BME 3005
Biostatistics

Lecture 3: *Probability*

Asst. Prof. Burcu Tunç Çamlıbel

Chapter 03

Probability

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**.

Table 3.1 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:

Let *A* = person has normotensive diastolic blood pressure (DBP)

B = person has borderline DBP readings (***90* ≤ DBP < 95**).

Pr(A) = 0.7 and Pr(B) = 0.1.

Z = event that person has DBP < 95.

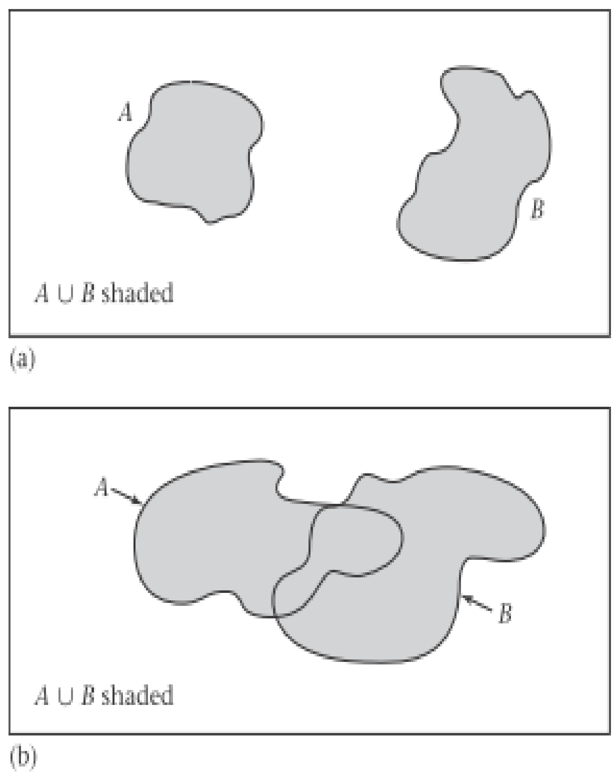
Then, Pr(Z) = Pr(A) + Pr(B) = 0.8.

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

Some Useful Probabilistic Notations

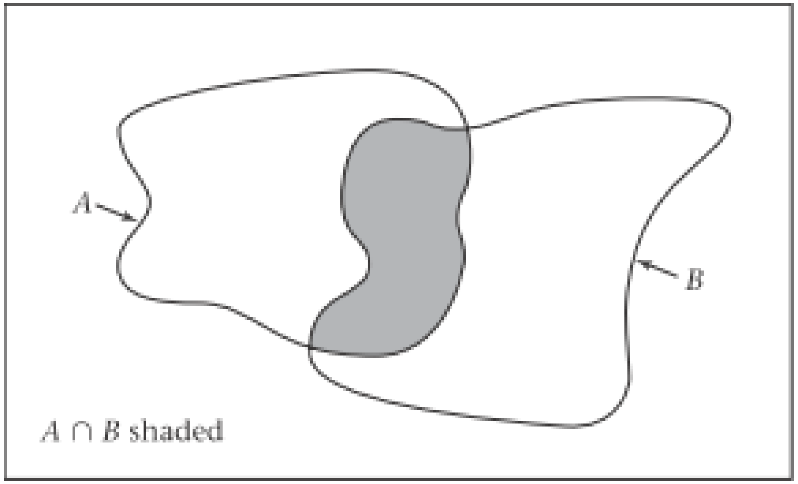
- The symbol { } is used as shorthand for the phrase "the event."
- A ∪ B is the event that either A or B occurs, or they both occur.

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



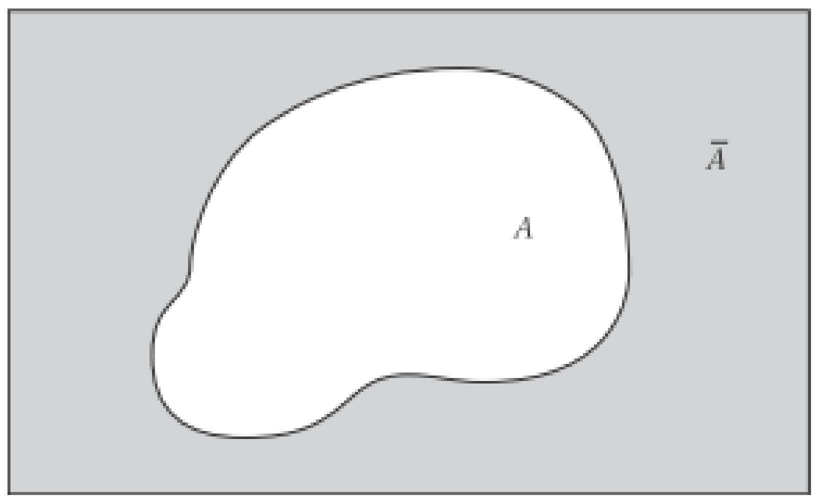
A $\cap$ B is the event that both A and B occur simultaneously.

