Conditional Probability

Formulas

CONDITIONAL PROBABILITY

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad (4.3)$$

or

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \quad (4.4)$$

Joint and Marginal Probabilities

Joint Probabilities

	No Default (D^c)	Default (D)	Total
Married (M)	0.2133	0.2633	0.4766
Single (S)	0.3867	0.1367	0.5234
Total	*0.6000*	*0.4000*	*1.0000*

Marginal Probabilities

Conditional Probabilities

Probability of Defaulting, Given the person is married:

$$P(D \mid M) = \frac{P(D \cap M)}{P(M)} = \frac{0.2633}{0.4766} = 0.5524$$

Probability of Defaulting, Given the person is single:

$$P(D \mid S) = \frac{P(D \cap S)}{P(S)} = \frac{0.1367}{0.5234} = 0.2611$$

Conditional Probability

Multiplication Law:

- **Multiplication law** can be used to calculate the probability of the intersection of two events.
- Based on the definition of conditional probability.
  - Rearrange formula to solve for P(A n B)

MULTIPLICATION LAW

$$P(A \cap B) = P(B)P(A \mid B) \quad \textbf{(4.7)}$$

or

$$P(A \cap B) = P(A)P(B \mid A) \quad \textbf{(4.8)}$$

Multiplication Law:

Joint Probabilities

	No Default (D^c)	Default (D)	Total
Married (M)	0.2133	0.2633	0.4766
Single (S)	0.3867	0.1367	0.5234
Total	0.6000	0.4000	1.0000

Marginal Probabilities

Example:

$$P(D|M) = \frac{P(D \cap M)}{P(M)} = \frac{0.2633}{0.4766} = 0.5524$$

$$P(D \cap M) = P(M)P(D|M) = (0.4766)(0.5524) = 0.2633$$

Conditional Probability

- Special case in which events *A* and *B* are independent.
- To compute the probability of the intersection of two independent events
  - Simply multiply the probabilities of each event.

MULTIPLICATION LAW FOR INDEPENDENT EVENTS

$$P(A \cap B) = P(A)P(B) \tag{4.9}$$

Conditional Probability

Bayes' Theorem:

- Way to update probabilities when we learn new information
- Example: A manufacturing firm receives shipments of parts from two different suppliers:
  - 65% of the parts purchased from supplier 1.
  - 35% of the parts purchased from supplier 2.
- These are called **Prior Probabilities**
  - estimates for specific events of interest

Conditional Probability

Historical data suggest the quality ratings of the two suppliers:

	% Good Parts	% Bad Parts
Advance Auto (A1)	98	2
Auto Zone (A2)	95	5

- This new information allows us to update the prior probabilities.
- To do so, we calculate conditional probabilities

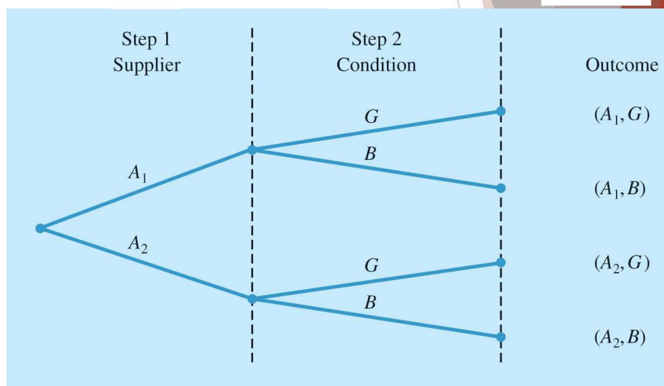
$$P(G|A_1) = 0.98 \qquad P(B|A_1) = 0.02$$

$$P(G|A_2) = 0.95 \qquad P(B|A_2) = 0.05$$

Conditional Probability

Diagram for Two-Supplier Example

$P(A_1)$ $P(G|A_1)$ $P(A_1,G)$



Note: Step 1 shows that the part comes from one of two suppliers and Step 2 shows whether the part is good or bad.

