Essence of Statistics

Tran Manh Cuong

Bài giảng của DSLab

Viện nghiên cứu cao cấp về Toán (VIASM)

Vietnam Institute for Advanced Study in Mathematics

Outline

1. Random Variables and distributions
2. Joint Distribution
3. Random sampling and Data description
4. Estimation
5. Hyphothesis Testing

Random Variables and Distributions

- What is a random variable?
- Range
- Two types of random variable:
- Discrete
- Continuous
- Why is the distribution important?

Discrete Random Variables

VIASM
Vietnam Institute for
Advanced Study in Mathematics

Probability Distributions and Probability Mass Functions

Definition

For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a **probability mass function** is a function such that

(1) $f(x_i) \geq 0$

(2) $\sum_{i=1}^{n} f(x_i) = 1$

(3) $f(x_i) = P(X = x_i)$ (3-1)

Vietnam Institute for Advanced Study in Mathematics

Cumulative Distribution Functions

Definition

The **cumulative distribution function** of a discrete random variable X, denoted as $F(x)$, is

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

For a discrete random variable X, $F(x)$ satisfies the following properties.

(1) $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$

(2) $0 \leq F(x) \leq 1$

(3) If $x \leq y$, then $F(x) \leq F(y)$ (3-2)

Mean and Variance of a Discrete Random Variable

Definition

The **mean** or **expected value** of the discrete random variable X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_x x f(x) \qquad (3\text{-}3)$$

The **variance** of X, denoted as σ^2 or $V(X)$, is

$$\sigma^2 = V(X) = E(X - \mu)^2 = \sum_x (x - \mu)^2 f(x) = \sum_x x^2 f(x) - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

Vietnam Institute for Advanced Study in Mathematics

Mean and Variance of a Discrete Random Variable

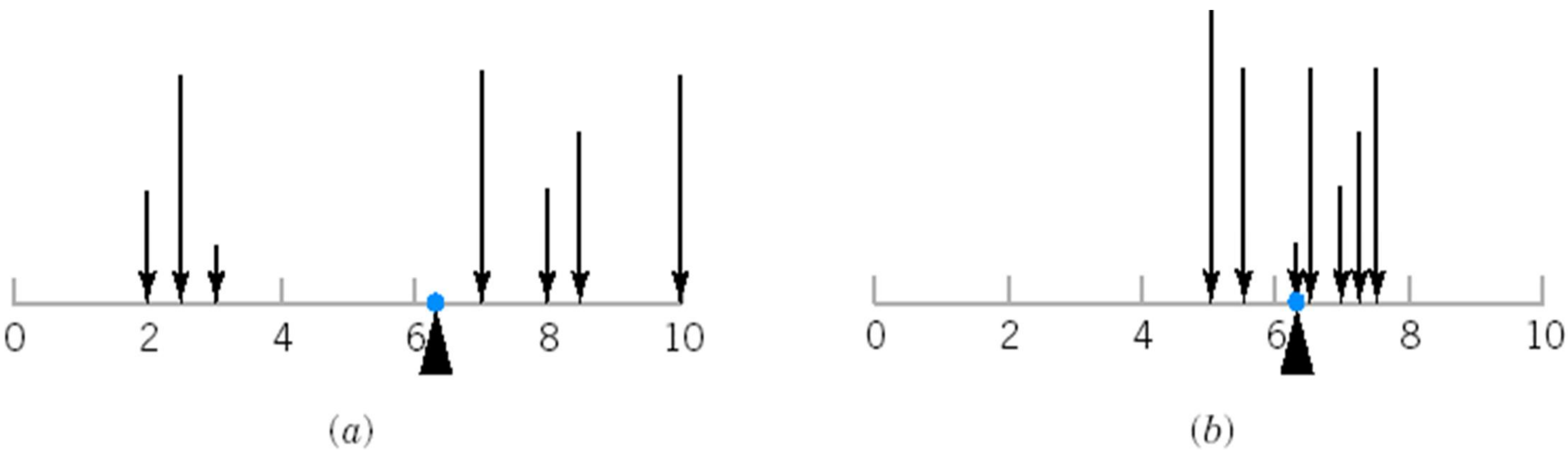


Figure A probability distribution can be viewed as a loading with the mean equal to the balance point. Parts (a) and (b) illustrate equal means, but Part (a) illustrates a larger variance.

Mean and Variance of a Discrete Random Variable

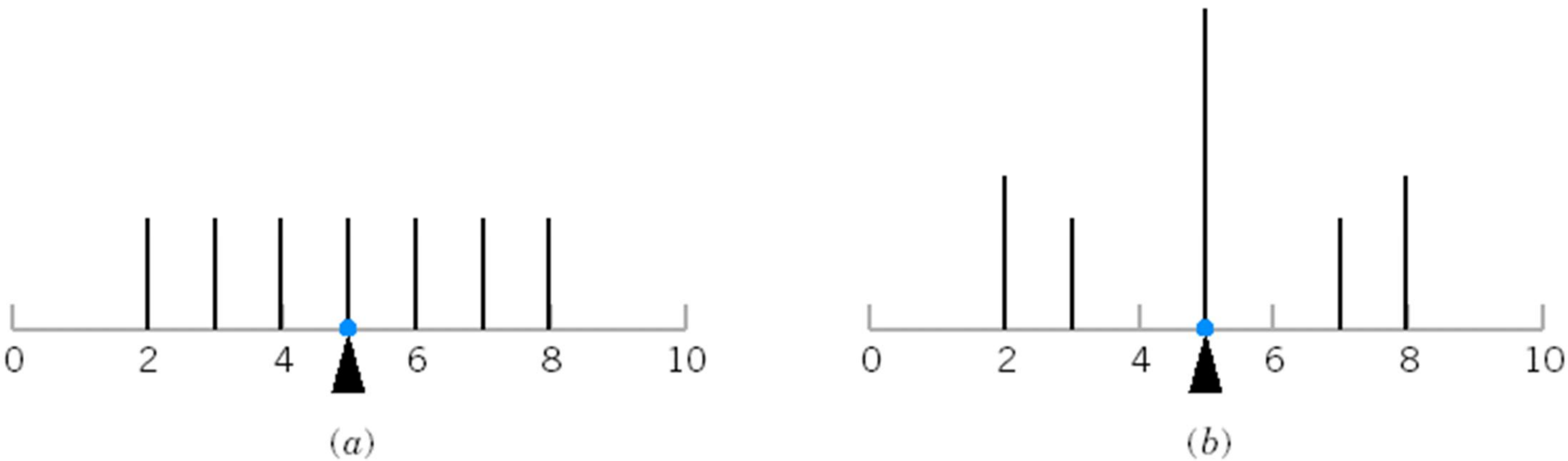


Figure The probability distribution illustrated in Parts (a) and (b) differ even though they have equal means and equal variances.

