MSBA7003 Quantitative Analysis Methods

01 Probability & Bayesian Learning

New Product Pricing with Limited Data

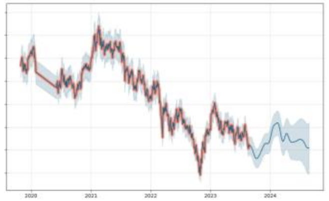
- Suppose Chow Tai Fook introduced a new gold ring. Historical sales data of similar rings suggest that customers' willingness to pay (WTP) follows a normal distribution with a standard deviation of $1,000. However, the mean WTP of this new ring is unknown. It can be anywhere between $2,000 to $5,000.
- The introductory price for this ring is $4,000. The cost for this ring is $2,000. There is sufficient inventory.
- If the first customer who showed interested in this ring did not buy it, how should the price be adjusted afterwards?

Agenda

- **Probability Concepts**
  - **Events and Venn Diagram**
  - **Conditional Probability and Independence**
- **Bayes' Theorem**
- **Random Variables and Distributions**
  - **Joint, Marginal, and Conditional Distributions**
- **Bayesian Inference & Application**
  - **The Authorship Problem**
  - **New Product Pricing**

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

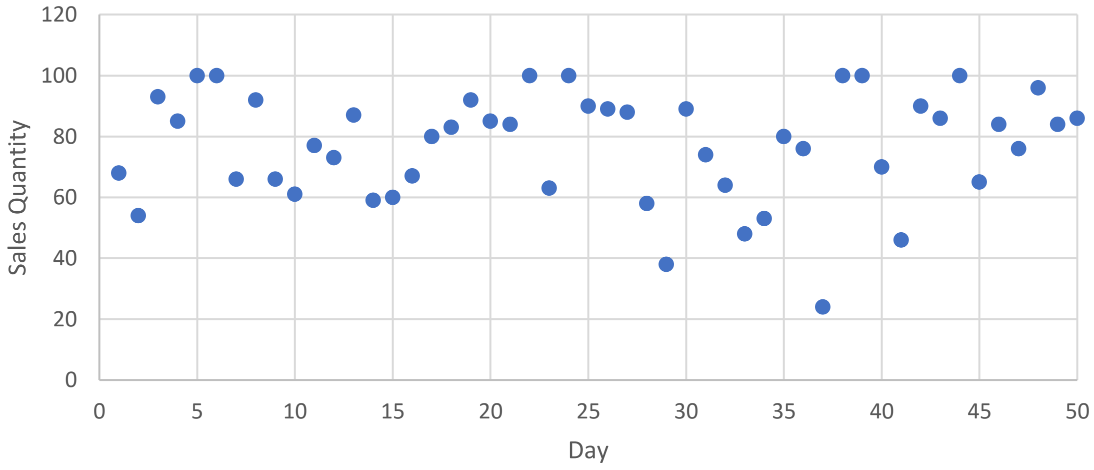


- Notation: $P(A) = Pr(A) =$ Probability of event A occurring.
- $0 \leq P(A) \leq 1$.

Determination of Probability

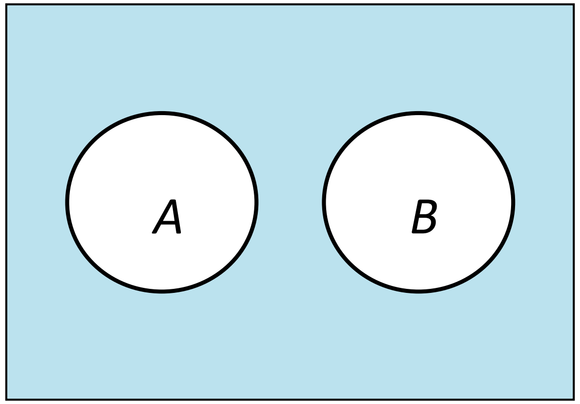
- Objective approach
  - Classical or logical method
    - $P(\text{head}) = 0.5$
    - $P(\text{spade}) = 0.25$
    - $P(\text{type AB blood given father type A \& mother type B}) = ?$
  - Relative frequency
    - Use data or experiments

