5.2 Testing a coin Your friend decides to flip a coin repeatedly to analyze whether the probability of a head on each flip is 1/2. He flips the coin 10 times and observes a head 7 times. He concludes that the probability of a head for this coin is $7/10 = 0.70$.

a. Your friend claims that the coin is not balanced, since the probability is not 0.50. What's wrong with your friend's claim?

b. If the probability of flipping a head is actually 1/2, what would you have to do to ensure that the cumulative proportion of heads falls very close to 1/2?

a. **The cumulative proportion o f heads fluctuates a lot with a short run.**

b. **We have to flip the coin many, many times.**

In the long run, the cumulative proportion approaches the actual probability of an outcome.

5.6 Random digits Consider a random number generator designed for equally likely outcomes. Which of the following is *not* correct, and why?

a. For each random digit generated, each integer between 0 and 9 has probability 0.10 of being selected.

b. If you generate 10 random digits, each integer between 0 and 9 must occur exactly once.

c. If you generated a very large number of random digits, then each integer between 0 and 9 would occur close to 10% of the time.

d. The cumulative proportion of times that a 0 is generated tends to get closer to 0.10 as the number of random digits generated gets larger and larger.

b is not correct because in the short run, probabilities of each digit being generated can fluctuate a lot.

5.8 Heart transplant Before the first human heart transplant, Dr. Christiaan Barnard of South Africa was asked to assess the probability that the operation would be successful. Did he need to rely on the relative frequency definition or the subjective definition of probability? Explain.

Subjective definition of probability.

He used his own judgment as there is no data available.

5.16 More true-false questions Your teacher gives a true-false pop quiz with 10 questions.

TRY

a. Show that the number of possible outcomes for the sample space of possible sequences of 10 answers is 1024.

b. What is the complement of the event of getting *at least* one of the questions wrong?

c. With random guessing, show that the probability of getting *at least* one question wrong is 0.999.

a. 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 = 1024

b. Getting all questions correctly answered.

c. P(at least one question wrong)

= 1 - P(all questions correct)

= 1 – 1/1024

= 0.999

5.20 Wrong sample space A couple plans on having four children. The father notes that the sample space for the number of girls the couple can have is 0, 1, 2, 3, and 4. He goes on to say that since there are five outcomes in the sample space, and since each child is equally likely to be a boy or girl, all five outcomes must be equally likely. Therefore, the probability of all four children being girls is 1/5. Explain the flaw in his reasoning.

The outcomes (numbers of girls) are not equally likely.

It is more likely to have 1, 2 or 3 girls than to have 0 or 4 girls.

5.24 Protecting the environment When the General Social Survey most recently asked subjects whether they are a member of an environmental group (variable GRNGROUP) and whether they would be very willing to pay higher prices to protect the environment (variable GRNPRICE), the results were as shown in the table. For a randomly selected American adult:

a. Estimate the probability of being (i) a member of an environmental group and (ii) willing to pay higher prices to protect the environment.

b. Estimate the probability of being both a member of an environmental group *and* very willing to pay higher prices to protect the environment.

c. Given the probabilities in part a, show that the probability in part b is larger than it would be if the variables were independent. Interpret.

d. Estimate the probability that a person is a member of an environmental group *or* very willing to pay higher prices to protect the environment. Do this (i) directly using the counts in the table and (ii) by applying the appropriate probability rule to the estimated probabilities found in parts a and b.

		Pay Higher Prices (GRNPRICE)		
		Yes	No	Total
Environmental Group Member (GRNGROUP)	Yes	69	15	84
	No	435	276	**711**
	Total	**504**	**291**	**795**

a. (i) 84/795 (ii) 504/795

b. 69/795 = 0.0868

c. If independent: it is (i) x (ii) = (84/795)(504/795) = 0.0670

d. (i) (69+15+435)/795 = 519/795 = 173/265 = 0.6528

(ii) (84+504-69) /795 = 519/795 = 173/265 = 0.6528

5.32 Cancer deaths Current estimates are that about 25% of all deaths are due to cancer, and of the deaths that are du[e] to cancer, 30% are attributed to tobacco, 40% to diet, an[d] 30% to other causes.

a. Define events, and identify which of these four probabilities refer to conditional probabilities.

b. Find the probability that a death is due to cancer and tobacco.

a. C = deaths that are due to cancer, T = tobacco, D = diet, O = others.
$P(C)= 0.25$, $P(T|C)=0.30$, $P(D|C)=0.40$, $P(O|C)=0.30$

b. $P(C \text{ and } T) = P(C) \times P(T|C) = (0.25)(0.30) = 0.075$

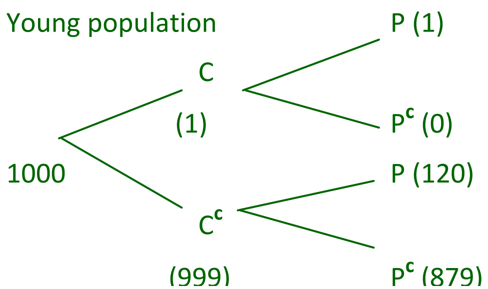
5.58 More screening for breast cancer Refer to the previous exercise. For young women, the prevalence of breast cancer is lower. Suppose the sensitivity is 0.86 and the specificity is 0.88, but the prevalence is only 0.001.