- There are 4 possible outcomes:
  - Each outcome is the intersection of 2 events
  - Example: Probability of getting a good part from Advanced Auto

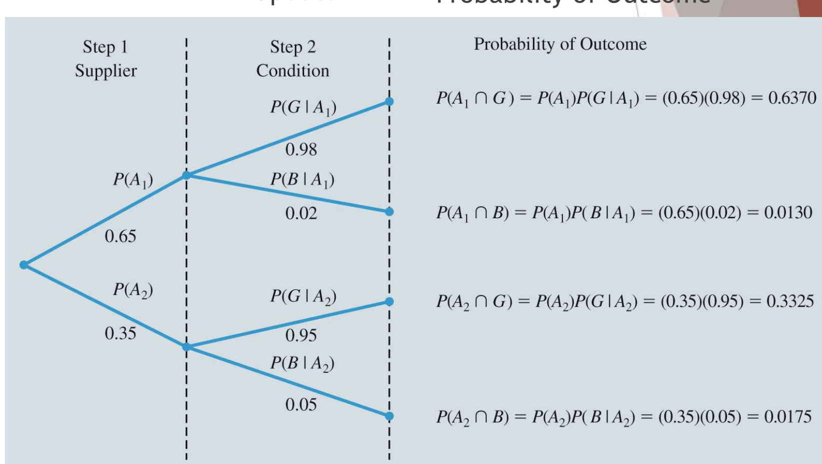
$$P(A_1,G) = P(A_1 \cap G) = P(A_1)P(G|A_1)$$

  - We use the multiplication rule to calculate

Conditional Probability

Figure 4.8: Probability Tree for Two-Supplier Example

Prior X Update = Probability of Outcome



- Simply multiply the probabilities of each branch to calculate the probability of the outcome.

Conditional Probability

- Suppose the parts from the two suppliers are used in the firm's manufacturing process and a machine breaks while attempting the process using a bad part:
  - Given the information that the part is bad, what is the probability that it came from supplier 1 and what is the probability that it came from supplier 2?
  - That is: What is
    - P(A1|B) = ?
    - P(A2|B) = ?
  - Using the prior probabilities and the conditional probabilities to update them, we can use Bayes Rule to calculate P(A1|B) and P(A2|B)
    - Known as **Posterior Probabilities**

Conditional Probability

BAYES' THEOREM (TWO-EVENT CASE)

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)} \quad \textbf{(4.10)}$$

$$P(A_2 \mid B) = \frac{P(A_2)P(B \mid A_2)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)} \quad \textbf{(4.11)}$$

BAYES' THEOREM

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots + P(A_n)P(B \mid A_n)} \quad \textbf{(4.12)}$$

- Bayes' theorem is applicable when are mutually exclusive and their union is the entire sample space.

Bayes Theorem

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2)}$$

$$= \frac{(0.65)(0.02)}{(0.65)(0.02) + (0.35)(0.05)} = \frac{0.0130}{0.0130 + 0.0175}$$

$$= \frac{0.0130}{0.0305} = 0.4262$$

$$P(A_2 \mid B) = \frac{P(A_2)P(B \mid A_2)}{P(A_2)P(B \mid A_1) + P(A_2)P(B \mid A_2)}$$

$$= \frac{(0.35)(0.05)}{(0.65)(0.02) + (0.35)(0.05)} = \frac{0.0175}{0.0130 + 0.0175}$$

$$= \frac{0.0175}{0.0305} = 0.5738$$

Random Variables

Discrete Random Variables

Continuous Random Variables

Random Variables

- A **random variable**
  - is a numerical description of the outcome of a random experiment.
- Random variables are quantities whose values are not known with certainty.
- A random variable can be classified as being either:
  - Discrete.
  - Continuous.

Random Variables

- **Why are they called Random Variables?!?**
  - A random variable is neither random, nor a variable!
- **Examples:**
  - the number of children in a family
  - the Friday night attendance at a cinema
  - the number of patients in a doctor's surgery
  - the number of defective light bulbs in a box of ten

Could there be patterns, reasons, or conscious decisions related to these “phenomena”?

