Chapter 3

Probability

3.1 Introduction

In addition to describing data, we might want to test specific inferences about the behavior of data.

Hypothesis: Women who have their first child after the age of 30 are more likely to develop breast cancer than those who have their first child before age 20.

Sample size: 2000 women; 45-54 years of age; of which, 1000 had their first child before age 20 and 1000 after age 30. This is a limited sample size and results may not be conclusive.

The sample size may be increased to 10,000 women; however, the apparent difference in rate of breast cancer occurrence may still be due to chance.

To set a framework for evaluating occurrence, we introduce the concept of **probability**.

3.2 Definition of Probability

Probability of a male livebirth during the period 1965–1974

Time period	Number of male livebirths (a)	Total number of livebirths (b)	Empirical probability of a male livebirth (a/b)
1965	1,927,054	3,760,358	.51247
1965–1969	9,219,202	17,989,361	.51248
1965–1974	17,857,857	34,832,051	.51268

The probability of live male births in 1965 was 0.51247 ; in 1965 – 69 was 0.51248; *in* 1965 – 74 *was* 0.51268.

These **empirical probabilities** are based on finite amount of data.

The sample size could be expanded indefinitely and an increasingly more precise estimate of the probability obtained.

- The **sample space** is the set of all possible outcomes.
- An **event** is any set of outcomes of interest.
- The **probability** of an event is the relative frequency of this set of outcomes over an indefinitely large (or infinite) number of trials.
- In real life, experiments cannot be performed an infinite number of times. Hence, probabilities are estimated from empirical probabilities obtained from larger samples.
- Theoretical probability models may also be constructed from which probabilities of many different kinds of events can be computed.
- Comparing empirical probabilities with theoretical probabilities enables us to assess the **goodness-of-fit** of probability models.

- The probability of an event *E*, denoted by *Pr(E)*, always satisfies

  $0 \leq Pr(E) \leq 1$

- If outcomes *A* and *B* are two events that cannot both happen at the same time, then *Pr(A or B occurs) = Pr(A) + Pr(B)*

EXAMPLE 3.6

Hypertension Let A be the event that a person has normotensive diastolic blood-pressure (DBP) readings (DBP < 90), and let B be the event that a person has borderline DBP readings ($90 \leq$ DBP < 95). Suppose that $Pr(A) = .7$, and $Pr(B) = .1$. Let Z be the event that a person has a DBP < 95. Then

$$Pr(Z) = Pr(A) + Pr(B) = .8$$

because the events A and B cannot occur at the same time.

Two events A and B are **mutually exclusive** if they cannot both happen at the same time.

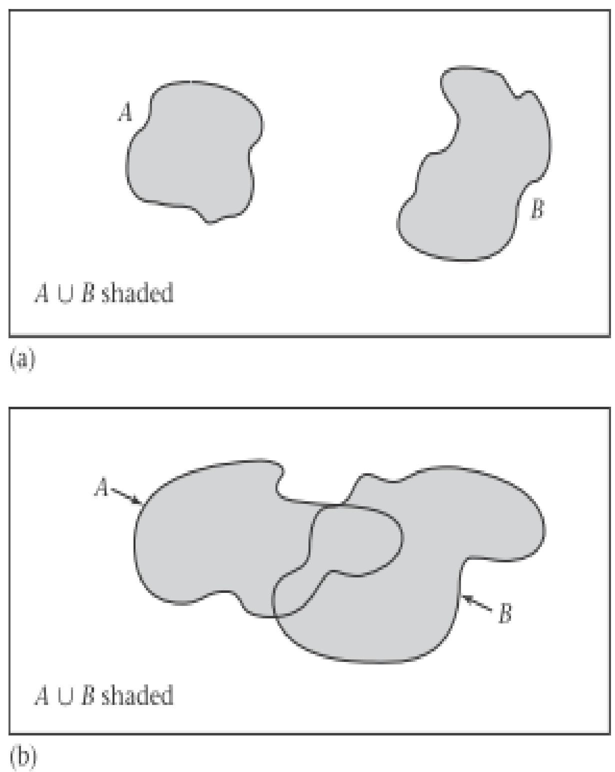
EXAMPLE 3.7

Hypertension Let X be DBP, C be the event $X \geq 90$, and D be the event $75 \leq X \leq 100$. Events C and D are *not* mutually exclusive, because they both occur when $90 \leq X \leq 100$.

