MSBA7003 Quantitative Analysis Methods

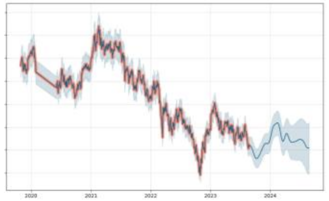
01 Probability & Bayesian Learning I

Agenda

- Probability Concepts
  - Events and Venn Diagram
  - Conditional Probability and Independence
- Bayes' Theorem
- Random Variables and Distributions
  - Joint, Marginal, and Conditional Distributions

Probability

- Probability is a numerical statement about the likelihood that an event will be seen.
    - 10% chance of rain tomorrow
    - 20% chance the Hang Seng Index will not go down next week
    - 30% chance there are aliens in the universe

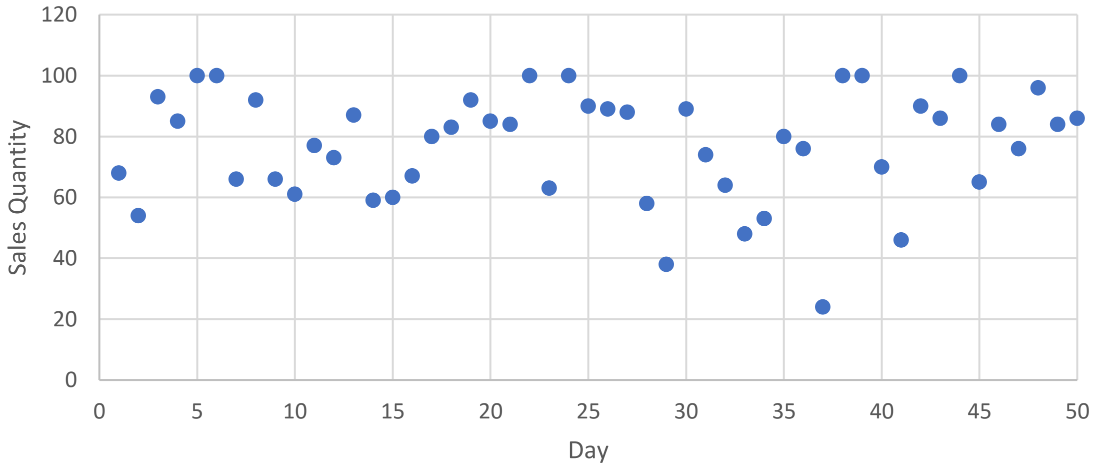
  

- Note: The event could have occurred, but we just do not know.
- Notation: $P(A) = Pr(A) =$ Probability of event A occurring.
- $0 \leq P(A) \leq 1$.

Determination of Probability

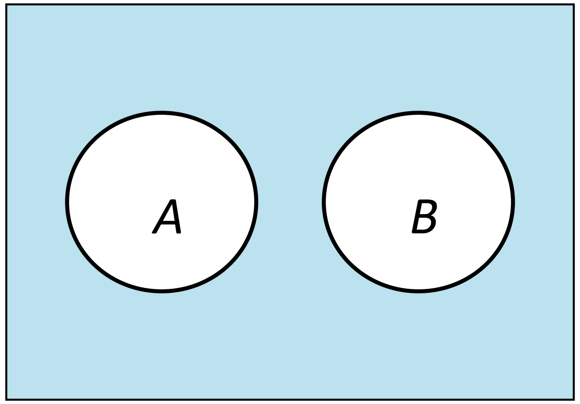
- Objective approach
  - Classical or logical method
    - $P(\text{head}) = 0.5$
    - $P(\text{spade}) = 0.25$
    - $P(\text{type AB blood given father type A \& mother type B}) = 9/16$
  - Relative frequency
    - Use data or experiments

