Chapter 16

Understanding Probability and Long-Term Expectations

Thought Question 1:

Two very different queries about probability:

a. If you **flip a coin** and do it fairly, what is the ***probability that it will land heads up***?

b. What is the ***probability that you will eventually own a home***; that is, how likely do you think it is? (If you already own a home, what is the probability that you will own a different home within the next 5 years?)

For which question was it *easier* to provide a *precise answer*? Why?

Thought Question 2:

Explain what it means for someone to say that the probability of his or her eventually owning a home is 70%.

Thought Question 3:

Explain ***what's wrong*** with this partial answer to Thought Question 1:

"The probability that I will eventually own a home, or of any other particular event happening, is 1/2 because either it will happen or it won't."

Thought Question 4:

Why do you think insurance companies charge young men more than they do older men for automobile insurance, but charge older men more for life insurance?

Thought Question 5:

How much would you be willing to pay for a ticket to a contest in which there was a ***1% chance that you would win $500*** and a ***99% chance that you would win nothing***?

Explain your answer.

16.1 Probability

What does the word ***probability*** mean?

Two distinct interpretations:

- For the probability of winning a lottery based on buying a single ticket -- we can quantify the chances exactly.
- For the probability that we will eventually buy a home -- we are basing our assessment on personal beliefs about how life will evolve for us.

16.2 The Relative-Frequency Interpretation

Relative-frequency interpretation: applies to situations that can be repeated over and over again.

Examples:

- Buying a weekly lottery ticket and observing whether it is a winner.
- Commuting to work daily and observing whether a certain traffic signal is red when we encounter it.
- Testing individuals in a population and observing whether they carry a gene for a certain disease.
- Observing births and noting if baby is male or female.

Idea of Long-Run Relative Frequency

Probability =
proportion of time it occurs over the long run

Long-run relative frequency of males born in the United States is about 0.512. (*Information Please Almanac,* 1991, p. 815)

Possible results for relative frequency of male births:

Weeks of Watching	1	4	12	24	36	52
Number of boys	12	47	160	310	450	618
Number of babies	30	100	300	590	880	1200
Proportion of boys	.400	.470	.533	.525	.511	.515

Proportion of male births jumps around at first but starts to settle down just above .51 in the ***long run***.

Determining the Probability of an Outcome

Method 1: Make an Assumption about the Physical World

Example: assume coins made such that they are equally likely to land with heads or tails up when flipped.

⇨ probability of a flipped coin showing heads up is ½.

Method 2: Observe the Relative Frequency

Example: observe the relative frequency of male births in a given city over the course of a year.

In 1987 there were a total of 3,809,394 live births in the U.S., of which 1,951,153 where males.

⇨ probability of male birth is 1,951,153/3,809,394 = 0.5122

Summary of Relative-Frequency Interpretation of Probability

- Can be applied when situation can be repeated numerous times and outcome observed each time.
- Relative frequency should settle down to constant value over long run, which is the probability.
- Does not apply to situations where outcome one time is influenced by or influences outcome the next time.
- Cannot be used to determine whether outcome will occur on a single occasion but can be used to predict long-term proportion of times the outcome will occur.

16.3 The Personal-Probability Interpretation

Person probability: the degree to which a given individual believes the event will happen.

They must be between 0 and 1 and be ***coherent***.

Examples:

- Probability of finding a parking space downtown on Saturday.
- Probability that a particular candidate for a position would fit the job best.

16.4 Applying Some Simple Probability Rules

Rule 1: If there are only two possible outcomes in an uncertain situation, then their probabilities must add to 1.

Example 1: If probability of a single birth resulting in a boy is 0.51, then the probability of it resulting in a girl is 0.49.

Simple Probability Rules

Rule 2:
If two outcomes cannot happen simultaneously, they are said to be **mutually exclusive.**
The **probability of one or the other** of two mutually exclusive outcomes happening is the **sum** of their individual probabilities.

Example 5: If you think probability of getting an A in your statistics class is 50% and probability of getting a B is 30%, then probability of getting either an A or a B is 80%. Thus, probability of getting C or less is 20% (using Rule 1).

Simple Probability Rules

Rule 3: If two events do not influence each other, and if knowledge about one doesn't help with knowledge of the probability of the other, the events are said to be **independent** of each other. If two events are independent, the **probability that they both happen** is found by **multiplying** their individual probabilities.

Example 7: Woman will have two children. Assume outcome of 2^{nd} birth independent of 1^{st} and probability birth results in boy is 0.51. Then probability of a boy followed by a girl is (0.51)(0.49) = 0.2499. About a 25% chance a woman will have a boy and then a girl.

Simple Probability Rules

Rule 4: If the ways in which one event can occur are a subset of those in which another event can occur, then the probability of the subset event *cannot* be higher than the probability of the one for which it is a subset.

Example 10: Suppose you are 18 and speculating about your future. You decide the probability you will eventually get married and have children is 75%. By Rule 4, probability that you will eventually get married is at least 75%.

16.5 When Will It Happen?

Probability an outcome will occur on any given instance is p.
Probability the outcome will not occur is $(1 - p)$.
Outcome each time is *independent* of outcome all other times.

Probability it *doesn't* occur on 1st try but *does* occur on 2nd try is $(1 - p)p$.