Figure 3.2 Diagrammatic representation of $A \cap B$



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

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

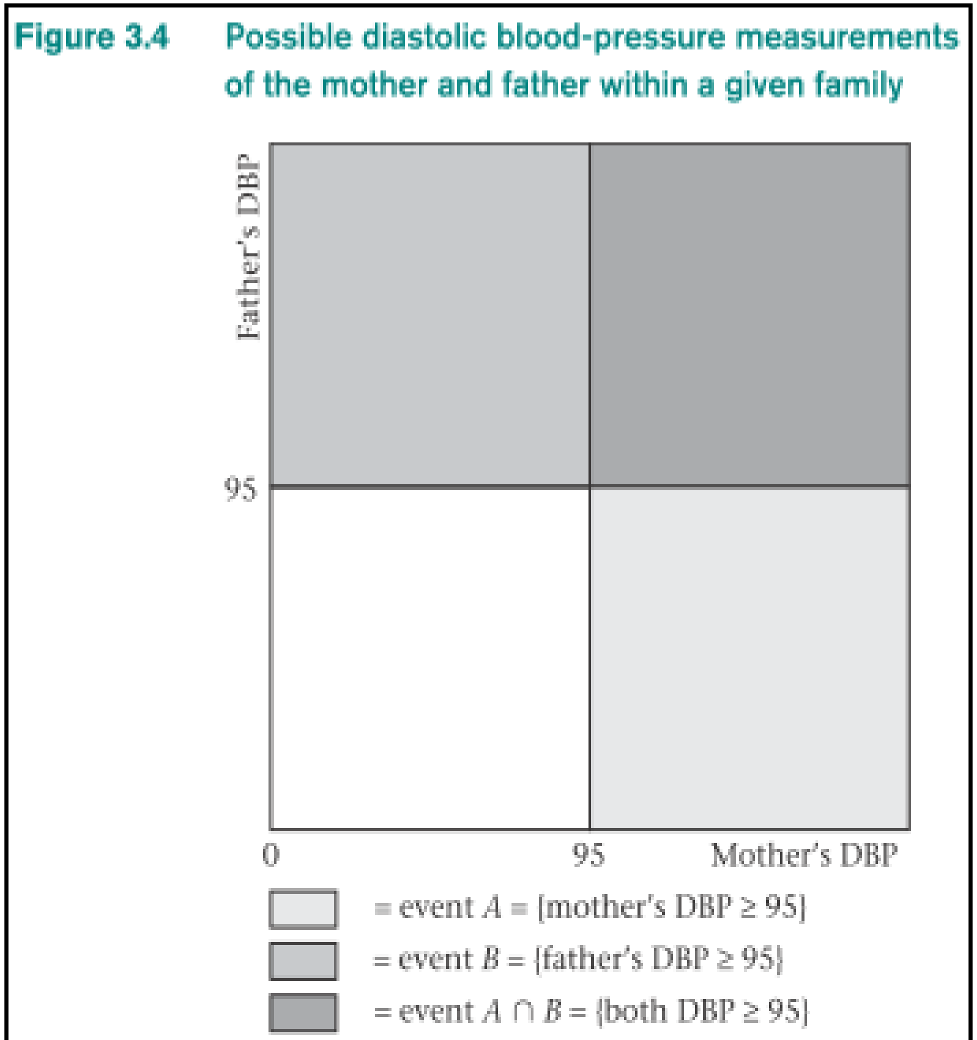


Example:

If A and C are two events. $A = \{ X < 90\}$; $C = \{ X \geq 90\}$.

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

Multiplication Law of Probability



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
$Pr(A_1 \cap A_2 \cap \ldots A_k) =$
$Pr(A_1) \times Pr(A_2) \ldots \times Pr(A_k)$

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

The probability of exactly obtaining any value is 0.