Daily Sales Statistics of A Newspaper at a Newsstand

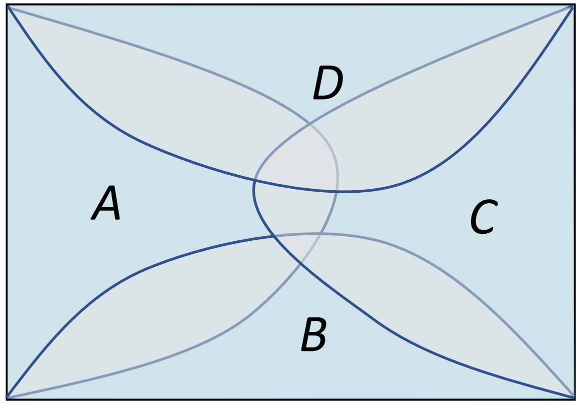


Events and Venn Diagram

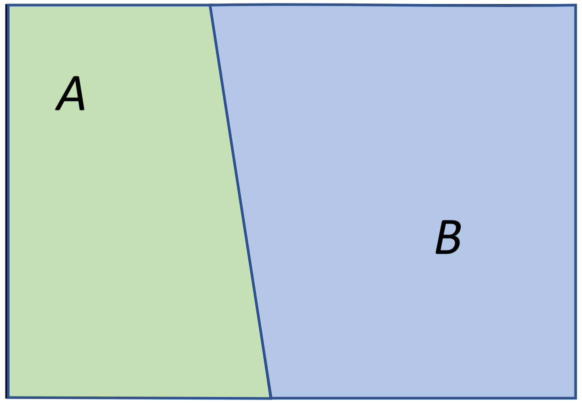
- **Mutually exclusive:** Events are *mutually exclusive* if only one of the events can occur in any one statistical trial or only one can occur at a time.
- **Collectively exhaustive:** Events are *collectively exhaustive* if they include every possible outcome in a statistical trial (i.e., they cover all the possibilities).



Events that are mutually exclusive



Events that are collectively exhaustive

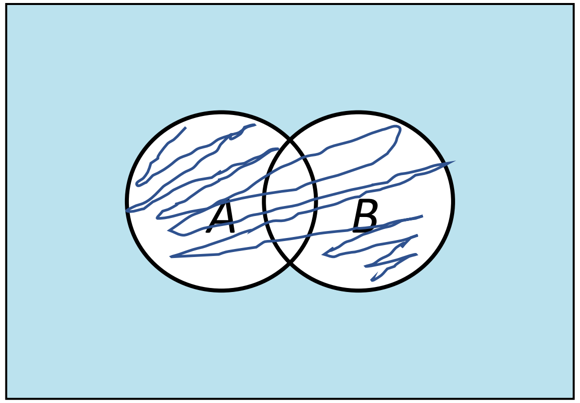


MECE events

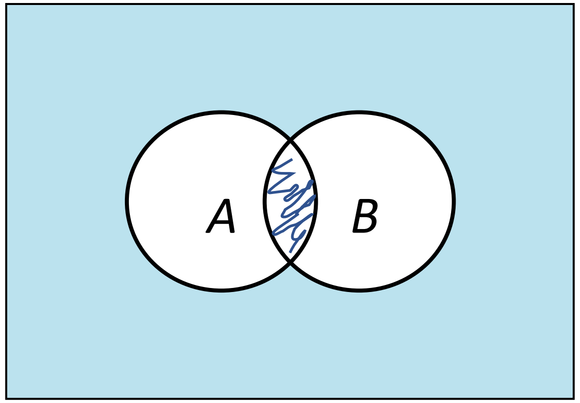
Note: The area of an event represents the probability.

Union and Intersection

- The **union** of two events is the set of all possible outcomes that are contained in either of the two events.
- $P(\text{Union of } A\&B) = P(A \text{ or } B) = P(A \cup B)$



- The **intersection** of two events is the set of all outcomes that are common to both events.
- $P(\text{Intersection of } A\&B) = P(A \text{ and } B) = P(A \cap B) = P(AB)$; it is called *joint probability*.



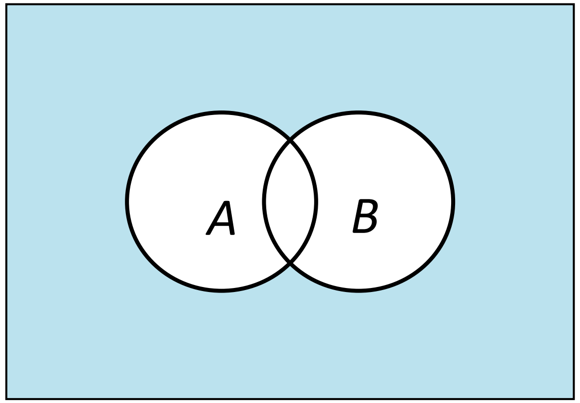
- $\boldsymbol{P(A \cup B) + P(A \cap B) = P(A) + P(B)}$

Conditional Probability

- A conditional probability is the probability of an event A occurring given that another event B has already happened.