Try on Which the Outcome First Happens	Probability
1	p
2	$(1 - p)p$
3	$(1 - p)(1 - p)p = (1 - p)^2p$
4	$(1 - p)(1 - p)(1 - p)p = (1 - p)^3p$
5	$(1 - p)(1 - p)(1 - p)(1 - p)p = (1 - p)^4p$

Example 11: Number of Births to First Girls

Probability of a birth resulting in a **boy is about .51**, and the probability of a birth resulting in a girl is about .49. Suppose couple will **continue** having children **until** have a **girl**. Assuming outcomes of births are **independent** of each other, probabilities of having the first girl on the first, second, third, fifth, and seventh tries are shown below.

Number of Births to First Girl	Probability
1	.49
2	(.51)(.49) = .2499
3	(.51)(.51)(.49) = .1274
5	(.51)(.51)(.51)(.51)(.49) = .0331
7	(.51)(.51)(.51)(.51)(.51)(.51)(.49) = .0086

Accumulated Probability

Probability of 1st occurrence *not happening* by occasion n is $(1 - p)^n$.

Probability 1st occurrence *has happened* by occasion n is $[1 - (1 - p)^n]$.

Example 12: Getting Infected with HIV

Suppose probability of infected is 1/500 = 0.002.
Probability of not infected is 0.998. So probability of infection *by 2nd encounter* is $[1 - (1 - 0.002)^2] = 0.003996$.

Number of Encounters	Probability of First Infection	Accumulated Probability of HIV
1	.002	.002
2	(.998)(.002) = .001996	.003996
4	$(.998)^3(.002)$ = .001988	.007976
10	$(.998)^9(.002)$ = .001964	.019821

16.6 Long-Term Gains, Losses, and Expectations

Long-Term Outcomes Can Be Predicted

Example 14: Insurance Policies

A simple case: All customers charged $500/year, 10% submit a claim in any given year and claims always for $1500.

How much can the company expect to make per customer?

Claim Paid?	Probability	Amount Gained
Yes	.10	–$1000
No	.90	+$ 500

Average Gain = .90 ($500) – .10 ($1000) = $350 per customer

Expected Value (EV)

EV = expected value = $A_1p_1 + A_2p_2 + A_3p_3 + \ldots + A_kp_k$
where A_1, A_2, A_3, …, A_k are the possible amounts
and $p_1, p_2, p_3, \ldots, p_k$ are the associated probabilities.

***Example 14, continued*:**

$A_1p_1 + A_2p_2$ = (.10)(−\$1000) + (.90)(\$500) = \$350
= expected value of company's profit per customer.

Key: Expected value is the average value per measurement ***over the long run*** and not necessarily a typical value for any one occasion or person.

Example 15: California Decco Lottery Game

Decco: player chooses one card from each of four suits. A winning card drawn from each suit. Prizes awarded based on the number of matches. Cost to play = $1.

Number of Matches	Prize	Net Gain	Probability
4	$5000	$4999	1/28,561 = .000035
3	$50	$49	1/595 = .00168
2	$5	$4	1/33 = .0303
1	Free ticket	0	.2420
0	None	–$1	.7260

EV = (\$4999)(1/28,561) + (\$49)(1/595) + (\$4)(.0303) + (–\$1)(.726)
= –$0.35

Over many repetitions, players will lose an average of 35 cents per play. The Lottery Commission will only pay out about 65 cents for each $1 ticket sold.

Expected Value as Mean Number

If measurement taken over a large group of individuals, the expected value can be interpreted as the **mean value** per individual.

Example:

Suppose 40% of people in a population smoke a pack of cigarettes a day (20 cigarettes) and remaining 60% smoke none.

Expected number of cigarettes smoked per day by one person:

EV = (0.40)(20 cigarettes) + (0.60)(0 cigarettes) = 8 cigarettes

On average 8 cigarettes are smoked per person per day.
Here again, the EV is not a value we actually *expect* to measure on any one individual.

Case Study 16.1: Birthdays and Death Days – Is There a Connection?

Study Details:

Data from death certificates of adults in California who died of natural causes between 1979 and 1990. Each death classified as to how many weeks after the birthday it occurred. Compared actual numbers of deaths to expected number (seasonally adjusted).

Results:

Women: biggest peak in Week 0; highest number of women died the week *after* their birthdays.

Men: biggest peak was in Week 51; highest number of men died the week *before* their birthdays.

Probability biggest peak for women in Week 0 ***AND*** biggest peak for men in Week 51 = (1/52)(1/52) = 0.0004

Source: Phillips, Van Voorhies, and Ruth, 1992

For Those Who Like Formulas

Notation

Denote "events" or "outcomes" with capital letters *A, B, C,* and so on.
If A is one outcome, all other possible outcomes are part of "A complement" = A^C.

$P(A)$ is the probability that the event or outcome A occurs. For any event A,
$0 \leq P(A) \leq 1$.

Rule 1

$$P(A) + P(A^C) = 1$$

A useful formula that results from this is

$$P(A^C) = 1 - P(A)$$

Rule 2

If events A and B are *mutually exclusive,* then

$$P(A \text{ or } B) = P(A) + P(B)$$

Rule 3

If events A and B are *independent,* then

$$P(A \text{ and } B) = P(A)\, P(B)$$

Rule 4

If the ways in which an event B can occur are a subset of those for event $A,$ then

$$P(B) \leq P(A)$$