Daily Sales Statistics of A Newspaper at a Newsstand

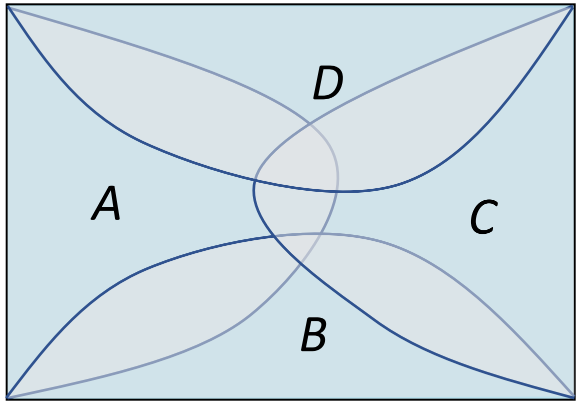


Events and Venn Diagram

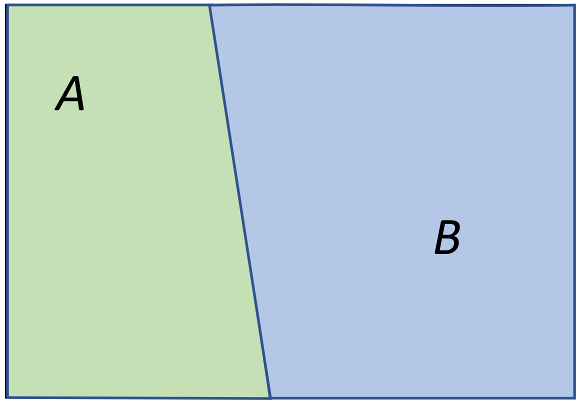
- **Mutually exclusive:** Events are *mutually exclusive* if only one of the events can occur in any one statistical trial or only one can occur at a time.
- **Collectively exhaustive:** Events are *collectively exhaustive* if they include every possible outcome in a statistical trial (i.e., they cover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



Events that are mutually exclusive



Events that are collectively exhaustive

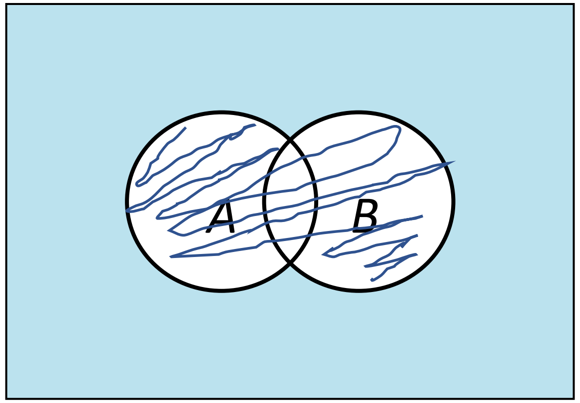


MECE events

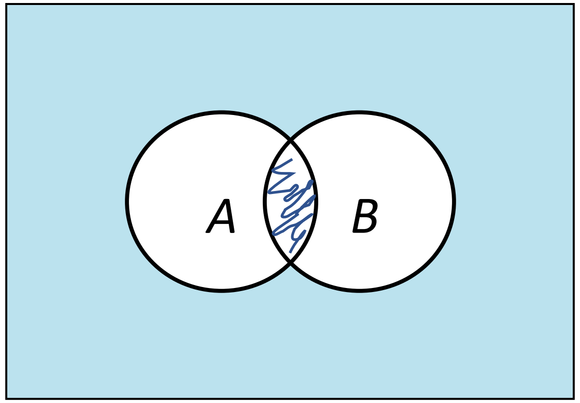
Note: The area of an event represents the probability.

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*.



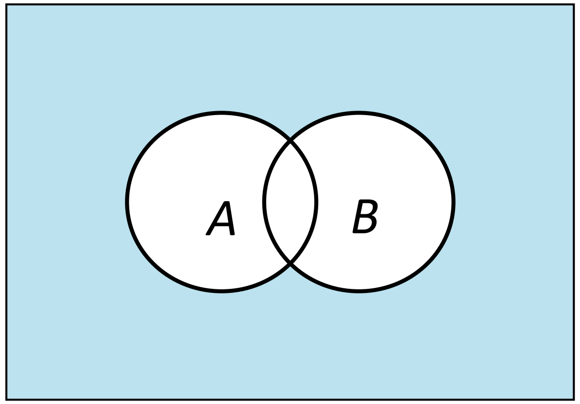
- $\boldsymbol{P(A \cup B) + P(A \cap B) = P(A) + P(B)}$

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?

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



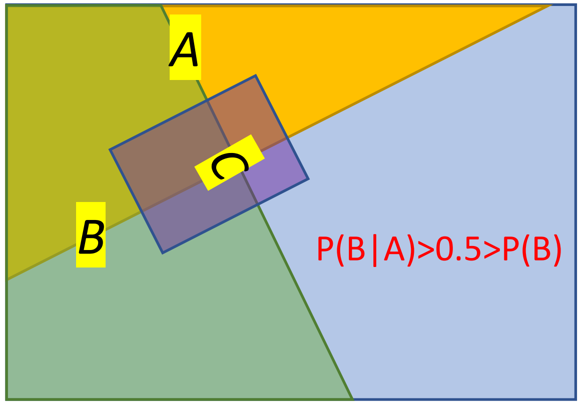
- **Independent** events:
- If $A \perp B$, then $P(A|B) = P(A)$.
- If $A \perp B$, then $\boldsymbol{P(AB) = P(A) \cdot P(B)}$.