- Notation: $P(A|B) = \frac{P(AB)}{P(B)}$. Why?

- $\boldsymbol{P(AB) = P(B) \cdot P(A|B) = P(A) \cdot P(B|A)}$.



- **Independent** events:
- If $A \perp B$, then $P(A|B) = P(A)$.
- If $A \perp B$, then $\boldsymbol{P(AB) = P(A) \cdot P(B)}$.

Basic Probability Rules

- $0 \leq P(A) \leq 1$ for any event A.
- $P(A \cap B) = 0$ if A and B are mutually exclusive.
- $P(A \cup B) = 1$ if A and B are collectively exhaustive.
- $P(A \cup B) + P(A \cap B) = P(A) + P(B)$.
- $P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$.
- $P(A|B) = P(A)$ if A and B are independent.
- $P(A \cap B) = P(A) \cdot P(B)$ if A and B are independent.

Probability Concepts: An Exercise

- Suppose A and B are mutually exclusive. In addition, $P(A) = 0.4$ and $P(B) = 0.6$. Suppose C and D are also mutually exclusive and collectively exhaustive. Further, $P(C|A) = 0.2$ and $P(D|B) = 0.4$. What are $P(C)$ and $P(D)$?

Bayes' Theorem

- How to revise your probability assessment when you have new information?

Diagnostic test for the Human Immuno-deficiency Virus (HIV)

	Infected	Not Infected
Test Positive	90% (conditional)	
Test Negative		95% (conditional)
HK Prevalence Rate	0.1% (marginal)	99.9% (marginal)

- Jack lives in Hong Kong, and he was randomly selected to take the test.
- $P(\text{Infected}|\text{Test Positive}) = ?$

Bayes' Theorem

- A = Infected; A' = Not Infected.
- B = Test Positive; B' = Test Negative.

- $P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B|A)P(A)}{P(BA)+P(BA')} = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A')P(A')}$
  - Note: $P(BA) + P(BA') = P\big((BA) \cup (BA')\big) + P\big((BA) \cap (BA')\big) = P(B) + 0$

- $P(\text{Infected}|\text{Test Positive}) = \frac{0.9 \times 0.001}{0.9 \times 0.001 + 0.05 \times 0.999}$

Bayes' Theorem: An Intuitive Way

- Draw a matrix for possible events related to two types (dimensions) of information.
- Calculate the joint probabilities
- Calculate conditional probabilities

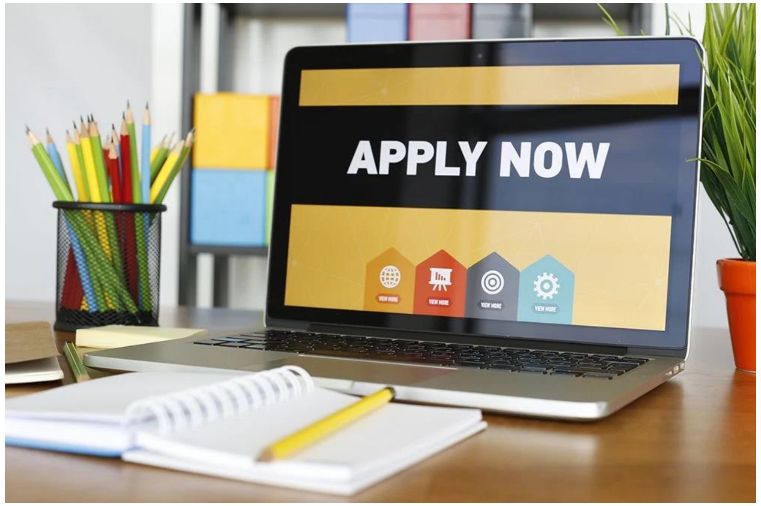
	Infected	Not Infected
Test Positive		
Test Negative		

Bayes' Theorem (Cont'd)

- If Jack took the test for the second time and got positive outcome again, what is the probability that Jack is infected?