Mean and Variance of a Discrete Random Variable

Expected Value of a Function of a Discrete Random Variable

If X is a discrete random variable with probability mass function $f(x)$,

$$E[h(X)] = \sum_x h(x)f(x) \tag{3-4}$$

Discrete Uniform Distribution

Definition

A random variable X has a **discrete uniform distribution** if each of the n values in its range, say, $x_1, x_2, \ldots, x_n$, has equal probability. Then,

$$f(x_i) = 1/n \tag{3-5}$$

Discrete Uniform Distribution

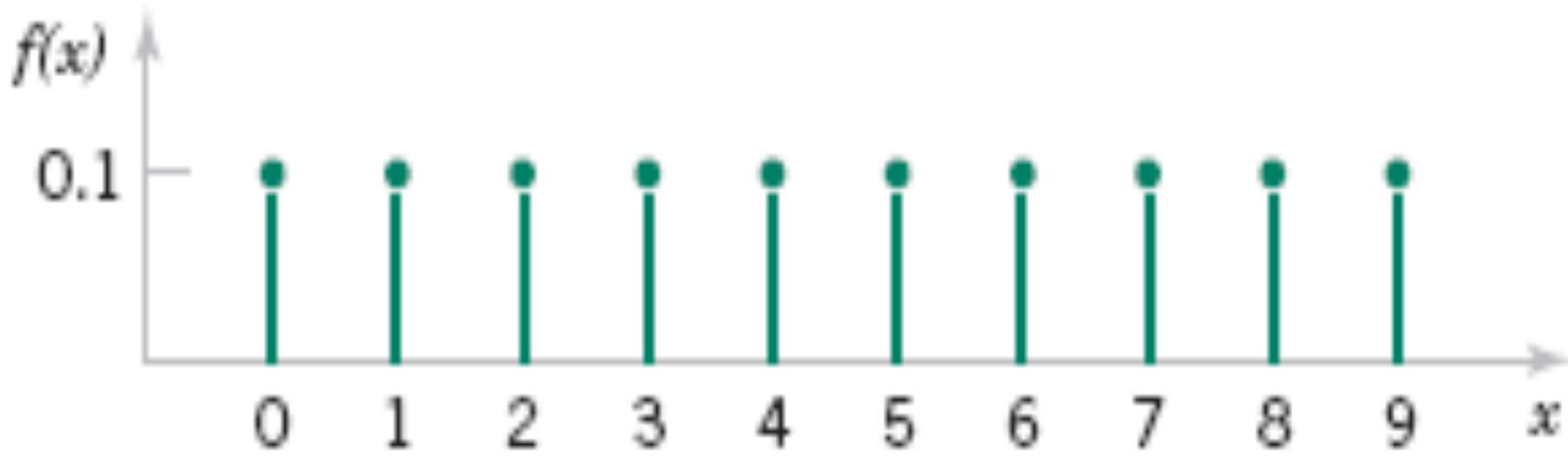


Figure Probability mass function for a discrete uniform random variable.

Vietnam Institute for Advanced Study in Mathematics

1-5 Discrete Uniform Distribution

Mean and Variance

Suppose X is a discrete uniform random variable on the consecutive integers $a, a+1, a+2, \ldots, b$, for $a \leq b$. The mean of X is

$$\mu = E(X) = \frac{b+a}{2}$$

The variance of X is

$$\sigma^2 = \frac{(b-a+1)^2 - 1}{12} \qquad (3\text{-}6)$$

Vietnam Institute for Advanced Study in Mathematics

Binomial Distribution

Random experiments and random variables

1. Flip a coin 10 times. Let X = number of heads obtained.
2. A worn machine tool produces 1% defective parts. Let X = number of defective parts in the next 25 parts produced.
3. Each sample of air has a 10% chance of containing a particular rare molecule. Let X = the number of air samples that contain the rare molecule in the next 18 samples analyzed.
4. Of all bits transmitted through a digital transmission channel, 10% are received in error. Let X = the number of bits in error in the next five bits transmitted.

Binomial Distribution

Random experiments and random variables

5. A multiple choice test contains 10 questions, each with four choices, and you guess at each question. Let X = the number of questions answered correctly.
6. In the next 20 births at a hospital, let X = the number of female births.
7. Of all patients suffering a particular illness, 35% experience improvement from a particular medication. In the next 100 patients administered the medication, let X = the number of patients who experience improvement.

Binomial Distribution

Definition

A random experiment consists of n Bernoulli trials such that

(1) The trials are independent

(2) Each trial results in only two possible outcomes, labeled as "success" and "failure"

(3) The probability of a success in each trial, denoted as p, remains constant

The random variable X that equals the number of trials that result in a success has a **binomial random variable** with parameters $0 < p < 1$ and $n = 1, 2, \ldots$. The probability mass function of X is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \ldots, n \qquad (3\text{-}7)$$

Binomial Distribution

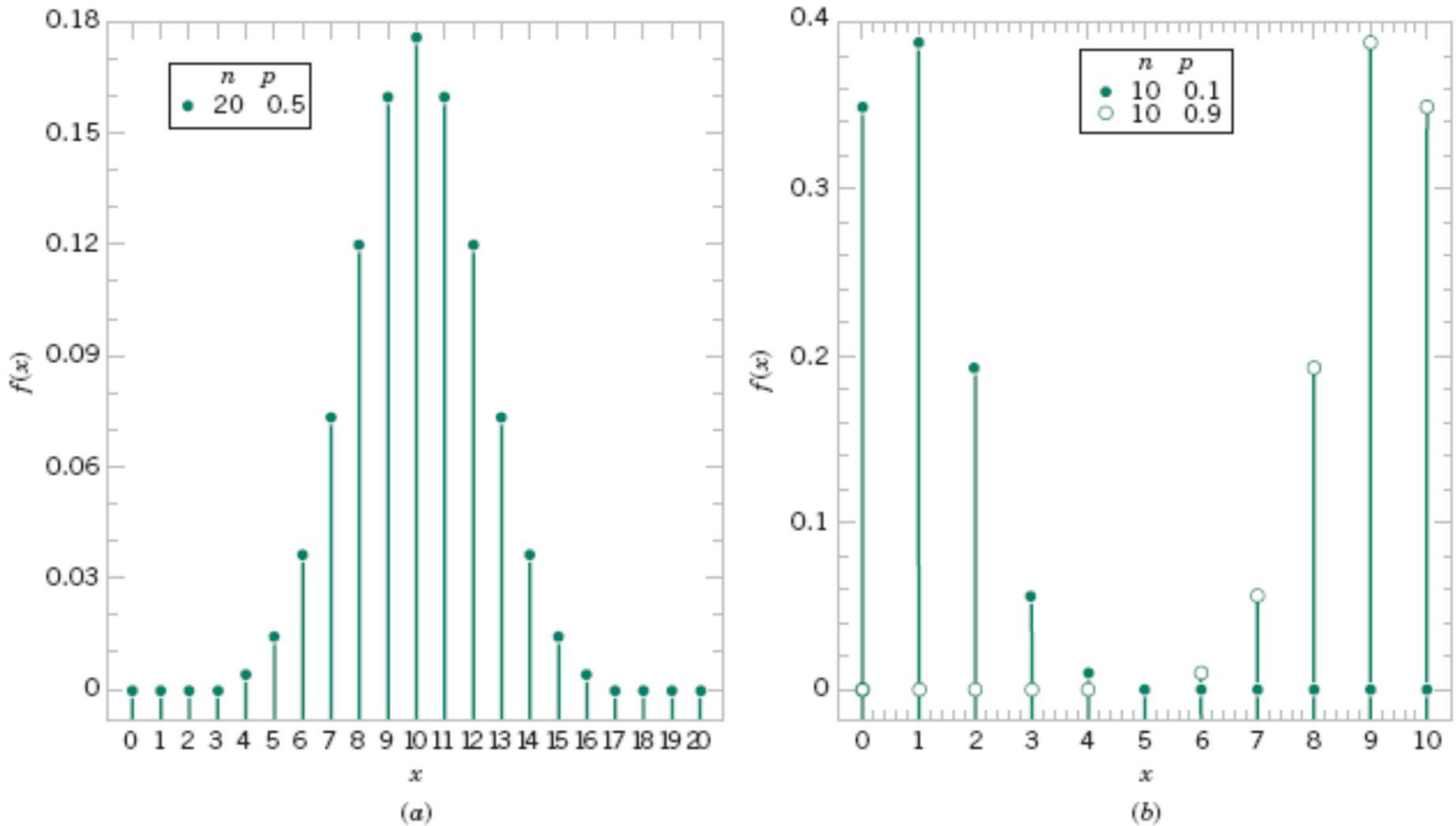


Figure Binomial distributions for selected values of *n* and *p*.