3.3 Some Useful Probabilistic Notations

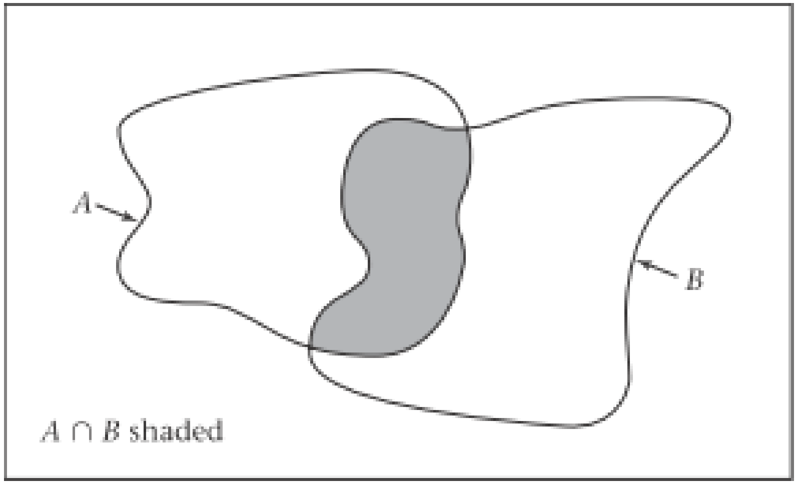
- The symbol { } is used as shorthand for the phrase “the event.”
- A ∪ B is the event that either A or B occurs, or they both occur.

Figure 3.1 Diagrammatic representation of $A \cup B$: (a) A, B mutually exclusive; (b) A, B not mutually exclusive



A $\cap$ B is the event that both A and B occur simultaneously.

Figure 3.2 Diagrammatic representation of $A \cap B$



EXAMPLE 3.8 **Hypertension** Let events A and B be defined as in Example 3.6: $A = \{X < 90\}$, $B = \{90 \le X < 95\}$, where X = DBP. Then $A \cup B = \{X < 95\}$.

EXAMPLE 3.9 **Hypertension** Let events C and D be defined as in Example 3.7:

$$C = \{X \ge 90\} \quad D = \{75 \le X \le 100\}$$

Then $C \cup D = \{X \ge 75\}$

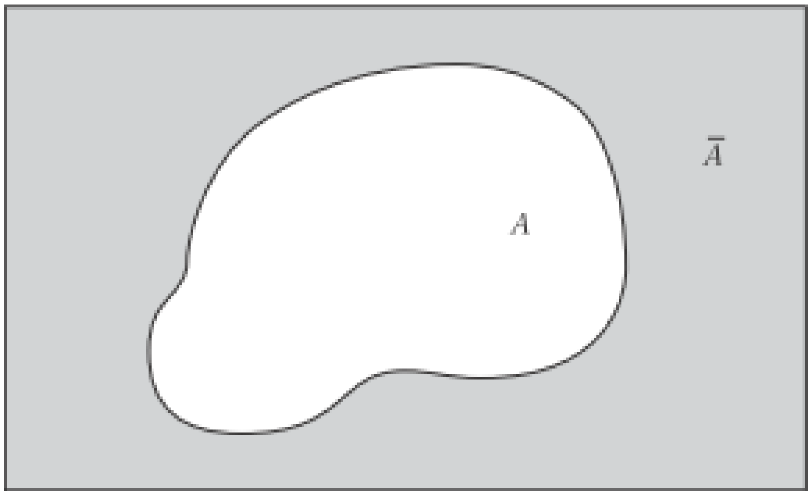
EXAMPLE 3.10 **Hypertension** Let events C and D be defined as in Example 3.7; that is,

$$C = \{X \ge 90\} \quad D = \{75 \le X \le 100\}$$

Then $C \cap D = \{90 \le X \le 100\}$

$\overline{A}$ is the event that A does not occur. It is called the complement of A. Notice that $Pr(\overline{A}) = 1 - Pr(A)$, because $\overline{A}$ occurs only when A does not occur.

Figure 3.3 Diagrammatic representation of $\overline{A}$



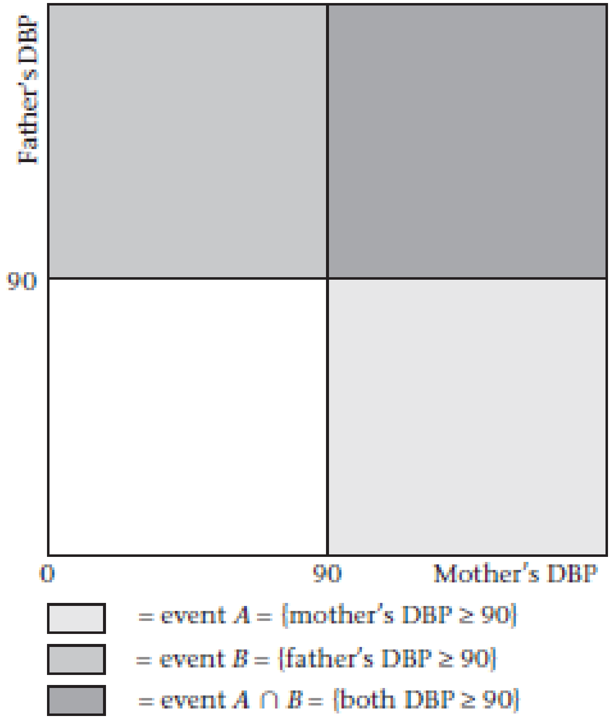
Example:

If A and C are two events. A = { X < 90}; C = { X ≥ 90}.

Then, C = $\overline{A}$, because C can only occur when A does not occur.

3.4 Multiplication Law of Probability

FIGURE 3.4 Possible diastolic blood-pressure measurements of the mother and father within a given family



Events A and B are called independent events if
$Pr(A \cap B) = Pr(A) \times Pr(B)$

If $Pr(A) = 0.1$ and $Pr(B) = 0.2$
Probability that both mother and father are hypertensive,

$Pr(A \cap B) = Pr(A) \times Pr(B)$
$= 0.1(0.2) = .02$

If $A_1, \ldots, A_k$ are mutually independent events, then
$Pr(A_1 \cap A_2 \cap \ldots A_k) =$
$Pr(A_1) \times Pr(A_2) \ldots \times Pr(A_k)$

EXAMPLE 3.12

Hypertension, Genetics Suppose we are conducting a hypertension-screening program in the home. Consider all possible pairs of DBP measurements of the mother and father within a given family, assuming that the mother and father are not genetically related. This sample space consists of all pairs of numbers of the form (X, Y) where $X > 0$, $Y > 0$. Certain specific events might be of interest in this context. In particular, we might be interested in whether the mother or father is hypertensive, which is described, respectively, by events $A = \{\text{mother's DBP} \ge 90\}$, $B = \{\text{father's DBP} \ge 90\}$. These events are diagrammed in Figure 3.4.

Suppose we know that $Pr(A) = .1$, $Pr(B) = .2$. What can we say about $Pr(A \cap B) = Pr(\text{mother's DBP} \ge 90 \text{ and father's DBP} \ge 90) = Pr(\text{both mother and father are hypertensive})$? We can say nothing unless we are willing to make certain assumptions.

Two events A, B are dependent if $Pr(A \cap B) \neq Pr(A) \times Pr(B)$

EXAMPLE 3.14

Hypertension, Genetics Consider all possible DBP measurements from a mother and her first-born child. Let

$$A = \{\text{mother's DBP} \ge 90\} \quad B = \{\text{first-born child's DBP} \ge 80\}$$

Suppose $\quad Pr(A \cap B) = .05 \quad Pr(A) = .1 \quad Pr(B) = .2$

Then $\quad Pr(A \cap B) = .05 > Pr(A) \times Pr(B) = .02$

and the events A, B would be dependent.

EXAMPLE 3.15

Sexually Transmitted Disease Suppose two doctors, A and B, test all patients coming into a clinic for syphilis. Let events $A^+ = \{\text{doctor A makes a positive diagnosis}\}$ and $B^+ = \{\text{doctor B makes a positive diagnosis}\}$. Suppose doctor A diagnoses 10% of all patients as positive, doctor B diagnoses 17% of all patients as positive, and both doctors diagnose 8% of all patients as positive. Are the events A^+, B^+ independent?

Solution: We are given that

$$Pr(A^+) = .1 \quad Pr(B^+) = .17 \quad Pr(A^+ \cap B^+) = .08$$

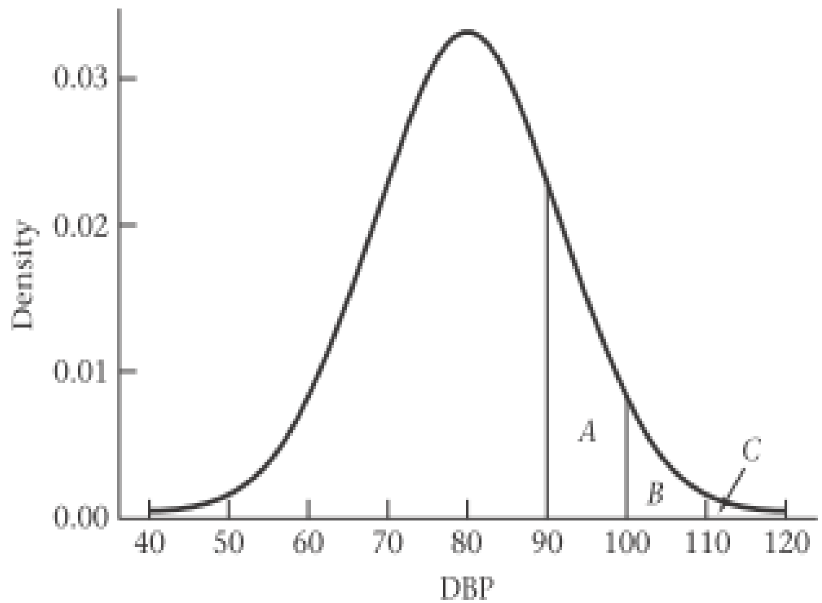
Thus, $Pr(A^+ \cap B^+) = .08 > Pr(A^+) \times Pr(B^+) = .1(.17) = .017$