Bayes' Theorem Exercise 1

- The Job Application Problem
  - In the past, the data analyst position at ByteWave.Com was publicly advertised, with an acceptance rate (number of admits/applicants) of 5%. According to historical data, 30% of the admitted candidates had a master's degree, while only 10% of the non-admitted candidates had a master's degree. Jack has a master's degree. If he applies for the data analyst position at ByteWave.Com, what is his probability of being admitted?



Bayes' Theorem Exercise 1

	Admitted	Rejected	Marginal
With Master Degree			
Without Master Degree			
Prior			
Posterior			

Bayes' Theorem Exercise 2

- The Presidential Election Problem
  - There are two experts on presidential election, A & B. According to historical data, A's predictions were correct in 90% cases, while B's predictions were correct only in 30% cases.
  - Pr(A predicts a candidate wins | the candidate wins) = 0.9
  - Pr(A predicts a candidate loses | the candidate loses) = 0.9
  - Pr(B predicts a candidate wins | the candidate wins) = 0.3
  - Pr(B predicts a candidate loses | the candidate loses) = 0.3
  - Now, without communicating with each other, both A & B predict that Donald Trump will be elected again. Without any information, your prior belief about Donald Trump being elected again is 0.5. Now knowing A & B's predictions, what should be your updated belief?

Bayes' Theorem Exercise 2

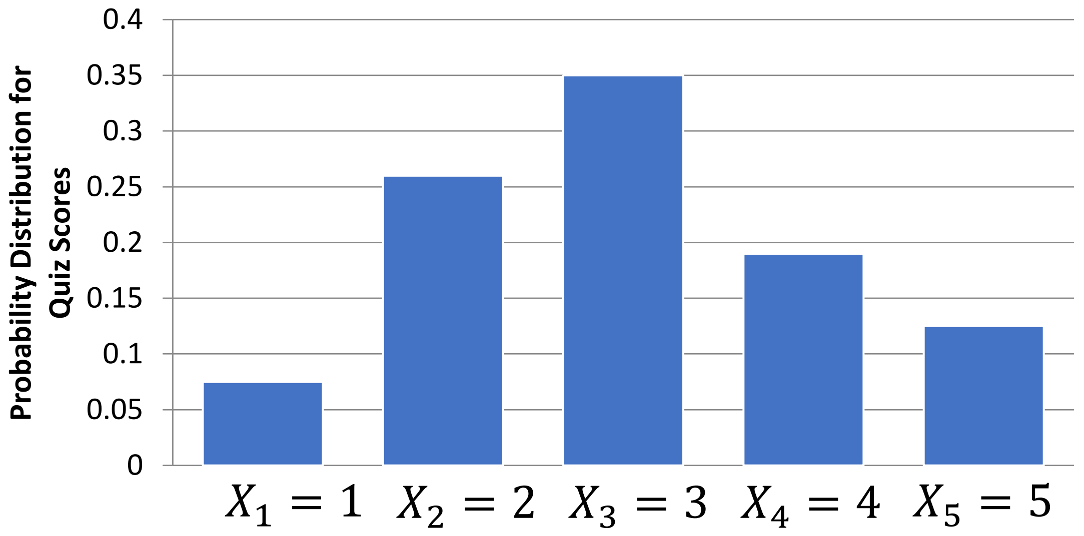
(A, B)	Trump Wins	Trump Loses	Marginal
Win, Win			
Win, Losing			
Losing, Win			
Losing, Losing			
Prior	0.5	0.5	
Posterior			

Random Variables

- For a set of events that are mutually exclusive and collectively exhaustive (MECE), if we assign a unique number/value to every possible event, then the number/value corresponding to the event occurring is a random variable (RV).
- A **discrete** RV can assume only a finite or countable set of values.
  - E.g., X = the number of newspapers sold during the day.
- A **continuous** RV has an uncountable set of possible values.
  - E.g., Y = the lifespan of a light bulb.
- When the outcome itself is not numerical or quantitative, it is necessary to define an RV that associates each outcome with a unique real number.
  - For tossing a coin, $X = 1$ if head and 0 if tail;
  - For consumers' response to how they like a product, $Y = 1$ if poor, 2 if average, and 3 if good;
  - For the brand of soda purchased by a consumer, $Z = 1$ if Pepsi, 2 if Coca-Cola, and 3 if Dr. Pepper.