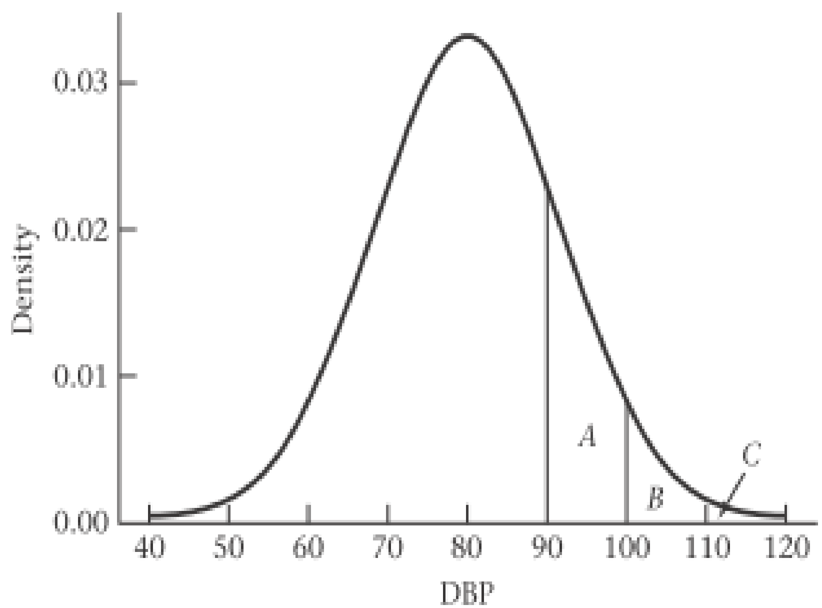
However, there exists the notion that certain ranges of values occur more frequently than others.

This notion can be quantified using the concept of a probability-density function (pdf).

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



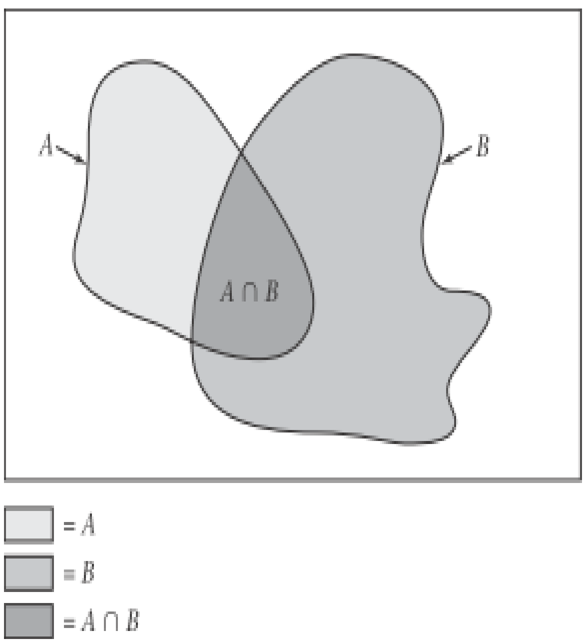
Areas A, B, C ≅ probabilities of being mildly, moderately, and severely hypertensive.

The pdf has large values in regions of high probability and small values in regions of low probability A.

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

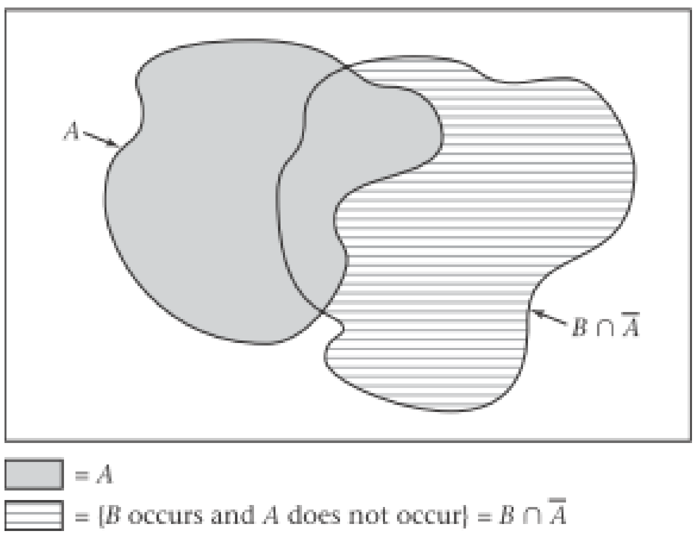


Special case: If events A and B are mutually exclusive, then $\Pr(A \cap B) = 0$
And the addition law reduces to
$\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

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 $\overline{A}$ 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 B: $Pr(A \cup B \cup C)$
$= Pr(A) + Pr(B) + Pr(C) - Pr(A \cap B) - Pr(A \cap C) - Pr(B \cap C) + Pr(A \cap B \cap C)$

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

The **relative risk (RR)** of B given A is $Pr(B|A)/Pr(B|\overline{A})$

- 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.