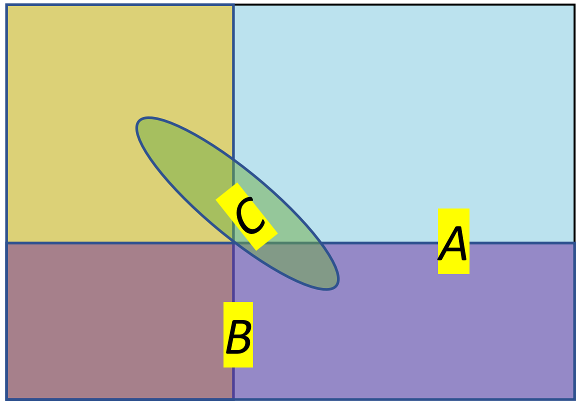
Conditional Independence

- For three different events, A, B, and C, if

$$P(AB|C) = P(A|C) \cdot P(B|C)$$

- Then A and B are ***conditionally independent*** given C.





- Example 1: A = lung cancer; B = yellow finger; C = smoking
- Example 2: A = Scotland; B = male; C = wearing a skirt

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.

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)

- $P(\text{Infected}|\text{Test Positive}) = ?$

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: $P(BA) + P(BA') = P\big((BA) \cup (BA')\big) + P\big((BA) \cap (BA')\big) = P(B) + 0$

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

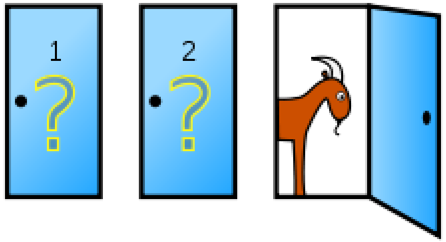
Bayes’ Theorem: An Intuitive Way

- 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: An Exercise

- The Monty Hall Problem
  - Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?



Bayes' Theorem: An Exercise

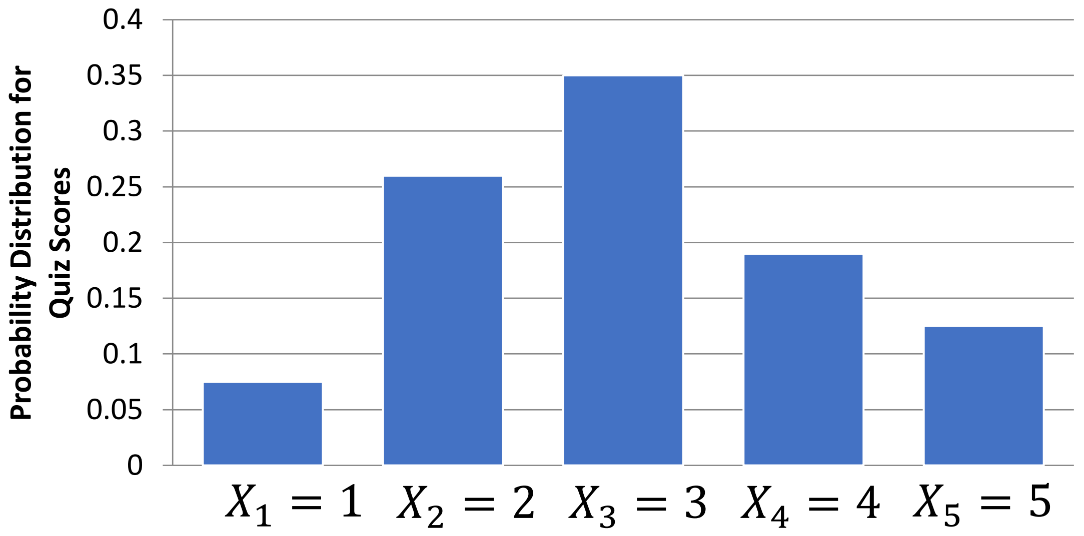
You picked door 1	(1)	(2)	(3)	
A				
B				
C				

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.