and the events are dependent. This result would be expected because there should be a similarity between how two doctors diagnose patients for syphilis.

The probability-density function of the random variable X is a function such that the area under the density-function curve between any two points a and b is equal to the probability that the random variable X falls between, a and b.

Thus, the total area under the density-function curve over the entire range of possible values for random variable is 1.

Figure 5.1 The pdf of DBP in 35- to 44-year-old men

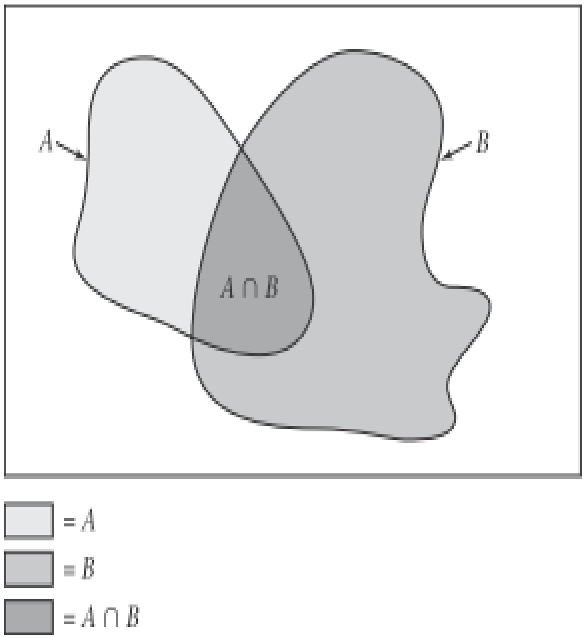


Areas A, B, C ≅ probabilities of being mildly, moderately, and severely hypertensive.

3.5 Addition Law of Probability

If A and B are any events, then $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

Figure 3.5 Diagrammatic representation of the addition law of probability



Special case: If events A and B are mutually exclusive, then $\Pr(A \cap B) = 0$
And the addition law reduces to
$\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

EXAMPLE 3.16

Sexually Transmitted Disease Consider the data in Example 3.15. Suppose a patient is referred for further lab tests if either doctor A or B makes a positive diagnosis. What is the probability that a patient will be referred for further lab tests?

Solution: The event that either doctor makes a positive diagnosis can be represented by $A^+ \cup B^+$. We know that

$$Pr(A^+) = .1 \qquad Pr(B^+) = .17 \qquad Pr(A^+ \cap B^+) = .08$$

Therefore, from the addition law of probability,

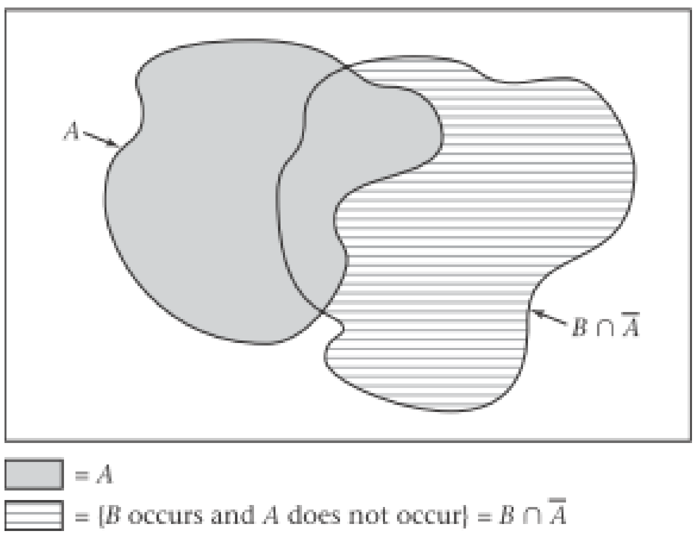
$$Pr(A^+ \cup B^+) = Pr(A^+) + Pr(B^+) - Pr(A^+ \cap B^+) = .1 + .17 - .08 = .19$$

Thus, 19% of all patients will be referred for further lab tests.