Random Variables

Table 4.7: Examples of Discrete Random Variables – only take on specified discreet values

Random Experiment	Random Variable (x)	Possible Values for the Random Variable
Flip a coin	Face of a coin showing	1 if heads; 0 if tails
Roll a die	Number of dots showing on top of die	1, 2, 3, 4, 5, 6
Contact five customers	Number of customers who place an order	0, 1, 2, 3, 4, 5
Operate a health care clinic for one day	Number of patients who arrive	0, 1, 2, 3, …
Offer a customer the choice of two products	Product chosen by customer	0 if none; 1 if choose product A; 2 if choose product B

Random Variables

- Lancaster Savings and Loan Example
  - Random Variable **x = 1** for Default on Mortgage or **x = 0** for No Default
  - Random Variable **y = 1** for Married or **y = 0** for Single
  - Random Variable **z = number of payments per year** (= 12 for 12 months)

Table 4.8: Joint Probability Table for Customer Mortgage Prepayments

	No Default ($x = 0$)	Default ($x = 1$)	$f(y)$
Married ($y = 1$)	0.2133	0.2633	*0.4766*
Single ($y = 0$)	0.3867	0.1367	*0.5234*
$f(x)$	*0.6000*	*0.4000*	*1.0000*

Random Variables

Continuous Random Variables:

- A random variable that may assume any numerical value in an interval or collection of intervals
  - Relatively few random variables are truly continuous
  - Many discrete random variables have a large number of potential outcomes
    - can be effectively modeled as continuous random variables.

Random Experiment	Random Variable (x)	Possible Values for the Random Variable
Customer visits a web page	Time customer spends on web page in minutes	$x \geq 0$
Fill a soft drink can (max capacity = 12.1 ounces)	Number of ounces	$0 \leq x \leq 12.1$
Test a new chemical process	Temperature when the desired reaction takes place (min temperature = 150°F; max temperature = 212°F)	$150 \leq x \leq 212$
Invest $10,000 in the stock market	Value of investment after one year	$x \geq 0$

Discrete Probability Distributions

Custom Discrete Probability Distribution

Expected Value and Variance

Discrete Uniform Probability Distribution

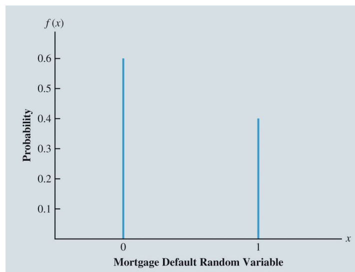
Binomial Probability Distribution

Poisson Probability Distribution

Discrete Probability Distributions

- The **probability distribution** for a random variable
  - describes the range and relative likelihood of possible values for a random variable.
- For a discrete random variable *x*, the probability distribution is defined by the **probability mass function**, denoted by $f(x)$.
- The probability mass function provides the probability for each value of the random variable.

Probability Distribution for Whether a Customer Defaults on a Mortgage



- f(0) = .6
- f(1) = .4

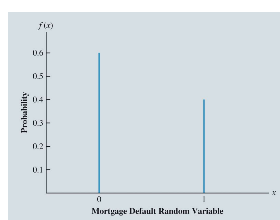
Requirements:

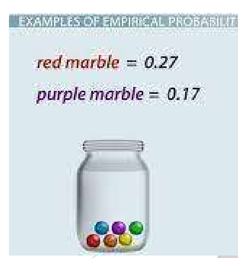
- $f(x) \geq 0$
- $\sum f(x) = 1$

Discrete Probability Distributions

- **Empirical Probability Distribution**
  - A probability that is generated from observations
  - is considered a **custom discrete probability distribution** if:
    - it is discrete
    - the possible values of the random variable have different values:

Useful for describing different possible scenarios that have different probabilities.





Discrete Probability Distributions

- Example: Lancaster Savings and Loan Example
  - The random variable describing the number of mortgage payments made per year by randomly chosen customers.