Binomial Distribution

Mean and Variance

If X is a binomial random variable with parameters p and n,

$$\mu = E(X) = np \qquad \text{and} \qquad \sigma^2 = V(X) = np(1-p) \tag{3-8}$$

Geometric and Negative Binomial Distributions

Definition

In a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until the first success. Then X is a **geometric random variable** with parameter $0 < p < 1$ and

$$f(x) = (1-p)^{x-1}p \qquad x = 1, 2, \ldots \tag{3-9}$$

Geometric and Negative Binomial Distributions

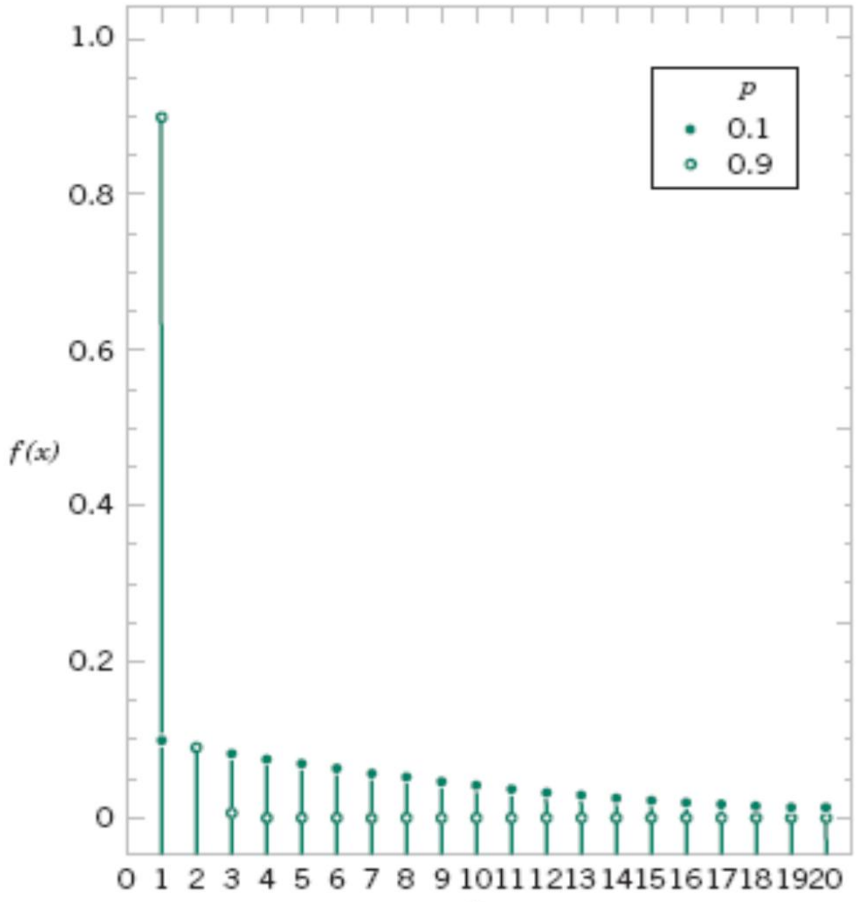


Figure. Geometric distributions for selected values of the parameter *p*.

Geometric and Negative Binomial Distributions

Definition

If X is a geometric random variable with parameter p,

$$\mu = E(X) = 1/p \qquad \text{and} \qquad \sigma^2 = V(X) = (1-p)/p^2 \tag{3-10}$$

Geometric and Negative Binomial Distributions

Negative Binomial Distribution

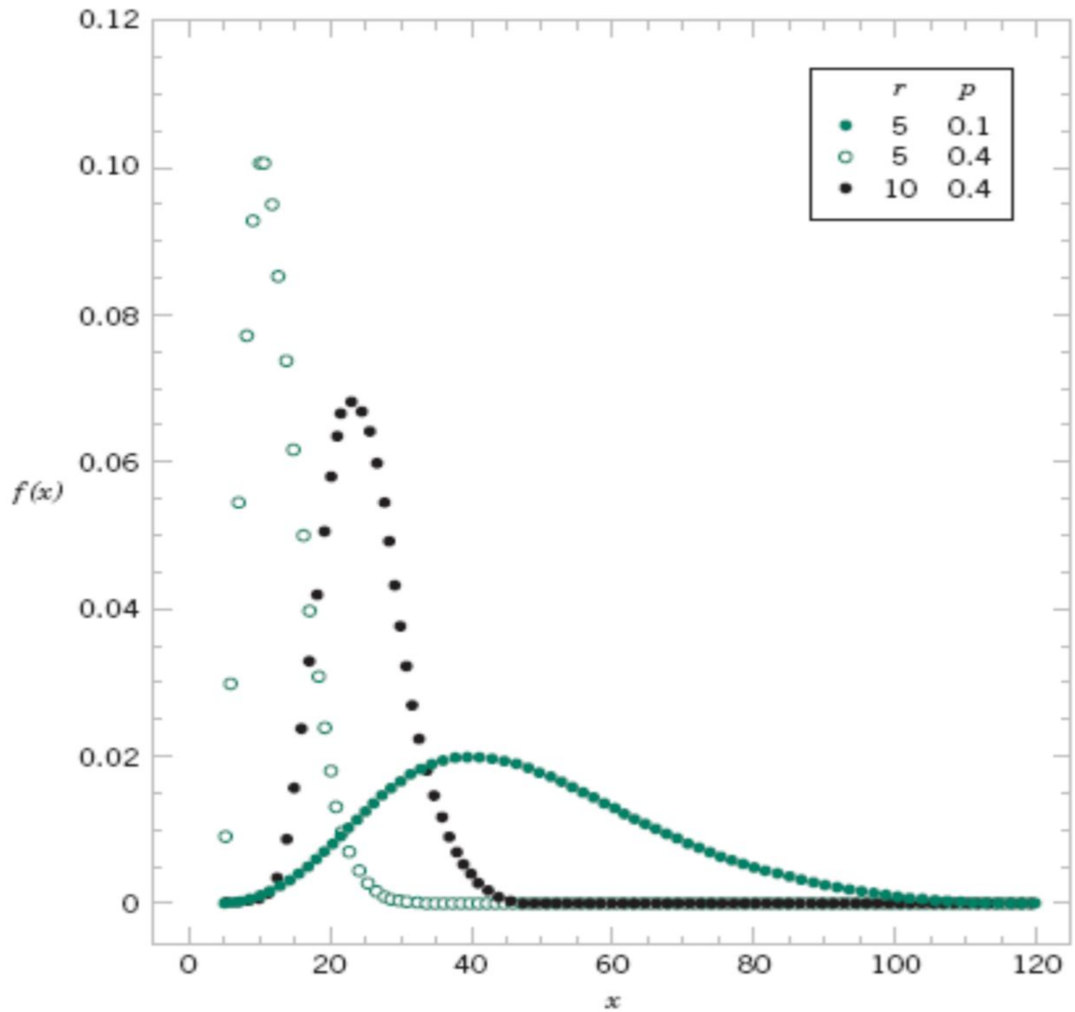
A generalization of a geometric distribution in which the random variable is the number of Bernoulli trials required to obtain r successes results in the **negative binomial distribution**.

In a series of Bernoulli trials (independent trials with constant probability p of a success), let the random variable X denote the number of trials until r successes occur. Then X is a **negative binomial random variable** with parameters $0 < p < 1$ and $r = 1, 2, 3, \ldots$, and

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r \qquad x = r, r+1, r+2, \ldots \qquad (3\text{-}11)$$

Geometric and Negative Binomial Distributions

Figure. Negative binomial distributions for selected values of the parameters *r* and *p*.