For independent events A and B: $\Pr(A \cup B) = \Pr(A) + \Pr(B) \times [1 - \Pr(A)]$

Two mutually exclusive events: {A occurs} and {B occurs and A does not occur}.

Figure 3.6 Diagrammatic representation of the addition law of probability for independent events



Similarly, for three events A, B, and C

$\Pr(A \cup B \cup C)$

$= \Pr(A) + \Pr(B) + \Pr(C) - \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C) + \Pr(A \cap B \cap C)$

EXAMPLE 3.17

Hypertension Look at Example 3.12, where

$$A = \{\text{mother's DBP} \geq 90\} \quad \text{and} \quad B = \{\text{father's DBP} \geq 90\}$$

$Pr(A) = .1$, $Pr(B) = .2$, and assume A and B are independent events. Suppose a "hypertensive household" is defined as one in which either the mother or the father is hypertensive, with hypertension defined for the mother and father, respectively, in terms of events A and B. What is the probability of a hypertensive household?

Solution: Pr(hypertensive household) is

$$Pr(A \cup B) = Pr(A) + Pr(B) \times [1 - Pr(A)] = .1 + .2(.9) = .28$$

Thus, 28% of all households will be hypertensive.

3.6 Conditional Probability

The quantity $Pr(A \cap B)/Pr(A)$ is defined as the **conditional probability of B given A**, which is written $Pr(B|A)$.

- If A and B are independent events, then $Pr(B|A) = Pr(B) = Pr(B|\bar{A})$
- If two events A, B are dependent, then $Pr(B|A) \neq Pr(B) \neq Pr(B|\bar{A})$ and $Pr(A \cap B) \neq Pr(A) \times Pr(B)$

Def.: The **relative risk (RR)** of B given A is $Pr(B|A)/Pr(B|\bar{A})$

- If A and B are independent, then RR is 1.
- If A and B are dependent, then RR is different from 1.
- The more the dependence between events increases, the further the RR will be from 1.

EXAMPLE 3.19

Cancer Suppose that among 100,000 women with negative mammograms 20 will be diagnosed with breast cancer within 2 years, or $Pr(B|\bar{A}) = 20/10^5 = .0002$, whereas 1 woman in 10 with positive mammograms will be diagnosed with breast cancer within 2 years, or $Pr(B|A) = .1$. The two events A and B would be highly dependent, because

$$RR = Pr\left(B|A\right)/Pr\left(B|\bar{A}\right) = .1/.0002 = 500$$

In other words, women with positive mammograms are 500 times more likely to develop breast cancer over the next 2 years than are women with negative mammograms. This is the rationale for using the mammogram as a screening test for breast cancer. If events A and B were independent, then the RR would be 1; women with positive or negative mammograms would be equally likely to have breast cancer, and the mammogram would not be useful as a screening test for breast cancer.

EXAMPLE 3.20

Sexually Transmitted Disease Using the data in Example 3.15, find the conditional probability that doctor B makes a positive diagnosis of syphilis given that doctor A makes a positive diagnosis. What is the conditional probability that doctor B makes a positive diagnosis of syphilis given that doctor A makes a negative diagnosis? What is the *RR* of B^+ given A^+?

Solution: $Pr(B^+|A^+) = Pr(B^+ \cap A^+)/Pr(A^+) = .08/.1 = .8$

Thus, doctor B will confirm doctor A's positive diagnoses 80% of the time. Similarly,

$$Pr(B^+|A^-) = Pr(B^+ \cap A^-)/Pr(A^-) = Pr(B^+ \cap A^-)/.9$$

We must compute $Pr(B^+ \cap A^-)$. We know that if doctor B diagnoses a patient as positive, then doctor A either does or does not confirm the diagnosis. Thus,

$$Pr(B^+) = Pr(B^+ \cap A^+) + Pr(B^+ \cap A^-)$$

because the events $B^+ \cap A^+$ and $B^+ \cap A^-$ are mutually exclusive. If we subtract $Pr(B^+ \cap A^+)$ from both sides of the equation, then

$$Pr(B^+ \cap A^-) = Pr(B^+) - Pr(B^+ \cap A^+) = .17 - .08 = .09$$

Therefore, $Pr(B^+ | A^-) = .09/.9 = .1$

Thus, when doctor A diagnoses a patient as negative, doctor B will contradict the diagnosis 10% of the time. The *RR* of the event B^+ given A^+ is

$$Pr(B^+|A^+)/Pr(B^+|A^-) = .8/.1 = 8$$

This indicates that doctor B is 8 times as likely to diagnose a patient as positive when doctor A diagnoses the patient as positive than when doctor A diagnoses the patient as negative. These results quantify the dependence between the two doctors' diagnoses.