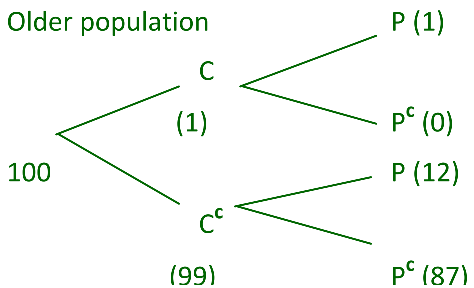
a. Given that a test comes out positive, find the probability that the woman truly has breast cancer.

b. Show how to use a tree diagram with frequencies for a typical sample of 1000 women to explain to someone who has not studied statistics why the probability found in part a is so low.

c. Of the cases that are positive, explain why the proportion in error is likely to be larger for a young population than for an older population.

Note: The prevalence is 0.01 in older population.





a. Let C = the event of having breast cancer = 0.001
P = the event of having a positive test outcome
Sensitivity = $P(P|C) = 0.86$; specificity = $P(P^c|C^c) = 0.88$
Thus $P(P) = P(C)\times P(P|C) + P(C^c)\times P(P|C^c)$
$= (0.001)(0.86)+(0.999)(1-0.88) = 0.12074$
$P(C|P) = P(C \text{ and } P)/ P(P) = [P(C)\times P(P|C)]/P(P)$
$= (0.001)(0.86)/(0.12074) = 0.0071227$

c. $P(C^c|P)$ is the probability of false positive.
Young population: $P(C^c|P) = 0.9929$ (higher)
Older population: $P(C^c|P) = 0.9310$ (lower)
Or from tree diagrams, 120 out of 121 were false positives among younger women while 12 out of 13 were false positives for older women. So the rate of false positives is higher in young population.

5.74 Independent on coffee? Students in a geography class are asked whether they've visited Europe in the past 12 months and whether they've flown on a plane in the past 12 months.

a. For a randomly selected student, would you expect these events to be independent or dependent? Explain.

b. How would you explain to someone who has never studied statistics what it means for these events to be either independent or dependent?

c. Students in a different class were asked whether they've visited Italy in the past 12 months and whether they've visited France in the past 12 months. For a randomly selected student, would you expect these events to be independent or dependent? Explain.

d. Students in yet another class were asked whether they've been to a zoo in the past 12 months and whether they drink coffee. For a randomly selected student, would you expect these events to be independent or dependent? Explain.

e. If you had to rank the pairs of events in parts a, c, and d in terms of the strength of any dependence, which pair of events is most dependent? Least dependent?

a. Dependent.
   If people has visited Europe, it is very likely people flew there.
b. The first event has no bearing on the second event.
c. Dependent. People are likely to visit more than one country.
d. Independent. They are unrelated.
e. Events in (a) most dependent,
   events in (d) the least dependent.

5.76 Teens and drugs In August 2006 the Center on Addiction and Substance Abuse (CASA) at Columbia University reported results of a survey of 1297 teenagers about their views on the use of illegal substances. Twenty percent of the teens surveyed reported going to clubs for music or dancing at least once a month. Of them, 26% said drugs were usually available at these club events. Which of these percentages estimates a conditional probability? For each that does, identify the event conditioned on and the event to which the probability refers.

Conditional probability = P(D|C) = 0.26
The probability refers to drug availability while going clubs is the condition

5.88 Screening for heart attacks Biochemical markers are used by emergency room physicians to aid in diagnosing patients who have suffered acute myocardial infarction (AMI), or what's commonly referred to as a heart attack. One type of biochemical marker used is creatine kinase (CK). Based on a review of published studies on the effectiveness of these markers (by E. M. Balk et al., *Annals of Emergency Medicine*, vol. 37, pp. 478–494, 2001), CK had an estimated sensitivity of 37% and specificity of 87%. Consider a population having a prevalence rate of 25%.

a. Explain in context what is meant by the sensitivity equaling 37%.

b. Explain in context what is meant by the specificity equaling 87%.

c. Construct a tree diagram for this diagnostic test. Label the branches with the appropriate probabilities.