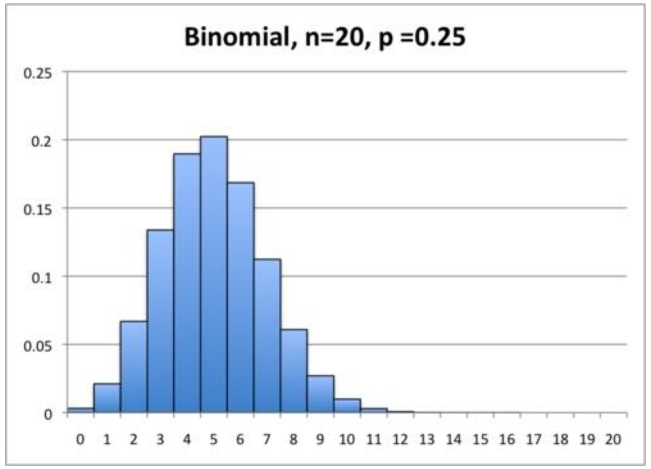
Discrete Distributions

- For each possible outcome X_i, there is a probability value $P(X_i)$.
- These values must be between 0 and 1: $0 \leq P(X_i) \leq 1$.
- They must sum up to 1: $\sum_{i=1}^{n} P(X_i) = 1$.

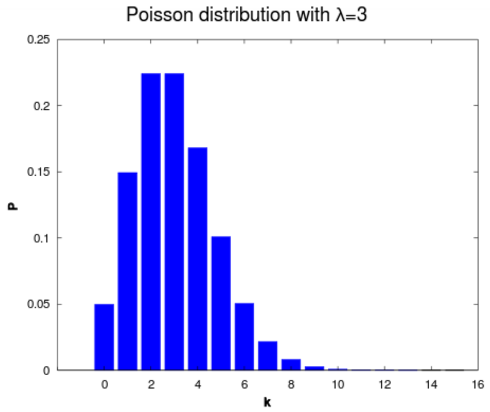


Discrete Distributions

- Binomial distribution
- Among N independent trials with the same success probability p, the number of successes follows Binomial distribution.



- Poisson distribution
- It is often used to describe the number of arrivals during a given period.
- If the average number of arrival during a unit time period is m, then the average is m×t during t units of time.



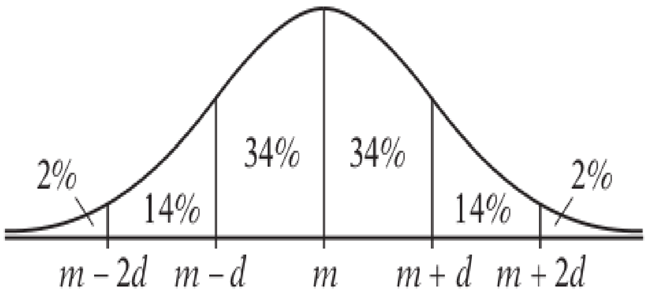
Continuous Distributions

- The sum of the probability values must equal 1.
- A continuous RV can take on an uncountable set of values such that the probability of each value must be 0.

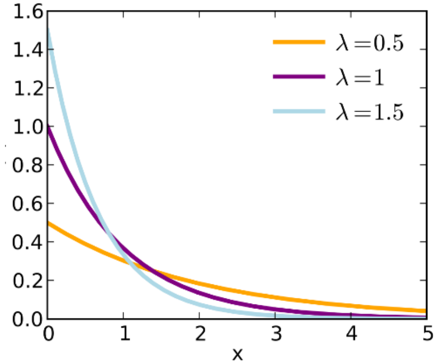
- The probability distribution is defined by continuous mathematical functions, the *cumulative distribution function* (CDF), and its derivative, the *probability density function* (PDF).
  - CDF is denoted by $F(\cdot)$ and $F(x) = P(X \leq x)$.
  - PDF is denoted by $f(\cdot) = F'(\cdot)$ and $f(x) \approx P(x < X \leq x + \Delta)/\Delta$.