Total-Probability Rule

For any events A and B, $\Pr(B) = \Pr(B|A) \times \Pr(A) + \Pr(B|\bar{A}) \times \Pr(\bar{A})$

If event B occurs, it must occur either with A or without A.

$\Pr(B) = \Pr(B \cap A) + \Pr(B \cap \bar{A})$

From definition of conditional probability,

$\Pr(B \cap A) = \Pr(A) \times \Pr(B|A)$ and $\Pr(B \cap \bar{A}) = \Pr(\bar{A}) \times \Pr(B|\bar{A})$

On substitution,

$\Pr(B) = \Pr(B|A)\Pr(A) + \Pr(B|\bar{A})\Pr(\bar{A})$

The unconditional probability of B is the sum of the conditional probability of B given A times the unconditional probability of A plus the conditional probability of B given A not occurring times the unconditional probability of A not occurring.

EXAMPLE 3.21 **Cancer** Let A and B be defined as in Example 3.19, and suppose that 7% of the general population of women will have a positive mammogram. What is the probability of developing breast cancer over the next 2 years among women in the general population?

Solution:

$$\begin{aligned} Pr(B) &= Pr(\text{breast cancer}) \\ &= Pr(\text{breast cancer} \mid \text{mammogram}^+) \times Pr(\text{mammogram}^+) \\ &+ Pr(\text{breast cancer} \mid \text{mammogram}^-) \times Pr(\text{mammogram}^-) \\ &= .1(.07) + .0002(.93) = .00719 = 719/10^5 \end{aligned}$$

Thus, the unconditional probability of developing breast cancer over the next 2 years in the general population ($719/10^5$) is a weighted average of the conditional probability of developing breast cancer over the next 2 years among women with a positive mammogram (.1) and the conditional probability of developing breast cancer over the next 2 years among women with a negative mammogram ($20/10^5$), with weights of 0.07 and 0.93 corresponding to mammogram$^+$ and mammogram$^-$ women, respectively.

EXAMPLE 3.22 **Ophthalmology** We are planning a 5-year study of cataract in a population of 5000 people 60 years of age and older. We know from census data that 45% of this population is 60–64 years of age, 28% are 65–69 years of age, 20% are 70–74 years of age, and 7% are 75 or older. We also know from the Framingham Eye Study that 2.4%, 4.6%, 8.8%, and 15.3% of the people in these respective age groups will develop cataract over the next 5 years [4]. What percentage of the population in our study will develop cataract over the next 5 years, and how many people with cataract does this percentage represent?

Solution: Let A_1 = {ages 60–64}, A_2 = {ages 65–69}, A_3 = {ages 70–74}, A_4 = {ages 75+}. These events are mutually exclusive and exhaustive because each person in our population must be in one and only one age group. Furthermore, from the conditions of the problem we know that $Pr(A_1) = .45$, $Pr(A_2) = .28$, $Pr(A_3) = .20$, $Pr(A_4) = .07$, $Pr(B|A_1) = .024$, $Pr(B|A_2) = .046$, $Pr(B|A_3) = .088$, and $Pr(B|A_4) = .153$, where B = {develop cataract in the next 5 years}. Finally, using the total-probability rule,

$$\begin{aligned} Pr(B) &= Pr(B|A_1) \times Pr(A_1) + Pr(B|A_2) \times Pr(A_2) \\ &+ Pr(B|A_3) \times Pr(A_3) + Pr(B|A_4) \times Pr(A_4) \\ &= .024(.45) + .046(.28) + .088(.20) + .153(.07) = .052 \end{aligned}$$

Thus 5.2% of this population will develop cataract over the next 5 years, which represents a total of $5000 \times .052 = 260$ people with cataract.

The probability of B is expressed in terms of two mutually exclusive events A and $\bar{A}$. In many instances, the probability of B will need to be expressed in terms of more than two mutually exclusive events, A_1, $A_2,\ldots,A_k$

A set of events $A_1,\ldots, A_k$ is exhaustive if at least one of the events must occur.