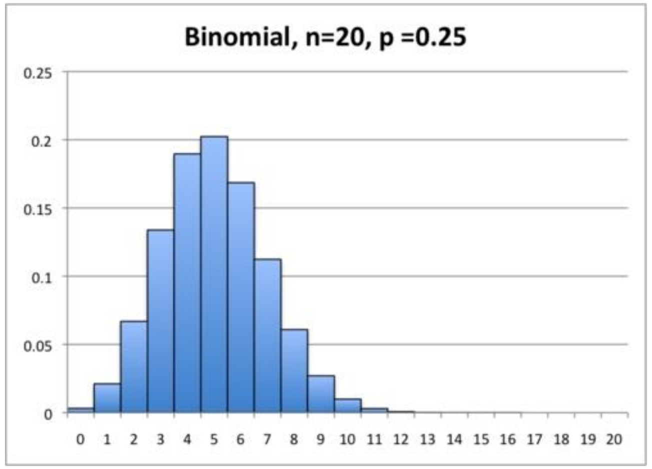
Discrete Distributions

- 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: $\sum_{i=1}^{n} P(X_i) = 1$.

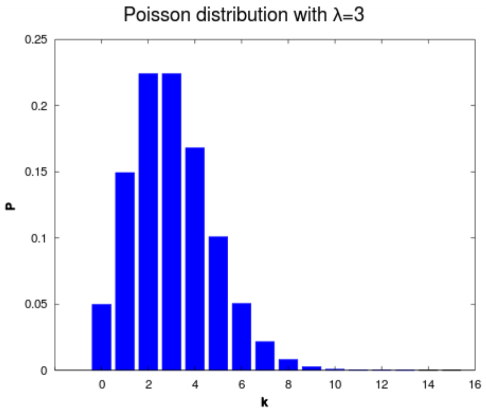


Discrete Distributions

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



- <u>Poisson distribution</u>
- 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 mt during t units of time.

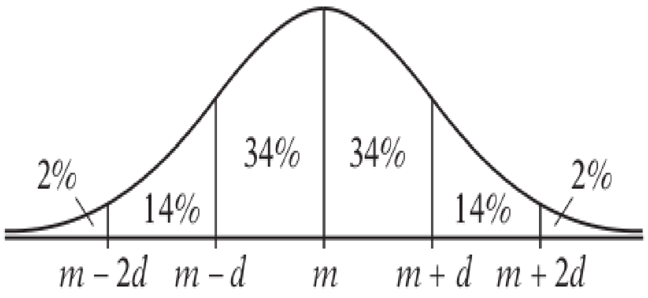


Continuous Distributions

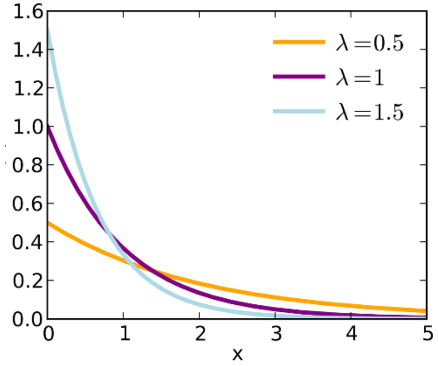
- The sum of the probability values must equal 1.
- A continuous RV can take on an uncountable set of values such that the probability of each value must be 0.
- The probability distribution is defined by continuous mathematical functions, the *cumulative distribution function* (CDF), and its derivative, 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

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



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

- Consider a high school playground. Suppose 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	

- 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?
- Suppose we <u>fix the value of X by a random draw</u>: only one class is on the playground. We do not know X. If the first student has Y = 1, then what is the probability of X = 1? What is the probability of the second student having Y=1 again?

Bayesian Inference

- We are interested in knowing the "state of the world X" (e.g., demand is high or low), and there are K possible states, which we call the ***Alternative Hypotheses***.
- The alternative hypotheses are mutually exclusive and collectively exhaustive. We have a prior subjective belief on each state (i.e., a marginal distribution of X).
- Under each hypothesis, a random variable Y will follow a known, distinct distribution.
- We wish to identify the state X by collecting samples of Y given the unknown state.
- After observing each value of Y, our subjective belief (marginal distribution) of X can be updated according to the Bayes' rule. *The posterior becomes the new prior*.
- With enough data points, we can use the posterior distribution of X to evaluate the probability of making the correct or wrong conclusion, and the posterior distribution of Y will converge to the conditional distribution given the "true" state X.

Bayesian Inference

- What is the probability of getting a head?
- Suppose there are three possible hypotheses: 1/3, 1/2, and 2/3.
- 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.

- What if we are allowed to toss the coin only once?