Geometric and Negative Binomial Distributions

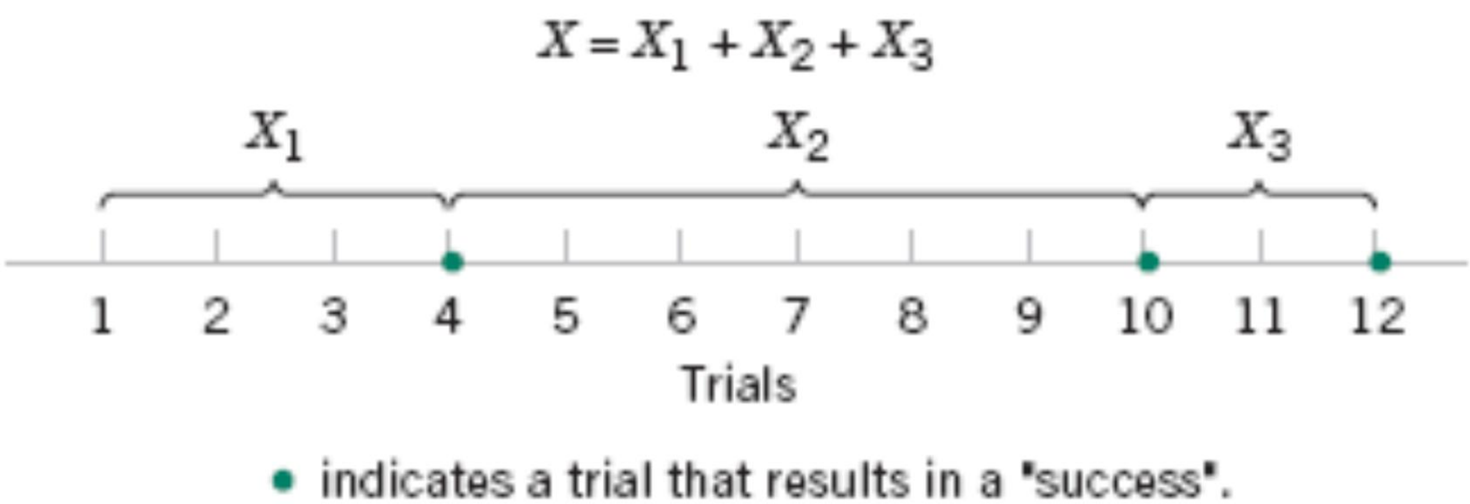


Figure . Negative binomial random variable represented as a sum of geometric random variables.

Geometric and Negative Binomial Distributions

Negative Binomial Distribution

If X is a negative binomial random variable with parameters p and r,

$$\mu = E(X) = r/p \qquad \text{and} \qquad \sigma^2 = V(X) = r(1-p)/p^2 \qquad (3\text{-}12)$$

Hypergeometric Distribution

Definition

A set of N objects contains

K objects classified as successes

$N - K$ objects classified as failures

A sample of size n objects is selected randomly (without replacement) from the N objects, where $K \le N$ and $n \le N$.

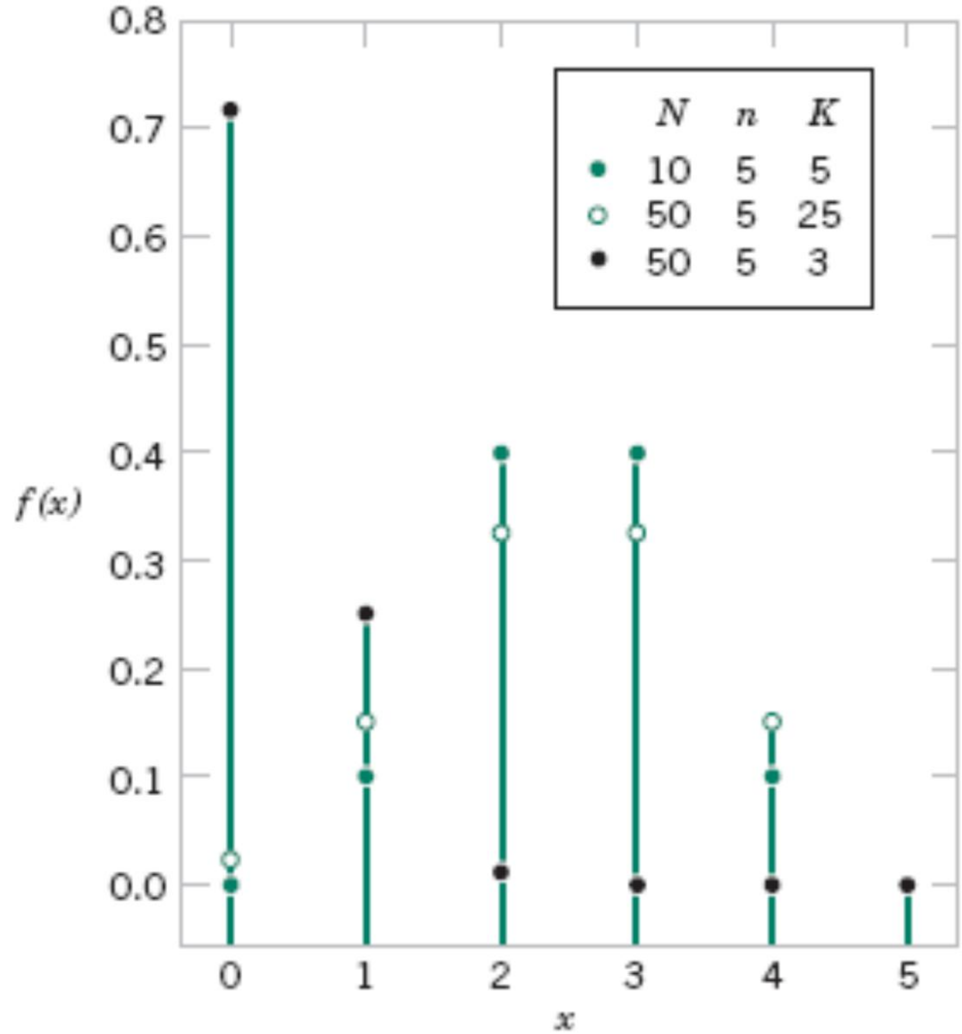
Let the random variable X denote the number of successes in the sample. Then X is a **hypergeometric random variable** and

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \qquad x = \max\{0, n + K - N\} \text{ to } \min\{K, n\} \qquad (3\text{-}13)$$

Vietnam Institute for Advanced Study in Mathematics

Hypergeometric Distribution

Figure. Hypergeometric distributions for selected values of parameters *N*, *K*, and *n*.



Advanced Study in Mathematics

Hypergeometric Distribution

Mean and Variance

If X is a hypergeometric random variable with parameters $N, K,$ and n, then

$$\mu = E(X) = np \qquad \text{and} \qquad \sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right) \quad \text{(3-14)}$$

where $p = K/N$.

Here p is interpreted as the proportion of successes in the set of N objects.