Let $A_1,\ldots A_k$ be mutually exclusive and exhaustive events. The unconditional probability of B (Pr(B)) can then be written as a weighted average of the conditional probabilities of B given A_i ($Pr(B|A_i)$) as follows: $Pr(B) = \sum^{k}_{i=1} Pr(B|A_i) \times Pr(A_i)$

Generalized Multiplication Law of Probability

If $A_1,\ldots, A_k$ are an arbitrary set of events, then $Pr(A_1 \cap A_2 \cap \ldots A_k)$

$= Pr(A_1) \times Pr(A_2|A_1) \times Pr(A_3|A_2 \cap A_1) \times \ldots . Pr(A_k|A_{k-1} \ldots . \cap A_2 \cap A_1)$

3.7 Baye's Rule and Screening Tests

The predictive value positive (PV^+) of a screening test is the probability that a person has a disease given that the test is positive.

$$Pr(\text{disease}|\text{test}^+)$$

The predictive value negative (PV^-) of a screening test is the probability that a person does not have a disease given that the test is negative.

$$Pr(\text{no disease}|\text{test}^-)$$

EXAMPLE 3.23

Cancer Find PV^+ and PV^- for mammography given the data in Example 3.19.

Solution: We see that $PV^+ = Pr(\text{breast cancer} \mid \text{mammogram}^+) = .1$

whereas $PV^- = Pr(\text{breast cancer}^- \mid \text{mammogram}^-)$

$$= 1 - Pr(\text{breast cancer} \mid \text{mammogram}^-) = 1 - .0002 = .9998$$

Thus, if the mammogram is negative, the woman is virtually certain *not* to develop breast cancer over the next 2 years ($PV^- \approx 1$); whereas if the mammogram is positive, the woman has a 10% chance of developing breast cancer ($PV^+ = .10$).

- A symptom or set of symptoms can be regarded as a screening test for a disease. Higher the PV of the symptoms, the more valuable the test will be.
- Clinicians often cannot directly measure the PV of a set of symptoms. However, they can measure how often specific symptoms occur in diseased and normal people.
- **Sensitivity** of a symptom (or a set of symptoms): the probability that the symptom is present given that the person has a disease.
- **Specificity** of a symptom (or a set of symptoms): the probability that the symptom is not present given that the person does not have a disease.
- **False negative**: negative test result when the disease or condition being tested for is actually present.
- **False positive**: positive test result when the disease or condition being tested for is not actually present.

EXAMPLE 3.24

Cancer Suppose the disease is lung cancer and the symptom is cigarette smoking. If we assume that 90% of people with lung cancer and 30% of people without lung cancer (essentially the entire general population) are smokers, then the sensitivity and specificity of smoking as a screening test for lung cancer are .9 and .7, respectively. Obviously, cigarette smoking cannot be used by itself as a screening criterion for predicting lung cancer because there will be too many false positives (people without cancer who are smokers).

EXAMPLE 3.25

Cancer Suppose the disease is breast cancer in women and the symptom is having a family history of breast cancer (either a mother or a sister with breast cancer). If we assume 5% of women with breast cancer have a family history of breast cancer but only 2% of women without breast cancer have such a history, then the sensitivity of a family history of breast cancer as a predictor of breast cancer is .05 and the specificity is .98 = (1 − .02). A family history of breast cancer cannot be used by itself to diagnose breast cancer because there will be too many false negatives (women with breast cancer who do not have a family history).

$$\text{Predictive value positive} = PV^+ = Pr\left(B \middle| A\right)$$

$$\text{Predictive value negative} = PV^- = Pr\left(\bar{B} \middle| \bar{A}\right)$$

$$\text{Sensitivity} = Pr\left(A \middle| B\right)$$

$$\text{Specificity} = Pr\left(\bar{A} \middle| \bar{B}\right)$$

Baye's Rule

Let A = symptom and B = disease.

$$PV^+ = Pr(B|A) = \frac{Pr(A|B)\times Pr(B)}{Pr(A|B)\times Pr(B) + Pr(A|\bar{B})\times Pr(\bar{B})}$$

In words, this can be written as

$$PV^+ = \frac{\text{Sensitivity}\times x}{\text{Sensitivity}\times x + (1-\text{Specificity})\times(1-x)}$$

where $x = Pr(B)$ = prevalence of disease in the reference population. Similarly,

$$PV^- = \frac{\text{Specificity}\times(1-x)}{\text{Specificity}\times(1-x) + (1-\text{Sensitivity})\times x}$$

Generalized Baye's Rule

Let B_1, B_2,..., B_k be a set of mutually exclusive and exhaustive disease states. Let A be a symptom or a set of symptoms.

$$Pr(B_i|A) = Pr(A|B_i)\times Pr(B_i) \Big/ \left[\sum_{j=1}^{k} Pr(A|B_j)\times Pr(B_j)\right]$$

EXAMPLE 3.26

Hypertension Suppose 84% of hypertensives and 23% of normotensives are classified as hypertensive by an automated blood-pressure machine. What are the PV^+ and PV^- of the machine, assuming 20% of the adult population is hypertensive?