Bayesian Inference

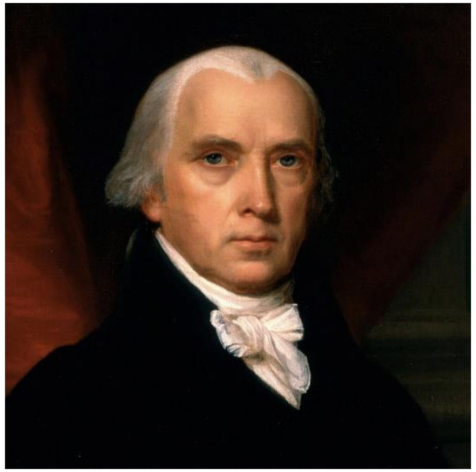
- Let p denote the probability of getting a head
- The three possible distributions are equally likely
- Suppose you toss the coin once and get a head.
- Which is the hypothesis with the largest posterior?

	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	
Conditional	2/9	1/3	4/9	

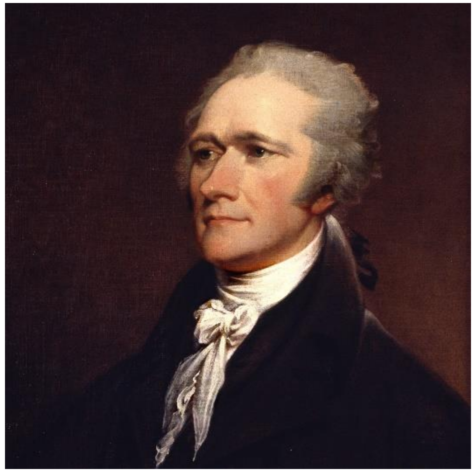
- $E[p|Head] = (1/3)*(2/9) + (1/2)*(1/3) + (2/3)*(4/9) = 29/54.$

Case: The Authorship Problem

- Federalist Papers (Published anonymously during 1787 - 1788)
  - Total: 77 papers
  - John Jay: 5
  - Alexander Hamilton: 43
  - James Madison: 14
  - Unknown: 12 + 3
- Bayesian Inference
  - Establish hypotheses: H_h vs. H_m
  - Determine the prior belief: 0.5 vs. 0.5 (or 0.75 vs. 0.25)
  - Collect data (the wording pattern in each paper)
  - Compute the probability of observing the data under each hypothesis
    - Using the papers with a known author
  - Compute the posterior of each hypothesis (for the papers in question)



James Madison (1751 - 1836)



Alexander Hamilton (1757 - 1840)

Case: The Authorship Problem

- Focus on non-contextual words
  - Rate of use is nearly invariant under change of topic.
  - Focus on the word [upon]
  - In paper 54, occurrence rate PTW = 0.996
- $\frac{\Pr(H\_h|data)}{\Pr(H\_m|data)} = \frac{\Pr(data|H\_h)\cdot\Pr(H\_h)}{\Pr(data|H\_m)\cdot\Pr(H\_m)}$

TABLE 2.3. FREQUENCY DISTRIBUTION FOR *upon*

Rate/1000	H	M
0 (exactly)	—	41
0+–1	1	7
1 –2	10	2
2 –3	11	
3 –4	11	
4 –5	10	
5 –6	3	
6 –7	1	
7 –8	1	
Totals	48	50

TABLE 2.5. FUNCTION WORDS AND THEIR CODE NUMBERS FOR THE FEDERALIST STUDY

1 a	8 as	15 do	22 has	29 is	36 no	43 or	50 than	57 this	64 when
2 all	9 at	16 down	23 have	30 it	37 not	44 our	51 that	58 to	65 which
3 also	10 be	17 even	24 her	31 its	38 now	45 shall	52 the	59 up	66 who
4 an	11 been	18 every	25 his	32 may	39 of	46 should	53 their	60 upon	67 will
5 and	12 but	19 for	26 if	33 more	40 on	47 so	54 then	61 was	68 with
6 any	13 by	20 from	27 in	34 must	41 one	48 some	55 there	62 were	69 would
7 are	14 can	21 had	28 into	35 my	42 only	49 such	56 thing	63 what	70 your