Continuous Distributions

- Normal distribution
- If X follows Normal distribution with mean m and s.d. s, then the random variable $Z = (X - m)/s$ follows standard normal distribution.



- Exponential distribution
- It is often used to describe time intervals and durations.
- If the time intervals follows exponential, the number of arrivals during a given period follows Poisson distribution.
- It is memoryless.



Multiple Random Variables

- Each random variable represents one way to divide the state of the world into a set of MECE events.
- Different ways of division can be either independent or correlated.

- For two random variables (RVs) to be independent, all the events represented by one RV should be independent of all the events represented by the other RV.
- For two RVs to be correlated, at least one event represented by one RV should be correlated with at least one event represented by the other RV.

Multiple Random Variables

- The collection of joint probabilities (or densities) between the two sets of MECE events respectively represented by two random variables is called the joint distribution between the two random variables.
- The marginal distribution of a random variable is the collection of probabilities for the associated MECE events without knowing other information.
- The conditional distribution of a random variable is the collection of probabilities for the associated MECE events given the information of a related event.

Joint, Marginal, & Conditional Distribution

Joint Distribution

$f(x, y)$	$X = 1$	$X = 2$	$X = 3$	f_Y
$Y = 1$	0.3	0.2	0.1	0.6
$Y = 2$	0.1	0.2	0.1	0.4
f_X	0.4	0.4	0.2	
$Y = 1 \mid X$	0.75	0.5	0.5	
$Y = 2 \mid X$	0.25	0.5	0.5	
$E[Y \mid X]$	1.25	1.5	1.5	$E[Y] = 1.4$

Marginal Distributions

Conditional Distributions

Law of iterative expectations: $E\left[E[Y|X]\right] = E[Y]$.

In-Class Exercise: Playground

- Suppose only one class is on a high school playground. And we know the following conditional distributions of Y (the gender) and the marginal distribution of X (the class number):

$f(x,y)$	$X=1$	$X=2$	$X=3$	f_Y
$Y=1\|X$	0.75	0.5	0.4	?
$Y=2\|X$	0.25	0.5	0.6	?
f_X	1/3	1/3	1/3	

- For a random student, what is the probability of Y = 1 (marginal probability)?
- For a random student, what is the probability of X = 1 (i.e., from class 1) given Y = 1 (posterior probability)?

In-Class Exercise: Playground

- Suppose only one class is on a high school playground. And we know the following conditional distributions of Y (the gender) and the marginal distribution of X (the class number):

$f(x,y)$	$X = 1$	$X = 2$	$X = 3$	f_Y
$Y = 1 \vert X$	0.75	0.5	0.4	?
$Y = 2 \vert X$	0.25	0.5	0.6	?
f_X	1/3	1/3	1/3	

- If the first random student has Y = 1, what is the probability of the second random student having Y=1 again?

In-Class Exercise: Coin Tossing

- There is a fair or biased coin. What is the probability of getting a head?
- Suppose there are three possible hypotheses: 1/3, 1/2, and 2/3. They are equally likely.
- We can think of the index of the true hypothesis as a random variable, the distribution of which will be updated as we collect more information.

- Before we do anything, the probability of getting a head = $(1/3 + 1/2 + 2/3)/3 = 1/2$.

In-Class Exercise: Coin Tossing

- What if we tossed the coin only once and we got a head?
- Let p denote the probability of getting a head.
- We compute the posterior probabilities:

	p = 1/3	p = 1/2	p = 2/3	Marginal
Head	(1/3)*(1/3)	(1/2)*(1/3)	(2/3)*(1/3)	1/2
Tail	(2/3)*(1/3)	(1/2)*(1/3)	(1/3)*(1/3)	1/2
Prior Prob.	1/3	1/3	1/3	
Posterior	2/9	1/3	4/9	

- $E[p \mid Head] = (1/3)*(2/9) + (1/2)*(1/3) + (2/3)*(4/9) = 29/54 > 1/2$.

Takeaways

- Venn diagram can be used to describe the relationship between different random events.
- Events that do not overlap are not independent.

- A random variable represents a set of MECE events.

- Conditional probability is important for statistical learning.
- It quantitatively describes the relationship between different random events.
- It is defined as the area of the overlap divided by the area of the condition.

Quiz