Hypergeometric Distribution

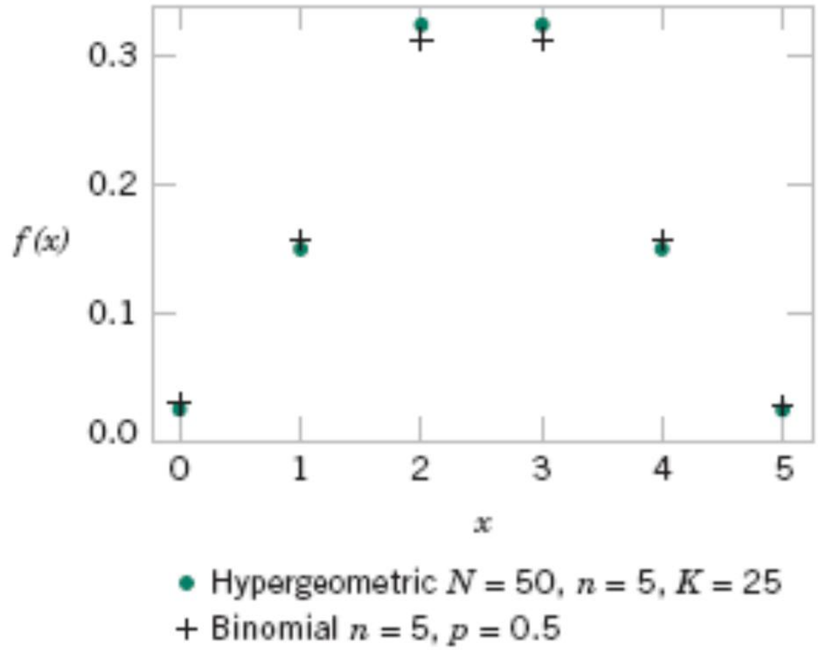
Finite Population Correction Factor

The term in the variance of a hypergeometric random variable

$$\frac{N-n}{N-1} \tag{3-15}$$

is called the finite population correction factor.

Hypergeometric Distribution



	0	1	2	3	4	5
Hypergeometric probability	0.025	0.149	0.326	0.326	0.149	0.025
Binomial probability	0.031	0.156	0.321	0.312	0.156	0.031

Figure. Comparison of hypergeometric and binomial distributions.

Poisson Distribution

Definition

Given an interval of real numbers, assume events occur at random throughout the interval. If the interval can be partitioned into subintervals of small enough length such that

(1) the probability of more than one event in a subinterval is zero,

(2) the probability of one event in a subinterval is the same for all subintervals and proportional to the length of the subinterval, and

(3) the event in each subinterval is independent of other subintervals, the random experiment is called a **Poisson process.**

The random variable X that equals the number of events in the interval is a **Poisson random variable** with parameter $0 < \lambda$, and the probability mass function of X is

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, \ldots \tag{3-16}$$

Poisson Distribution

Consistent Units

It is important to **use consistent units** in the calculation of probabilities, means, and variances involving Poisson random variables. The following example illustrates unit conversions. For example, if the

average number of flaws per millimeter of wire is 3.4, then the
average number of flaws in 10 millimeters of wire is 34, and the
average number of flaws in 100 millimeters of wire is 340.

Vietnam Institute for Advanced Study in Mathematics

Poisson Distribution

Mean and Variance

If X is a Poisson random variable with parameter λ, then

$$\mu = E(X) = \lambda \quad \text{and} \quad \sigma^2 = V(X) = \lambda \tag{3-17}$$

Continuous Random Variables

VIASM
Vietnam Institute for
Advanced Study in Mathematics

Probability Distributions and Probability Density Functions

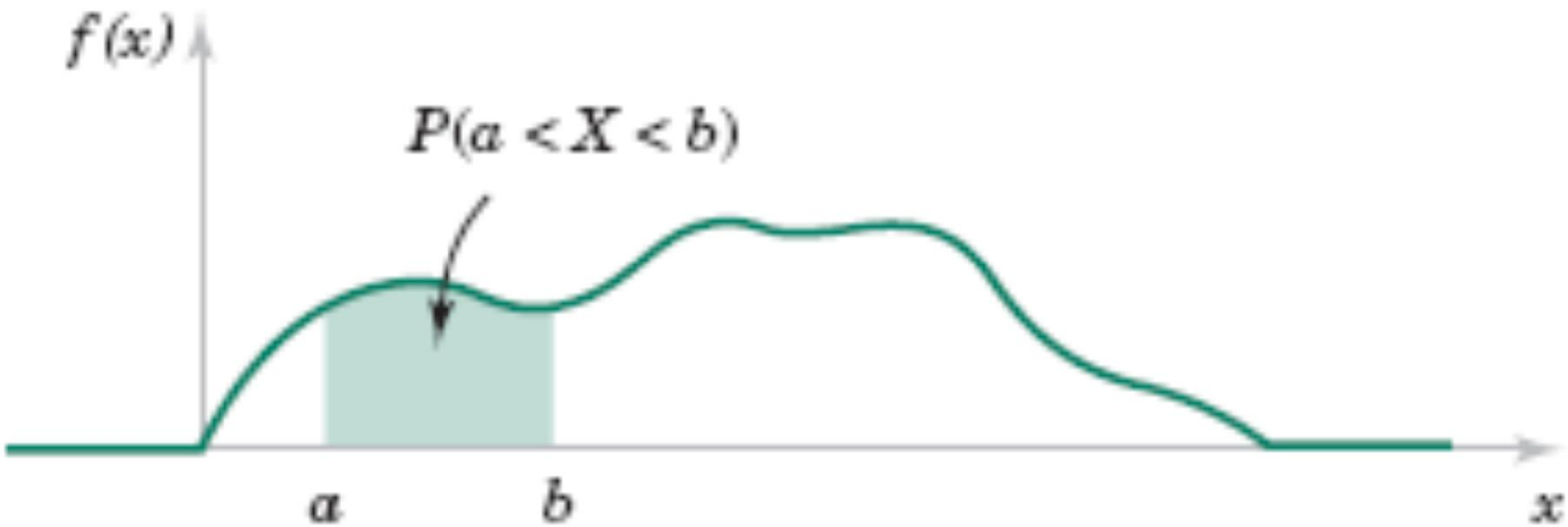


Figure Probability determined from the area under $f(x)$.

Probability Distributions and Probability Density Functions

Definition

For a continuous random variable X, a **probability density function** is a function such that

(1) $f(x) \geq 0$

(2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

(3) $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx =$ area under $f(x)$ from a to b

for any a and b (4-1)

Probability Distributions and Probability Density Functions

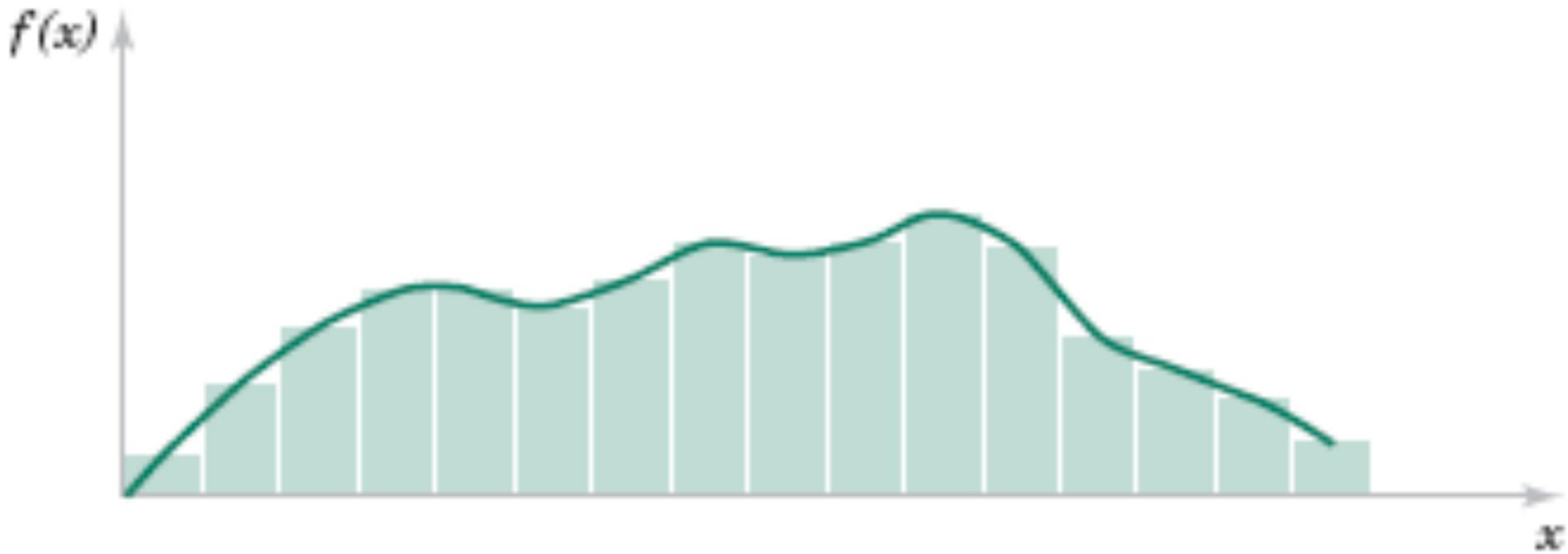


Figure Histogram approximates a probability density function.