TABLE 2.6. ADDITIONAL WORDS AND CODE NUMBERS FOR THE FEDERALIST STUDY

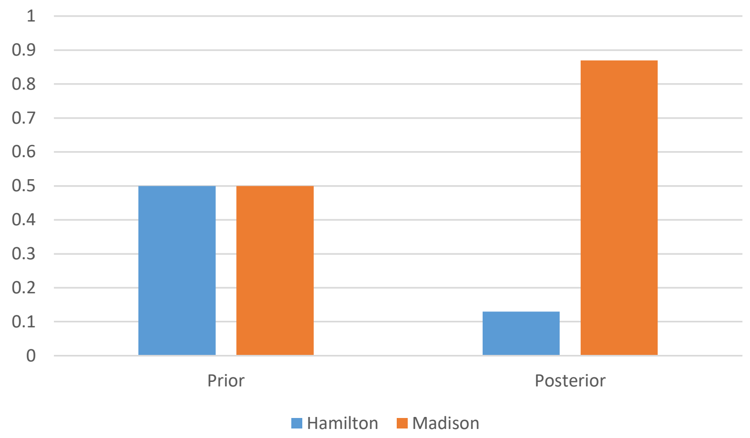
*71 affect	*79 city	*87 direction	*94 innovation	102 perhaps	*110 vigor
*72 again	*80 commonly	*88 disgracing	*95 join	*103 rapid	*111 violate
*73 although	*81 consequently	89 either	*96 language	104 same	*112 violence
74 among	*82 considerable	*90 enough (and in sample of 20)	97 most	105 second	*113 voice
75 another	*83 contribute	*91 fortune	98 nor	106 still	114 where
76 because	*84 defensive	*92 function	*99 offensive	107 those	115 whether
77 between	*85 destruction	93 himself	100 often	*108 throughout	*116 while
78 both	86 did		*101 pass	109 under	*117 whilst

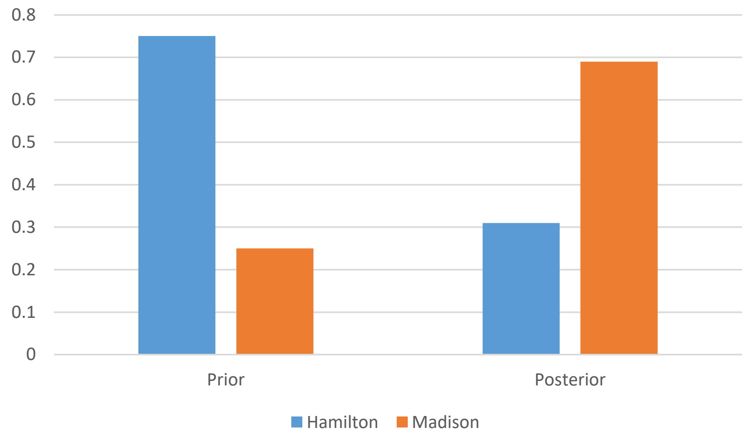
TABLE 2.7. NEW WORDS FROM THE WORD INDEX STUDY TOGETHER WITH THEIR CODE NUMBERS

118 about	130 choice	142 intrust+s+ed+ing	154 proper
119 according	131 common	143 kind	155 propriety
120 adversaries	132 danger	144 large	156 provision+s
121 after	133 decide+s+ed+ing	145 likely	157 requisite
122 aid	134 degree	146 matter+s	158 substance
123 always	135 during	147 moreover	159 they
124 apt	136 expence+s	148 necessary	160 though
125 asserted	137 expense+s	149 necessity+ies	161 truth+s
126 before	138 extent	150 others	162 us
127 being	139 follow+s+ed+ing	151 particularly	163 usage+s
128 better	140 I	152 principle	164 we
129 care	141 imagine+s+ed+ing	153 probability	165 work+s