Table 4.10: Summary Table of Number of Payments Made per Year

	Number of Payments Made per Year			
	$x = 4$	$x = 12$	$x = 24$	Total
Number of observations	45	180	75	300
$f(x)$	0.15	0.60	0.25	

Discrete Probability Distributions

Expected Value and Variance:

- The **expected value**, or mean, of a random variable
  - is a measure of the central location for the random variable.
  - The weighted average of the values of the random variable, where weights are probabilities

EXPECTED VALUE OF A DISCRETE RANDOM VARIABLE

$$E(x) = \mu = \Sigma x f(x) \quad \textbf{(4.13)}$$

Discrete Probability Distributions

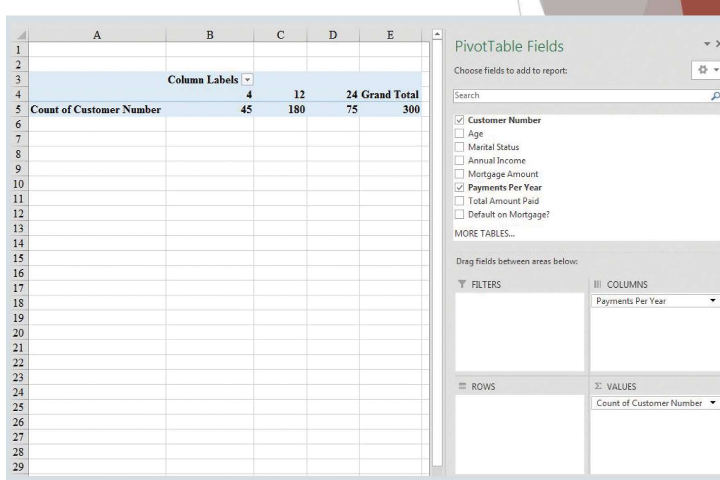
Expected Value for Number of Payments Made per Year by a Lancaster Savings and Loan Mortgage Customer

x	$f(x)$	$xf(x)$
4	0.15	$(4)(0.15) = 0.6$
12	0.60	$(12)(0.60) = 7.2$
24	0.25	$(24)(0.25) = 6.0$
		13.8 ← $E(x) = \mu = \sum xf(x)$

If Lancaster Savings and Loan signs a new mortgage customer, the expected number of payments per year for this customer is 13.8.

Discrete Probability Distributions

Figure 4.10: Excel PivotTable for Number of Payments Made per Year



	A	B	C	D	E
3		Column Labels			
4		4	12	24	Grand Total
5	Count of Customer Number	45	180	75	300

Discrete Probability Distributions

Figure 4.11: Using Excel SUMPRODUCT Function to Calculate the Expected Value for Number of Payments Made per Year by a Lancaster Savings and Loan Mortgage Customer

	A	B	C	D
1	*x*	*f(x)*		
2	4	0.15		
3	12	0.6		
4	24	0.25		
5				
6	**Expected Value:**	=SUMPRODUCT(A2:A4,B2:B4)		
7				
8				
9				
10				

	A	B	C	D
1	*x*	*f(x)*		
2	4	0.15		
3	12	0.60		
4	24	0.25		
5				
6	**Expected Value:**	13.8		
7				
8				
9				
10				

Discrete Probability Distributions

Figure 4.12: Excel Calculation of the Expected Value for Number of Payments Made per Year by a Lancaster Savings and Loan Mortgage Customer

	A	B	C	D	E	F	G	H
1	**Customer Number**	**Age**	**Marital Status**	**Annual Income**	**Mortgage Amount**	**Payments Per Year**	**Total Amount Paid**	**Prepay Mortgage?**
2	1	37	SINGLE	172125.7	473402.96	24	581885.13	YES
3	2	31	SINGLE	108571.04	300468.6	12	489320.38	NO
4	3	37	MARRIED	124136.41	330664.24	24	493541.93	YES
5	4	24	MARRIED	79614.04	230222.94	24	449682.09	YES
6	5	27	SINGLE	68087.33	282203.53	12	520581.82	NO
296	295	37	MARRIED	84791.08	179676.63	24	256361.65	YES
297	296	33	MARRIED	83498.89	235907.5	12	437145.85	NO
298	297	41	SINGLE	16597.53	151972.2	4	171289.87	YES
299	298	30	SINGLE	49293.95	186043.13	12	376694.27	NO
300	299	35	SINGLE	84241.8	194417.84	12	352597.79	NO
301	300	31	MARRIED	94428.15	264175.55	24	434102.49	YES
302								
304					**Expected Value:**	=AVERAGE(F2:F301)		
305					**Variance:**	=VAR.P(F2:F301)		
306					**Standard Deviation:**	=STDEV.P(F2:F301)		