Probability Distributions and Probability Density Functions

If X is a **continuous random variable**, for any x_1 and x_2,

$$P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2) \quad (4\text{-}2)$$

Cumulative Distribution Functions

Definition

The **cumulative distribution function** of a continuous random variable X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du \qquad (4\text{-}3)$$

for $-\infty < x < \infty$.

Mean and Variance of a Continuous Random Variable

Definition

Suppose X is a continuous random variable with probability density function $f(x)$. The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx \qquad (4\text{-}4)$$

The **variance** of X, denoted as $V(X)$ or σ^2, is

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

Mean and Variance of a Continuous Random Variable

Expected Value of a Function of a Continuous Random Variable

If X is a continuous random variable with probability density function $f(x)$,

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x)\, dx \qquad (4\text{-}5)$$

Continuous Uniform Random Variable

Definition

A continuous random variable X with probability density function

$$f(x) = 1/(b - a), \qquad a \le x \le b \tag{4-6}$$

is a **continuous uniform random variable.**

Continuous Uniform Random Variable

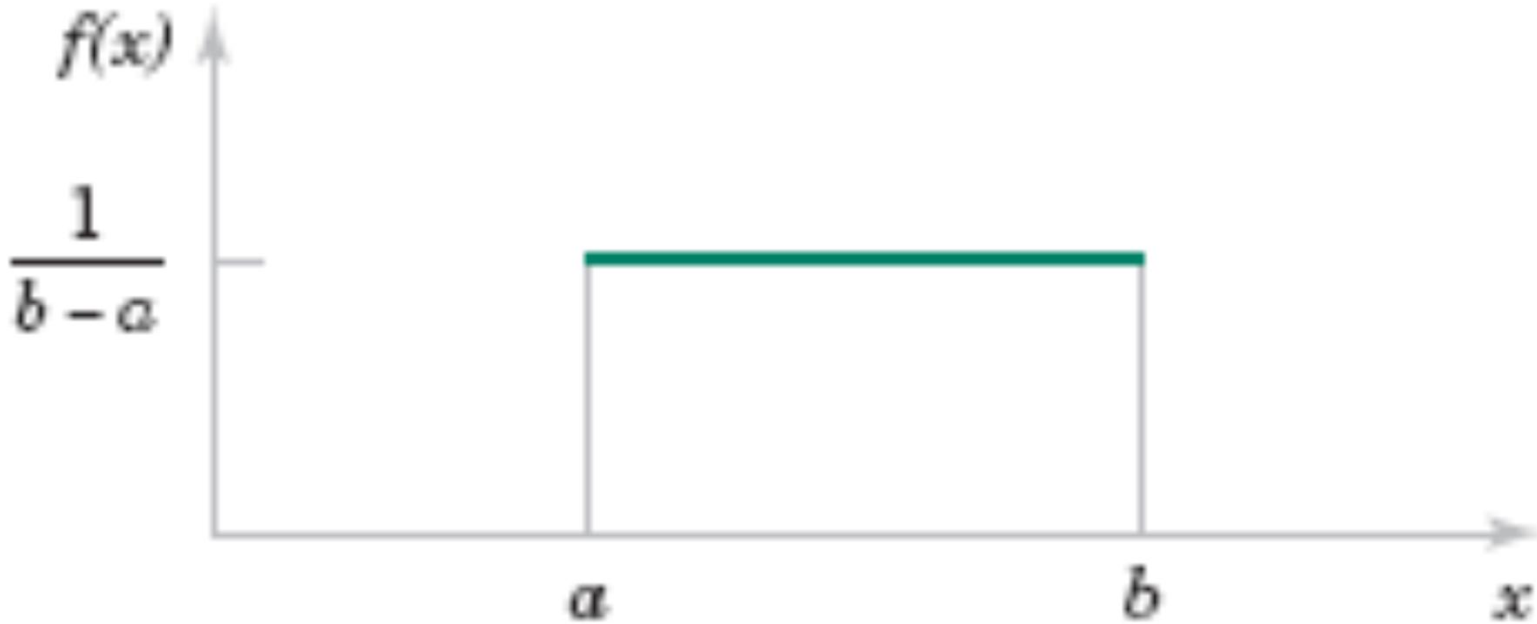


Figure Continuous uniform probability density function.

Continuous Uniform Random Variable

Mean and Variance

If X is a continuous uniform random variable over $a \le x \le b$,

$$\mu = E(X) = \frac{(a+b)}{2} \quad \text{and} \quad \sigma^2 = V(X) = \frac{(b-a)^2}{12} \tag{4-7}$$

Normal Distribution

Definition

A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \tag{4-8}$$

is a **normal random variable** with parameters μ, where $-\infty < \mu < \infty$, and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2 \tag{4-9}$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution. The mean and variance of X are shown to equal μ and σ^2, respectively, at the end of this Section 5-6.

Normal Distribution

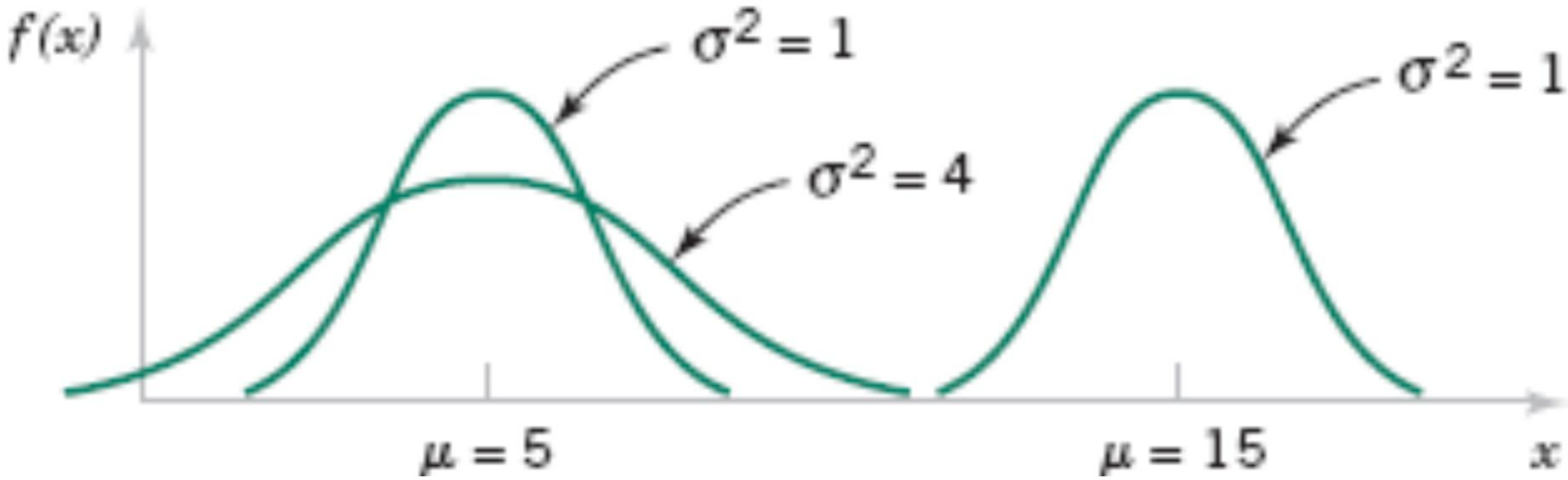


Figure Normal probability density functions for selected values of the parameters μ and σ^2.