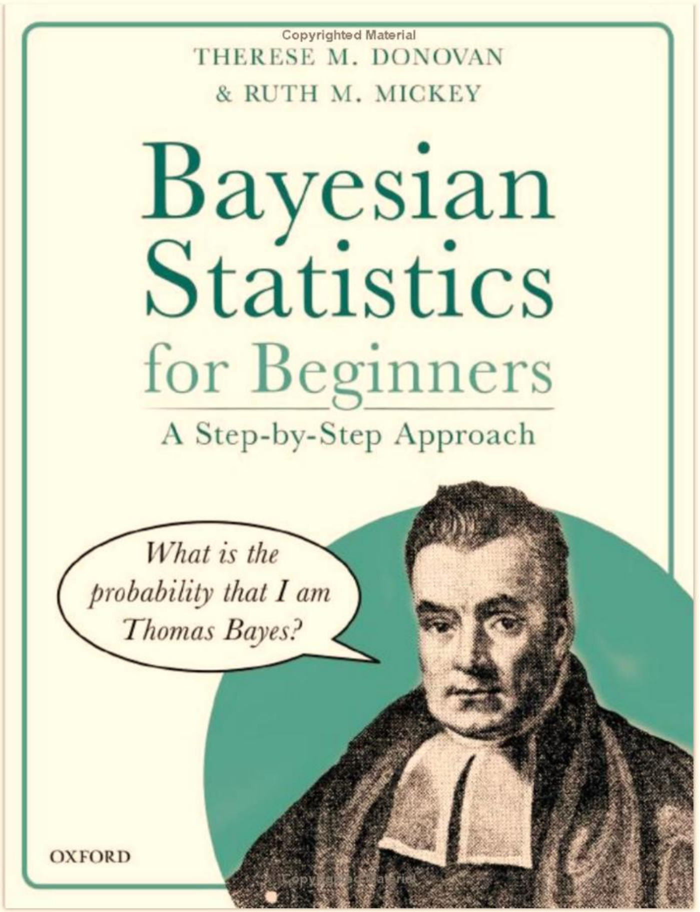
Case: The Authorship Problem

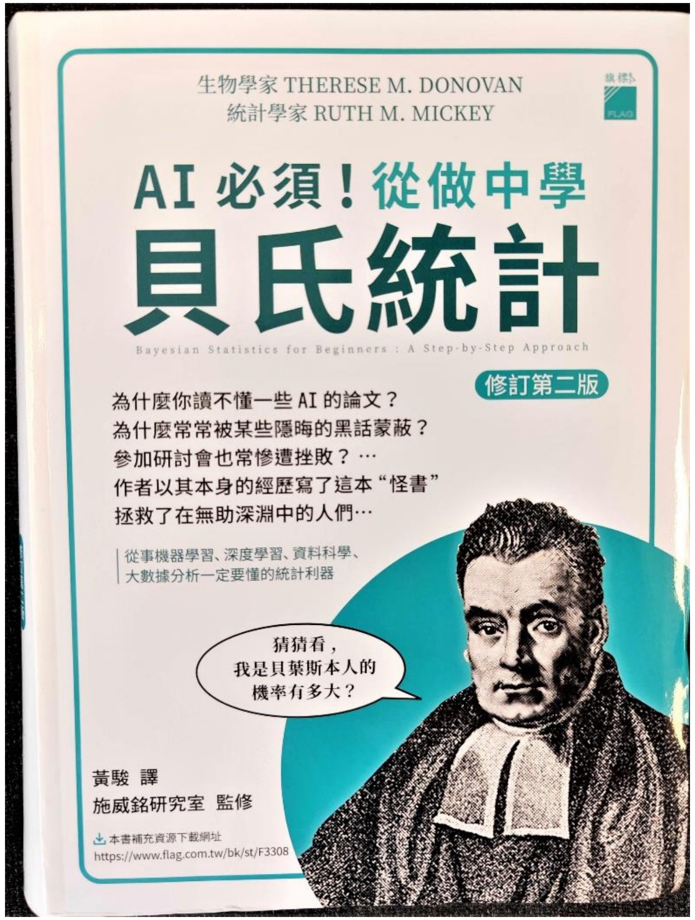
- $\Pr(data|H\_h) = 1/48$; $\Pr(data|H\_m) = 7/50$
- Hence, if $\Pr(H\_h) : \Pr(H\_m)$ = 1:1, then
    - $\Pr(H\_h|data)$ = 0.13
    - $\Pr(H\_h|data)$ / $\Pr(H\_m|data)$ = 0.13:0.87 = 0.15 < 1
- If $\Pr(H\_h) : \Pr(H\_m)$ = 0.75:0.25, then
    - $\Pr(H\_h|data)$ = 0.31
    - $\Pr(H\_h|data)$ / $\Pr(H\_m|data)$ = 0.31:0.69 = 0.4464 < 1
- Conclusion: the author is more likely to be Madison.





More on Bayesian Inference





Pricing with Unknown Demand

- A new product is introduced but demand is unknown.
- There are two possible price points: 0.5 and 1.0.
- But the distribution of customer willingness-to-pay (i.e., the highest acceptable price) is unknown. There are three possible distributions, which are equally likely:
  - (i) $P(wtp<0.5) = P(0.5<=wtp<1.0) = P(wtp>=1.0) = 1/3$
  - (ii) $P(wtp<0.5) = 1/2$; $P(0.5<=wtp<1.0) = 1/3$; $P(wtp>=1.0) = 1/6$
  - (iii) $P(wtp<0.5) = 1/6$; $P(0.5<=wtp<1.0) = 1/3$; $P(wtp>=1.0) = 1/2$
- Suppose the population size is large.

Pricing with Unknown Demand

- The price was set at 1.0 at the beginning.
- Two customer arrived. The first left and the second bought the product.
- Should you lower the price to 0.5 to maximize your profit?

- First update:

Scenarios	(i)	(ii)	(iii)	Marginal
Buy at 1.0	(1/3) * (1/3)	(1/6) * (1/3)	(1/2) * (1/3)	1/3
Leave	(2/3) * (1/3)	(5/6) * (1/3)	(1/2) * (1/3)	2/3
Prior	1/3	1/3	1/3	
Conditional	**1/3**	**5/12**	**1/4**	

Note that we do not observe WTP.