	A	B	C	D	E	F	G	H
1	**Customer Number**	**Age**	**Marital Status**	**Annual Income**	**Mortgage Amount**	**Payments Per Year**	**Total Amount Paid**	**Prepay Mortgage?**
2	1	37	SINGLE	$ 172,125.70	$ 473,402.96	24	$ 581,885.13	YES
3	2	31	SINGLE	$ 108,571.04	$ 300,468.60	12	$ 489,320.38	NO
4	3	37	MARRIED	$ 124,136.41	$ 330,664.24	24	$ 493,541.93	YES
5	4	24	MARRIED	$ 79,614.04	$ 230,222.94	24	$ 449,682.09	YES
6	5	27	SINGLE	$ 68,087.33	$ 282,203.53	12	$ 520,581.82	NO
296	295	37	MARRIED	$ 84,791.08	$ 179,676.63	24	$ 256,361.65	YES
297	296	33	MARRIED	$ 83,498.89	$ 235,907.50	12	$ 437,145.85	NO
298	297	41	SINGLE	$ 16,597.53	$ 151,972.20	4	$ 171,289.87	YES
299	298	30	SINGLE	$ 49,293.95	$ 186,043.13	12	$ 376,694.27	NO
300	299	35	SINGLE	$ 84,241.80	$ 194,417.84	12	$ 352,597.79	NO
301	300	31	MARRIED	$ 94,428.15	$ 264,175.55	24	$ 434,102.49	YES
302								
304					**Expected Value:**	13.8		
305					**Variance:**	42.360		
306					**Standard Deviation:**	6.508		

Discrete Probability Distributions

- **Variance** is a measure of variability in the values of a random variable:

VARIANCE OF A DISCRETE RANDOM VARIABLE

$$\text{Var}(x) = \sigma^2 = \Sigma(x - \mu)^2 f(x) \tag{4.14}$$

- An essential part of the variance formula is the deviation, $x - \mu$,
  - which measures how far a particular value of the random variable is from the expected value, or mean, μ.

Discrete Probability Distributions

Calculation of the Variance for Number of Payments Made per Year by a Lancaster Savings and Loan Mortgage Customer

x	$x - \mu$	$f(x)$	$(x - \mu)^2 f(x)$
4	$4 - 13.8 = -9.8$	0.15	$(-9.8)^{2*}\ 0.15 = 15.606$
12	$12 - 13.8 = -1.8$	0.60	$(-1.8)^{2*}\ 0.60 = 2.904$
21	$21 - 13.8 = 10.2$	0.25	$(10.2)^{2*}\ 0.25 = 24.010$
			$42.360 \leftarrow \sigma^2 = \sum(x - \mu)^2 f(x)$

- The standard deviation, σ, is defined as the positive square root of the variance.
- The standard deviation for the payments made per year by a mortgage customer is $\sqrt{42.360} = 6.508$.

Discrete Probability Distributions

Figure 4.13: Excel Calculation of the Variance for Number of Payments Made per Year by a Lancaster Savings and Loan Mortgage Customer

	A	B	C	D
1	x	$f(x)$	$(x - \mu)^2$	
2	4	0.15	=(A2-B6)^2	
3	12	0.6	=(A3-B6)^2	
4	24	0.25	=(A4-B6)^2	
5				
6	**Expected Value:**	=SUMPRODUCT(A2:A4,B2:B4)		
7				
8	**Variance:**	=SUMPRODUCT(B2:B4,C2:C4)		
9				
10	**Stadard Deviation:**	=SQRT(B8)		

	A	B	C	D
1	x	$f(x)$	$(x - \mu)^2$	
2	4	0.15	96.04	
3	12	0.60	3.24	
4	24	0.25	104.04	
5				
6	**Expected Value:**	13.8		
7				
8	**Variance:**	42.360		
9				
10	**Stadard Deviation:**	6.508		