Vietnam Institute for Advanced Study in Mathematics

Normal Distribution

Some useful results concerning the normal distribution

For any normal random variable,

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$

Normal Distribution

Definition : Standard Normal

A normal random variable with

$$\mu = 0 \quad \text{and} \quad \sigma^2 = 1$$

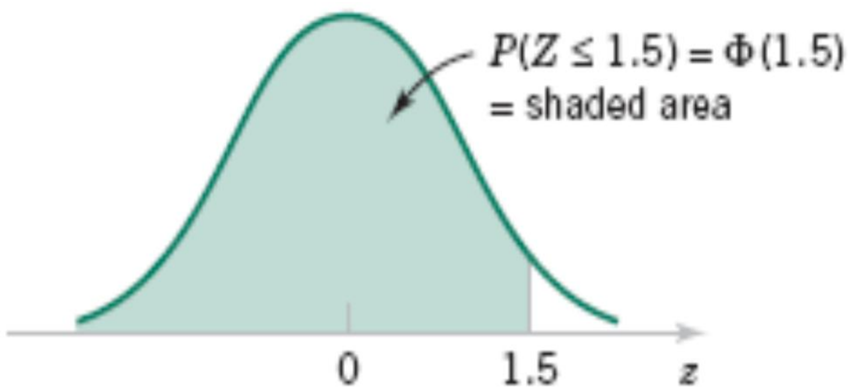
is called a **standard normal random variable** and is denoted as Z.

The cumulative distribution function of a standard normal random variable is denoted as

$$\Phi(z) = P(Z \le z)$$

Normal Distribution

Example



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Figure Standard normal probability density function.

Normal Distribution

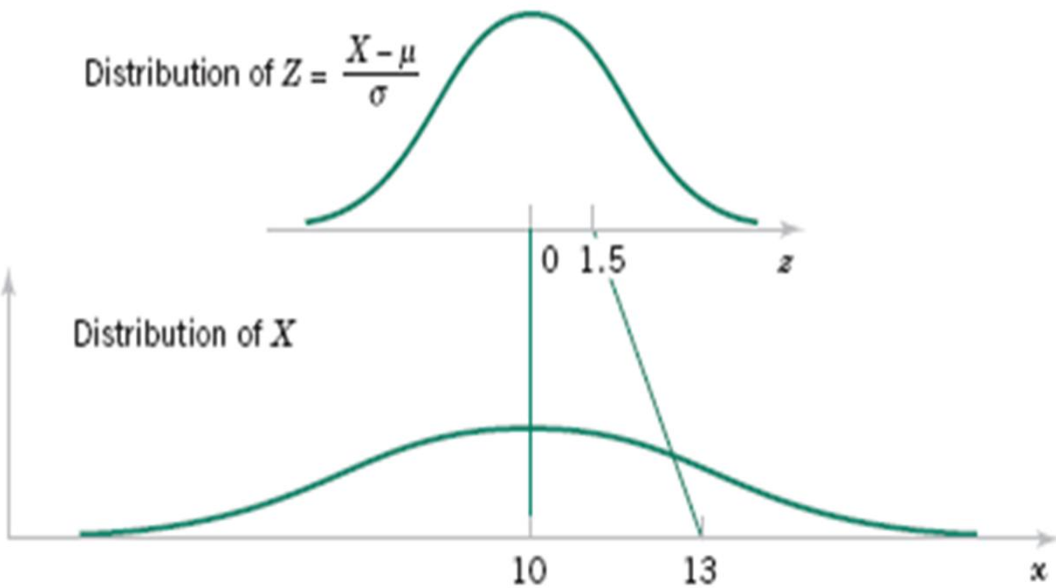
Standardizing

If X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, the random variable

$$Z = \frac{X - \mu}{\sigma} \tag{4-10}$$

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$. That is, Z is a standard normal random variable.

Normal Distribution



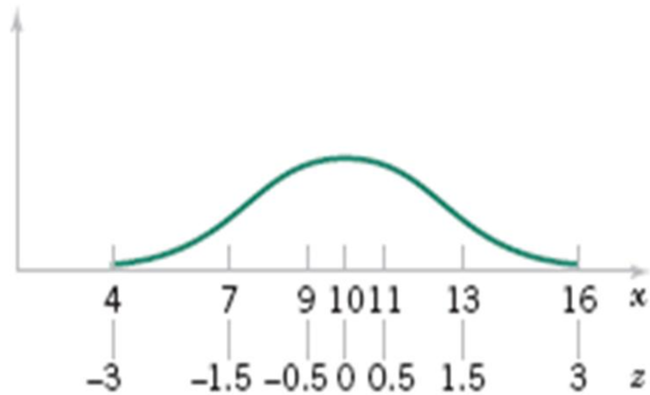


Figure Standardizing a normal random variable.

Normal Distribution

To Calculate Probability

Suppose X is a normal random variable with mean μ and variance σ^2. Then,

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P(Z \leq z) \qquad (4\text{-}11)$$

where Z is a **standard normal random variable,** and $z = \dfrac{(x-\mu)}{\sigma}$ is the **z-value** obtained by **standardizing** X.

The probability is obtained by entering Appendix Table II with $z = (x - \mu)/\sigma$.

Normal Approximation to the Binomial and Poisson Distributions

- Under certain conditions, the normal distribution can be used to approximate the binomial distribution and the Poisson distribution.