Pricing with Unknown Demand

- Second update:

Scenarios	(i)	(ii)	(iii)	Marginal
Buy at 1.0	(1/3) * (1/3)	(1/6) * (5/12)	(1/2) * (1/4)	11/36
Leave	(2/3) * (1/3)	(5/6) * (5/12)	(1/2) * (1/4)	25/36
Prior	1/3	5/12	1/4	
Conditional	**4/11**	**5/22**	**9/22**	

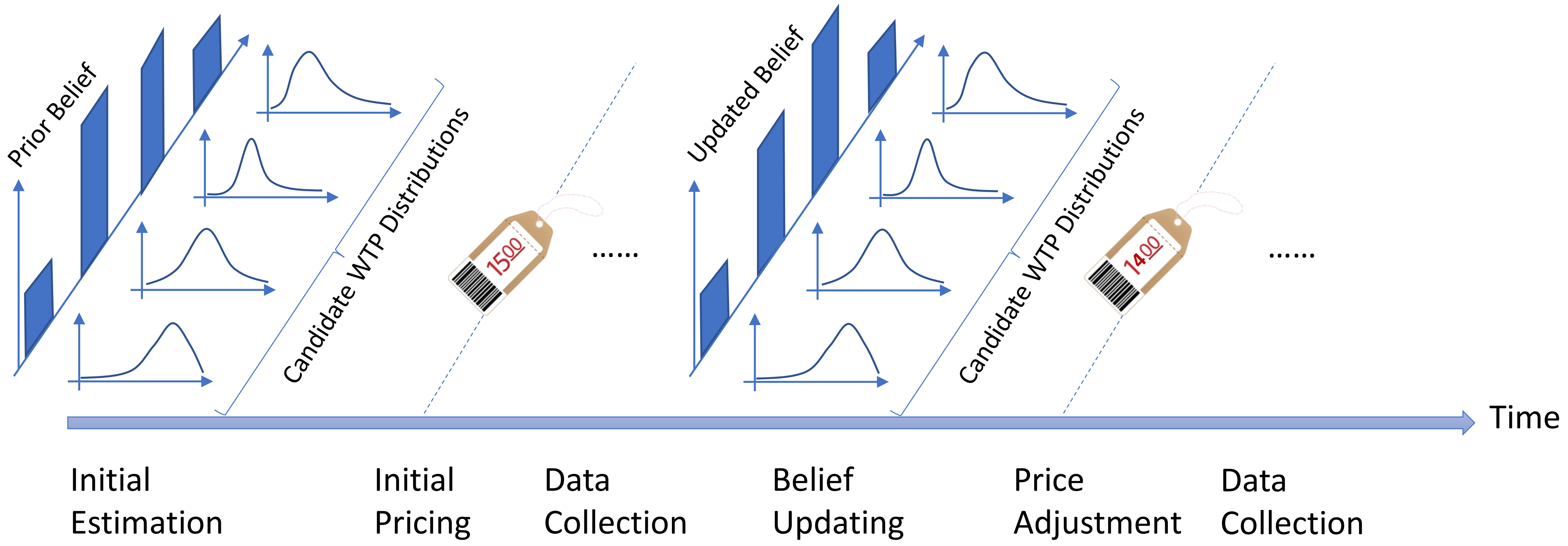
- Expected revenue of pricing at 1.0 =

$$(4/11)*(1/3)+(5/22)*(1/6)+(9/22)*(1/2) = 4/11 = 24/66$$

- Expected revenue of pricing at 0.5 =

$$0.5*[(4/11)*(2/3)+(5/22)*(1/2)+(9/22)*(5/6)] = 23/66$$

Pricing with Unknown Demand



After-Class Exercise

- Suppose $P(A) = 0.4$ and $P(B) = 0.6$. Are A and B mutually exclusive?
- Suppose A and B are mutually exclusive and $P(A) = 0.4$. Then $P(B) =$?
- 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 $P(D)$?
- There are two fortune tellers, A & B. According to historical data, A's predictions were correct in 90% cases, while B's predictions were correct only in 30% cases. 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 corrected belief?

After-Class Exercise

- When a man passes the airport security check, they discover a bomb in his bag. He explains. "Statisticians show that the probability of a bomb being on an airplane is 1/10,000. However, the chance that there are two bombs on one plane is 1/10,000,000. So, I am much safer …"
- Suppose the statisticians are right and it is impossible to have more than two bombs on an airplane. Do you agree with the man?

- If event A and B are independent given event C happens, then A and B are also independent given C does not happen. True or False?

- In the introductory example, what is the best price to set if the first customer did not buy? Consider three possible means of WTP: \$2,000, \$3,500, and \$5,000.