Solution: The sensitivity = .84 and specificity = 1 – .23 = .77. Thus, from Bayes' rule it follows that

$$PV^+ = (.84)(.2)/[(.84)(.2)+(.23)(.8)]$$
$$= .168/.352 = .48$$

Similarly, $PV^- = (.77)(.8)/[(.77)(.8)+(.16)(.2)]$
$$= .616/.648 = .95$$

Thus, a negative result from the machine is reasonably predictive because we are 95% sure a person with a negative result from the machine is normotensive. However, a positive result is not very predictive because we are only 48% sure a person with a positive result from the machine is hypertensive.

Example 3.26 considered only two possible disease states: hypertensive and normotensive. In clinical medicine there are often more than two possible disease states. We would like to be able to predict the most likely disease state given a specific symptom (or set of symptoms). Let's assume that the probability of having these symptoms among people in each disease state (where one of the disease states may be normal) is known from clinical experience, as is the probability of each disease state in the reference population. This leads us to the generalized Bayes' rule:

EXAMPLE 3.27 **Pulmonary Disease** Suppose a 60-year-old man who has never smoked cigarettes presents to a physician with symptoms of a chronic cough and occasional breathlessness. The physician becomes concerned and orders the patient admitted to the hospital for a lung biopsy. Suppose the results of the lung biopsy are consistent either with lung cancer or with sarcoidosis, a fairly common, usually nonfatal lung disease. In this case

$$A = \{\text{chronic cough, results of lung biopsy}\}$$

$$\text{Disease state} \begin{cases} B_1 = \text{normal} \\ B_2 = \text{lung cancer} \\ B_3 = \text{sarcoidosis} \end{cases}$$

Suppose that $Pr(A|B_1) = .001$ $Pr(A|B_2) = .9$ $Pr(A|B_3) = .9$

and that in 60-year-old, never-smoking men

$$Pr(B_1) = .99 \quad Pr(B_2) = .001 \quad Pr(B_3) = .009$$

The first set of probabilities $Pr(A|B_i)$ could be obtained from clinical experience with the previous diseases, whereas the latter set of probabilities $Pr(B_i)$ would have to be obtained from age-, gender-, and smoking-specific prevalence rates for the diseases in question. The interesting question now becomes what are the probabilities $Pr(B_i|A)$ of the three disease states given the previous symptoms?

Solution: Bayes' rule can be used to answer this question. Specifically,

$$Pr\left(B_1|A\right) = Pr\left(A|B_1\right) \times Pr\left(B_1\right) \Big/ \left[\sum_{j=1}^{3} Pr\left(A|B_j\right) \times Pr\left(B_j\right)\right]$$

$$= .001(.99)/[.001(.99) + .9(.001) + .9(.009)]$$

$$= .00099 / .00999 = .099$$

$$Pr\left(B_2|A\right) = .9(.001)/[.001(.99) + .9(.001) + .9(.009)]$$

$$= .00090 / .00999 = .090$$

$$Pr\left(B_3|A\right) = .9(.009)/[.001(.99) + .9(.001) + .9(.009)]$$

$$= .00810 / .00999 = .811$$

Thus, although the unconditional probability of sarcoidosis is very low (.009), the conditional probability of the disease given these symptoms and this age-gender-smoking group is .811. Also, although the symptoms and diagnostic tests are consistent with both lung cancer and sarcoidosis, the latter is much more likely among patients in this age-gender-smoking group (i.e., among never- smoking men).

EXAMPLE 3.28 **Pulmonary Disease** Now suppose the patient in Example 3.27 smoked two packs of cigarettes per day for 40 years. Then assume $Pr(B_1) = .98$, $Pr(B_2) = .015$, and $Pr(B_3) = .005$ in this type of person. What are the probabilities of the three disease states for this type of patient, given these symptoms?

Solution: $Pr\left(B_1|A\right) = .001(.98)/[.001(.98) + .9(.015) + .9(.005)]$

$$= .00098 / .01898 = .052$$

$$Pr\left(B_2|A\right) = .9(.015) / .01898 = .01350 / .01898 = .711$$

$$Pr\left(B_3|A\right) = .9(.005) / .01898 = .237$$

Thus, in this type of patient (i.e., a heavy-smoking man) lung cancer is the most likely diagnosis.

Summary

- Probabilities may be calculated using addition and multiplication laws.
- Independent events are unrelated to each other as opposed to dependent events.
- Conditional probability and RR may be used quantify the dependence between two events.
- Sensitivity, specificity, and PV are used to define the accuracy of screening tests.
- Baye's rule may be used to compute the PV of screening tests.

The End