.
Additionally, the adaptability of the model shines through its parameters. Probabilities like p_{growth} and p_{ignite} can be calibrated to mirror real-world observations or tweaked to simulate myriad scenarios, enhancing the model's versatility.

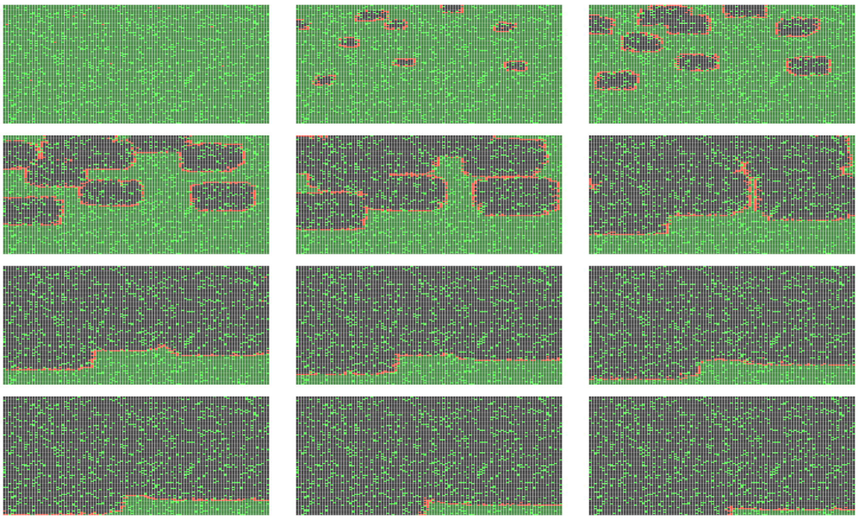

Figure 4: 12 snapshots in the forest fire simulation, showing the dynamic spread and aftermath of the fire.

In summation, Markov Chains furnish a potent methodology for simulating and deciphering the multifaceted dynamics of forest fires. Harnessing their mathematical prowess, we can derive insights into the trajectories of wildfires, optimize resource deployment, and conceive strategies to curtail the ravages of these natural calamities.

9.2 Agent-Based Models

Agent-Based Modeling (ABM) is a computational method used to model and analyze systems composed of individual agents interacting with each other and possibly with their environment. Each agent is typically defined by a set of characteristics and rules governing its behavior. Over time, these individual interactions can lead to the emergence of complex system-wide patterns and behaviors.
In an ABM, agents are often represented as entities with:

- **States**: Described by variables that can change over time, e.g., position, velocity, health status.
- **Behaviors**: Rules or strategies that determine how an agent will act based on its current state and the state of its surroundings.

The environment in which agents operate can be represented as a grid or a network, with agents moving between and interacting within these discrete spaces.
The dynamics of an ABM can often be described by a set of equations. Let $S_i(t)$ represent the state of agent i at time t. The behavior of agent i can then be represented as:

$$S_i(t+1) = f(S_i(t), N_i(t))$$

where:

- f is a function describing the agent's behavior.

- $N_i(t)$ represents the neighborhood or set of agents that agent i interacts with at time t.

The agents' behaviors and interactions lead to emergent properties of the system, which can be analyzed at the macro level.
The power of ABM lies in its ability to simulate individual behaviors to observe emergent macro-level outcomes. Traditional modeling techniques, such as differential equations, often describe systems at an aggregate level, making it difficult to capture localized interactions and heterogeneity. ABM, on the other hand, provides a bottom-up approach, making it particularly suited for systems where individual behaviors play a critical role in shaping collective dynamics.

Example. Ant Foraging Simulation
Consider a grid where ants search for food. The grid has cells representing either food, the nest, pheromones left by other ants, or empty cells. Each ant is an agent with the following behaviors:

1. If an ant finds food, it picks it up and returns to the nest.

2. If an ant carrying food encounters another ant, it deposits pheromones to signal the presence of food.

3. Ants follow pheromone trails to locate food sources more efficiently.

We provide a computational model to simulate the foraging behavior of ants. The ants' movements are influenced by their environment, primarily the presence of food and pheromones. This document formalizes the rules governing these behaviors based on the provided R code.
Let G be the grid where each cell (i, j) represents a position in the 2D space, and A be the set of ants. Each ant $a \in A$ has a state S_a defined as follows:

$$S_a = \{X, Y, F\}$$

Where:

- X, Y are the coordinates of the ant on the grid.

- F is a boolean flag indicating whether the ant is carrying food $(F = 1)$ or not $(F = 0)$.

The state of an ant at time $t + 1$ is given by:

$$S_{a,t+1} = P(S_{a,t})$$

Where P is the transition function defined as:

$$T(S_a) = \begin{cases} (X + \Delta x, Y + \Delta y, 1) & \text{if ant is at food source and } F = 0 \\ (X + \Delta x, Y + \Delta y, 0) & \text{if ant is at nest and } F = 1 \\ (X + \Delta x, Y + \Delta y, F) & \text{otherwise} \end{cases}$$

Δx and Δy are random variables that represent the ant's movement in each time step. Ants move based on their current state, previous state, and the state of the cells around them. The movement can be represented as:

$$M(a_t, a_{t-1}, N(a_t)) \rightarrow a_{t+1}$$

where a_t is the ant's current state at time t, a_{t-1} is the previous state, $N(a_t)$ denotes the neighboring cells' states, and a_{t+1} is the resulting state in the next time step.
The primary factor affecting the ants' movement is the presence of pheromones. Let's denote pheromones as Ph_1 and Ph_2. The transition probabilities associated with these pheromones are:

$$\begin{aligned} P(\text{grass} \rightarrow Ph_1) &= p_{\text{grasstopher1}} \\ P(Ph_1 \rightarrow Ph_2) &= p_{\text{pher1topher2}} \\ P(Ph_2 \rightarrow \text{grass}) &= p_{\text{pher2tograss}} \\ P(\text{grass} \rightarrow P_2) &= p_{\text{grasstopher2}} \end{aligned}$$

When an ant, not carrying food, moves to a cell with food, it picks up the food. This can be represented as:

$$F(a_t, \text{food}) \rightarrow (a_{t+1}, \text{carry})$$

where F is the function governing food pickup, a_t is the ant's current state, and 'carry' indicates the ant is now carrying food.

Ants carrying food, when they move and are not at the nest or on a food source, drop a pheromone. This can be represented as:

$$P(a_t, \text{carry}) \rightarrow P_1$$

Over time, pheromones on the grid undergo transitions:

$$\begin{aligned} Ph_1 &\rightarrow Ph_2 \quad \text{with probability } p_{\text{pher1topher2}} \\ Ph_2 &\rightarrow \text{grass} \quad \text{with probability } p_{\text{pher2tograss}} \end{aligned}$$

The randomness in ant movement is influenced by transition probabilities. Given a certain probability p, an ant will transition from state x to state y. This is represented by:

$$P(x, y, p) = \begin{cases} y & \text{with probability } p \\ x & \text{otherwise} \end{cases}$$

The simulation environment, including the ant positions, food sources, and nest, is initialized. The grid contains various entities: grass, ants, nest, and food. $\square$