

**Class 5: Probability (Text: Sections 4.1 and 4.2)****WHAT IS PROBABILITY?**

**Probability** measures the likelihood of an event whose outcome cannot be predicted. For example, saying

$$P(\text{rain}) = 30\%$$

means that on 30% of the days where the weather looked like this, it rained.

“Rain” is an *event*; the probability of an event is a number between 0 and 1.

*Sample space* is all possible events and *Outcomes* is group of events

**Long-run frequency interpretation of probability:**

$$P(\text{rain}) = \frac{\# \text{ days rain}}{\# \text{ days}} = 30\%$$

Ex: For fair coin:

$$P(\text{heads}) = \frac{\# \text{ heads}}{\# \text{ tosses}} = 50\%$$

Proportion heads = 50% in the long run. It may be different in the short run. Why?

Random variation averages out in the long run, not in the short run.

**MUTUALLY EXCLUSIVE AND COMPLEMENTARY EVENTS**

Example: One step to improving children’s dental care is to understand how often they go to the dentist. A 2005 report from National Center for Health Statistics<sup>1</sup> found the following distribution of times since the last visit.

(Assume there is no overlap between categories.)

Length of time	Less than 6 mos	6 mos - 1 year	1 year - 2 years	2 years - 5 years	More than 5 years
Proportion	0.57	0.18	0.08	0.03	?

- What is the probability that a randomly chosen child has been to the dentist within the previous year?
- What is the probability that a randomly chosen child has not been to a dentist within the last 6 months?
- Find the probability that a randomly chosen child has not gone to the dentist for more than 5 years.

(a) We have  $P(\text{within 1 year}) = 0.57 + 0.18 = 0.75$ .

(b) We have  $P(\text{not in last 6 mos}) = 1 - 0.57 = 0.43$ .

(c) To sum to 1, we have  $P(\text{more than 5 years}) = 1 - (0.57 + 0.18 + 0.08 + 0.03) = 0.14$ .

Notice that in part (a) the probabilities “less than 6 mos” and “6 mos to 1 year” add because the events cannot happen at once; we say they are **mutually exclusive events**.

In general if  $A$  and  $B$  are mutually exclusive events

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

In part (b), we say that “in last 6 mos” and “not in last 6 mos” are **complementary events** and their probabilities add to 1, because one of the two events has to happen. Thus

In general if  $A$  and  $A^c$  are complementary events

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

<sup>1</sup> “Summary Health Statistics for US Children: National Health Interview Survey, 2003” by A.D. Det et al, 2005, Table 17. Reported by Baldi and Moore in IPS Life Sciences.

**INDEPENDENT EVENTS**

Events are called **independent** if they “don’t affect one another.”

Ex: Rain in Tucson and whether or not your cousin catches the train in New York are independent. (Unless you call her because of the rain and delay her.....)

Ex: Rain and sun in Tucson are not independent. Also, rain and no rain are not independent.

**Gender and handedness:**

Ex: Since 13% of men and 13% of women are left handed (LH), we say handedness is *independent* of gender. This tells that the two conditional distributions in the table to the left below are equal.

Ex: In a population that is 50% male, find  $P(LH \text{ and } M)$ . It may be helpful to imagine a population of 100 people and fill in the figure to the right. This tells us

$$P(LH \text{ and } M) = 6.5/100 = 0.065$$

**Handedness: Conditional Distribution**

	Men	Women
LH	0.13	0.13
RH	0.87	0.87
	1	1

**Handedness: Number of people**

LH	6.5	6.5
RH	43.5	43.5
	Men: 50	Women: 50

Two events are **independent** if their conditional distributions are the same.

Notice that the fact that gender and handedness are independent tell us that 13% of the men *and* 13% of the women are left handed. Thus we have  $P(LH) = 0.13$  and  $P(M) = 0.5$ , so

$$P(LH \text{ and } M) = (0.13)(0.5) = 0.065$$

**Test for Independence:** In general if events  $A, B$  are independent,

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

This *only* works if the events are independent!

**Gender and Colorblindness**

Ex: Given that 0.4% of women, 7% of men are color blind (CB), are gender and color blindness independent? No; see the two different conditional distributions in the table to the left.

Ex: There are 50% women in the population. Find  $P(CB)$ , the proportion of color blind people in the population. For example, do this by filling out the figure to the right below:

**Color Blindness: Conditional Distribution**

	Men	Women
CB	0.07	0.004
Not CB	0.93	0.996
	1	1

**Colorblindness: Number of people**

CB	3.5	0.2
Not CB	46.5	49.8
	Men: 50	Women: 50

Thus  $P(CB) = \frac{(3.5 + 0.2)}{100} = 0.037 = 3.7\%$

Notice that the value is between 7% and 0.4%. Since there equal proportions of the two genders, the proportion of CB in the population is the average:

$$P(CB) = \frac{0.07 + 0.004}{2} = \frac{0.074}{2} = 0.037.$$

Ex: Find  $P(CB \text{ and } M)$  and  $P(CB) \cdot P(M)$ . Comment: What do you notice? Are color blindness and gender independent?