Total-Probability Rule

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

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

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

From definition of conditional probability,

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

On substitution,

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

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.

The probability of B is expressed in terms of two mutually exclusive events A and $\overline{A}$. 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 $Pr(A_1 \cap A_2 \cap \ldots A_k)$

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

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.

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

- 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 system 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.

For a symptom to be effective in predicting a disease, the specificity and sensitivity must be high.

Table 3.2 Association between PSA and prostate cancer

PSA test result	Prostate cancer	Frequency
+	+	92
+	−	27
−	+	46
−	−	72

The level of prostrate-specific antigen (PSA) in the blood is frequently used as a screening test for prostrate cancer.

Baye's Rule

Let A = symptom and B = disease.

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

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.

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

Bayesian Inference

- It is an alternative definition of probability and inference.
- It rejects the idea of the definition of probability sometimes called the **frequency definition of probability** (a theoretical concept).
- Bayesians conceive of two types of probability:
- **Prior probability** of an event is the best guess by the observer of an event's likelihood in the absence of data. This may be a single number, or a range of likely values, perhaps with weights attached to each possible value.
- **Posterior probability** of an event is the likelihood that an event will occur after collecting some empirical data. It is obtained by integrating information from the prior probability with additional data related to the event in question.

Receiver Operating Characteristic (ROC) Curves

A test may provide several categories of response rather than simply positive or negative results. In some cases, the results of the test are reported as a continuous variable.

In either case, the designation of a cutoff point for distinguishing a test result as positive versus negative is arbitrary.

Table 3.3 Ratings of 109 CT images by a single radiologist vs. true disease status

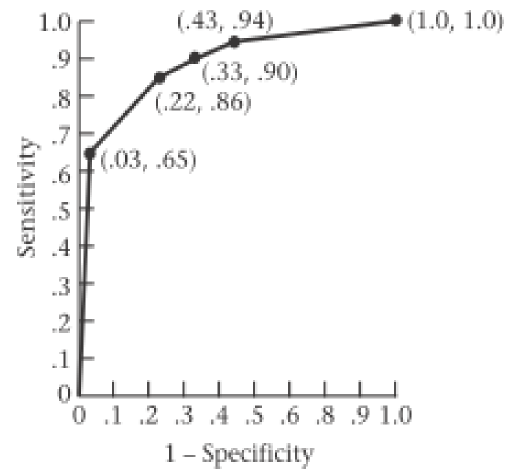
	CT rating					
True disease status	Definitely normal (1)	Probably normal (2)	Questionable (3)	Probably abnormal (4)	Definitely abnormal (5)	Total
Normal	33	6	6	11	2	58
Abnormal	3	2	2	11	33	51
Total	36	8	8	22	35	109

This example has no obvious cutoff point to use for designating a subject as positive for disease based on the CT scan.

Table 3.4 Sensitivity and specificity of the radiologist's ratings according to different test-positive criteria based on the data in Table 3.3		
Test-positive criteria	Sensitivity	Specificity
1 +	1.0	0
2 +	.94	.57
3 +	.90	.67
4 +	.86	.78
5 +	.65	.97
6 +	0	1.0

To display above data, we construct a **receiver operating characteristic (ROC) curve**, which is a plot of the sensitivity versus (1 – specificity) of a screening test, where the different points on the curve correspond to different cutoff points used to designate test-positive.

Figure 3.7 ROC curve for the data in Table 3.4*



*Each point represents (1 – specificity, sensitivity) for different test-positive criteria.

Prevalence and Incidence

- **Prevalence** of a disease is the probability of currently having the disease regardless of the duration of time one has had the disease. Prevalence is obtained by dividing the number of people who currently have the disease by the number of people in the study population.

- **Cumulative incidence**, also referred to as incidence, of a disease is the probability that a person with no prior disease will develop a new case of the disease over some specified time period.

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.
- ROC curve is used when the designation of the cutoff point for test-positive versus test-negative is arbitrary.
- Baye’s rule may be used to compute the PV of screening tests.
- In Bayesian inference, we specify a prior probability for an event, which after data are collected, is then modified to a posterior probability.