We have  $P(CB \text{ and } M) = \frac{3.5}{100} = 0.035$  and  $P(CB) \cdot P(M) = (0.037) \cdot (0.5) = 0.0185$ .

Thus  $P(CB \text{ and } M) \neq P(CB) \cdot P(M)$ , so color blindness and gender are **not** independent. Different proportions of men and of women are color blind.

**USING INDEPENDENCE TO CALCULATE PROBABILITIES**

Ex: The sex of babies in successive births are independent. (Why?) Find the probability of all girls in family of three children. Assume the probability of a girl is  $\frac{1}{2}$ :

$$P(GGG) = (P(G))^3 = \left(\frac{1}{2}\right)^3 = \frac{1}{8} = 0.125$$

Ex: Now assume the probability of a girl is 48%. (This is the actual figure.) Find the probability of all girls in a family of three.

$$P(GGG) = (0.48)^3 = 0.1106$$

Ex: What is the probability of two girls and a boy in a family of three children, if the boy is the youngest? (Use  $P(G) = 0.48$ .) The probability of a boy is  $P(B) = 1 - 0.48 = 0.52$ . Thus

$$P(GGB) = (0.48)^2(0.52) = 0.1198$$

Ex: What is the probability of two girls and one boy in a family of three children, in any order. (Again use  $P(G) = 0.48$ .)

Now there are three possibilities, with the boy the youngest, the middle, or the oldest. Each of the three possibilities has that same probability, 0.1198. Thus

$$\begin{aligned} P(\text{one boy}) &= P(GGB) + P(GBG) + P(BGG) = (0.48)^2(0.52) + (0.48)(0.52)(0.48) + (0.52)(0.48)^2 \\ &= 3(0.48)^2(0.52) = 3(0.1198) = 0.3594. \end{aligned}$$

Thus there is about a 35% chance of having one boy and two girls in the family.

Ex: What is the probability of one girl and two boys in a family of three children, in any order. (Again use  $P(G) = 0.48$ .)

There are three possibilities, but the probabilities are reversed, so

$$\begin{aligned} P(\text{one girl}) &= P(GBB) + P(BGB) + P(BBG) = (0.48)(0.52)^2 + (0.52)(0.48)(0.52) + (0.52)^2(0.48) \\ &= 3(0.52)^2(0.48) = 3(0.1298) = 0.3894. \end{aligned}$$

As expected, the result is slightly higher as boys are slightly more likely.

Ex: The natural sex ratio of births is not 1:1; there are more boys born. The actual ratio is therefore something like 1.05:1 (These ratios are given as Male: Female.) What is the sex ratio if 48% of births are girls?

Suppose there are 100 girls born and  $x$  boys born. Then the girls are 48% of the births and the boys are 52%, so

$$\frac{x}{100} = \frac{0.52}{0.48}$$

so

$$x = \frac{52}{0.48} = 1.0833.$$

Thus the ratio is 1.0833 to 1 (male:female).

Ex: If people who do not have HIV are given an HIV test, there is a small probability that the test will come back positive. (Why?) These are called *false positive* test results, FP. Given that  $P(FP) = 0.4\%$ , find the probability that a person without HIV gets a negative test result.

$$P(\text{negative}) = P(\text{no } FP) = 1 - P(FP) = 1 - 0.004 = 0.996 = 99.6\%$$

Ex: If 1000 people who do not have HIV are screened, find the probability that there is at least one false positive.

$$P(\text{at least one FP}) = 1 - P(\text{no FP among 1000}) = 1 - (P(\text{no FP}))^{1000} = 1 - (0.996)^{1000} = 0.982. \text{ Almost certain!}$$

Ex: China's "one child per couple" policy has been in effect since 1979. This means that families that have more than one child can be fined. However, in rural areas where couples particularly want a boy, families may be allowed to have two children if the first one is a girl. Assume the sex ratio of births<sup>2</sup> is 1.06:1 (male to female).

- (a) What is the probability that a baby is a boy?  
 (b) Consider 1000 rural families who will have a second child if the first is a girl.  
 (i) How many of the families have just one child?  
 (ii) Of the families with two children, how many have two girls and how many have one child of each sex?

(a) For every 106 male births, there are 100 female births. Thus

$$\text{Probability of boy} = \frac{106}{106 + 100} = 0.515 = 51.5\%.$$

(b)

- i. There are  $(0.515)1000 = 515$  families that have a boy first and so have only one child.  
 ii. There are  $1000 - 515 = 485$  families that have two children. Of these,  $(0.515)485 = 250$  have a boy for their second child, and  $485 - 250 = 235$  families have two girls.

<sup>2</sup> Industrialized countries generally have a natural ratio between 1.03 and 1.07 to 1, not 1:1.