Normal Approximation to the Binomial and Poisson Distributions

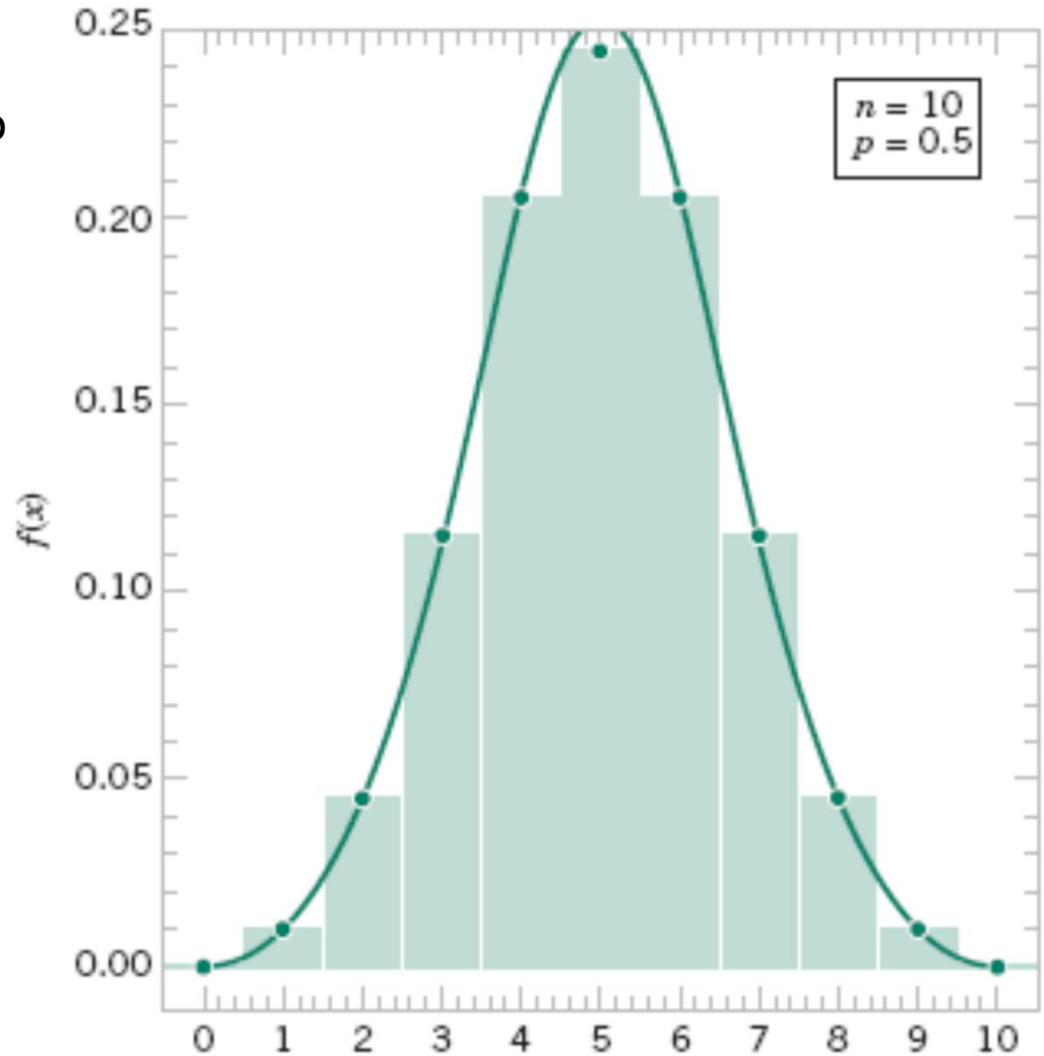


Figure Normal approximation to the binomial.

Normal Approximation to the Binomial and Poisson Distributions

Normal Approximation to the Binomial Distribution

If X is a binomial random variable, with parameters n and p

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \tag{4-12}$$

is approximately a standard normal random variable. To approximate a binomial probability with a normal distribution a **continuity correction** is applied as follows

$$P(X \le x) = P(X \le x + 0.5) \cong P\left(Z \le \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

and

$$P(x \le X) = P(x - 0.5 \le X) \cong P\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}} \le Z\right)$$

The approximation is good for $np > 5$ and $n(1-p) > 5$.

Normal Approximation to the Binomial and Poisson Distributions

hypergometric distribution	$\approx$ $\dfrac{n}{N} < 0.1$	binomial distribution	$\approx$ $np > 5$ $n(1-p) > 5$	normal distribution

Figure Conditions for approximating hypergeometric and binomial probabilities.

4-7 Normal Approximation to the Binomial and Poisson Distributions

Normal Approximation to the Poisson Distribution

If X is a Poisson random variable with $E(X) = \lambda$ and $V(X) = \lambda$,

$$Z = \frac{X - \lambda}{\sqrt{\lambda}} \tag{4-13}$$

is approximately a standard normal random variable. The approximation is good for

$$\lambda > 5$$

Vietnam Institute for Advanced Study in Mathematics

Exponential Distribution

Definition

The random variable X that equals the distance between successive events of a Poisson process with mean $\lambda > 0$ is an **exponential random variable** with parameter λ. The probability density function of X is

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \le x < \infty \tag{4-14}$$

4-8 Exponential Distribution

Mean and Variance

If the random variable X has an exponential distribution with parameter λ,

$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2} \qquad (4\text{-}15)$$

It is important to **use consistent units** in the calculation of probabilities, means, and variances involving exponential random variables. The following example illustrates unit conversions.

Exponential Distribution

Example 4-22 illustrates the **lack of memory property** of an exponential random variable and a general statement of the property follows. In fact, the exponential distribution is the only continuous distribution with this property.

Lack of Memory Property

For an exponential random variable X,

$$P(X < t_1 + t_2 | X > t_1) = P(X < t_2) \tag{4-16}$$

Exponential Distribution

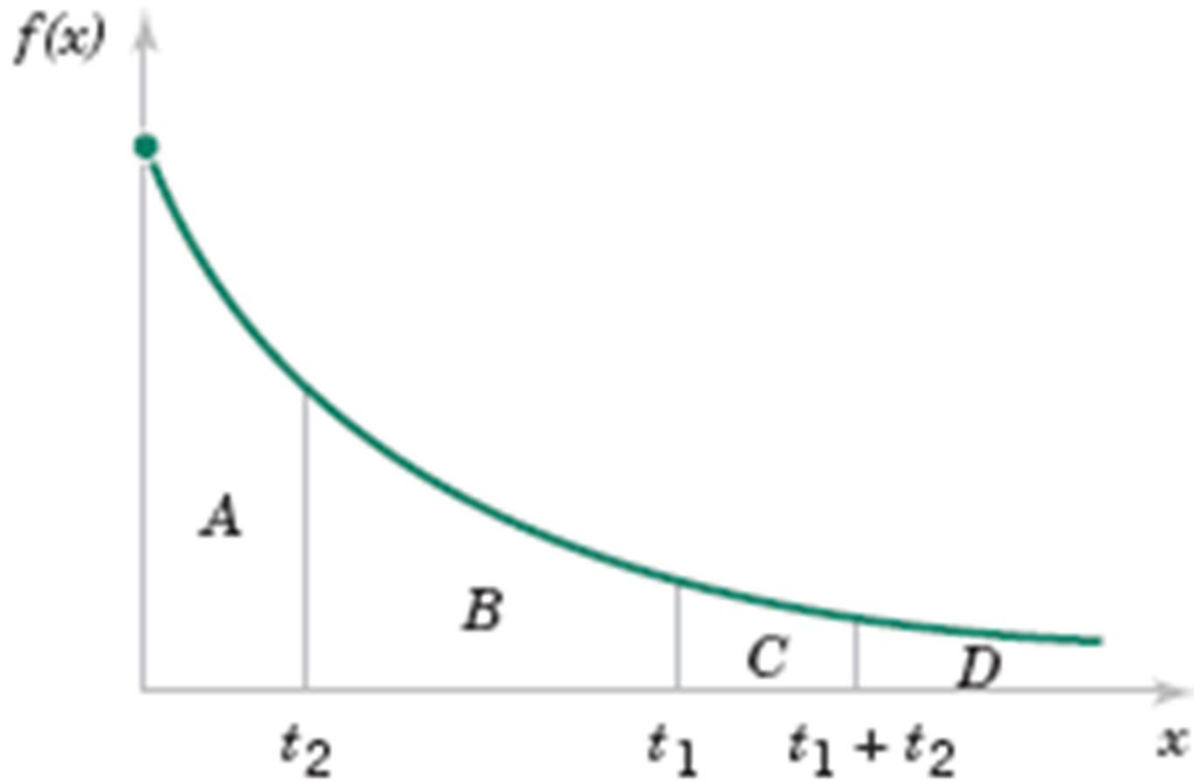


Figure Lack of memory property of an Exponential distribution.

Erlang and Gamma Distributions

Erlang Distribution

The random variable *X* that equals the interval length until *r* counts occur in a Poisson process with mean λ > 0 has and **Erlang random variable** with parameters λ and *r*. The probability density function of *X* is

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}$$

for *x* > 0 and *r* =1, 2, 3, ….

Erlang and Gamma Distributions

Gamma Distribution

The **gamma function** is

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x}\, dx, \quad \text{for } r > 0 \tag{4-17}$$

Erlang and Gamma Distributions

Gamma Distribution

The random variable X with probability density function

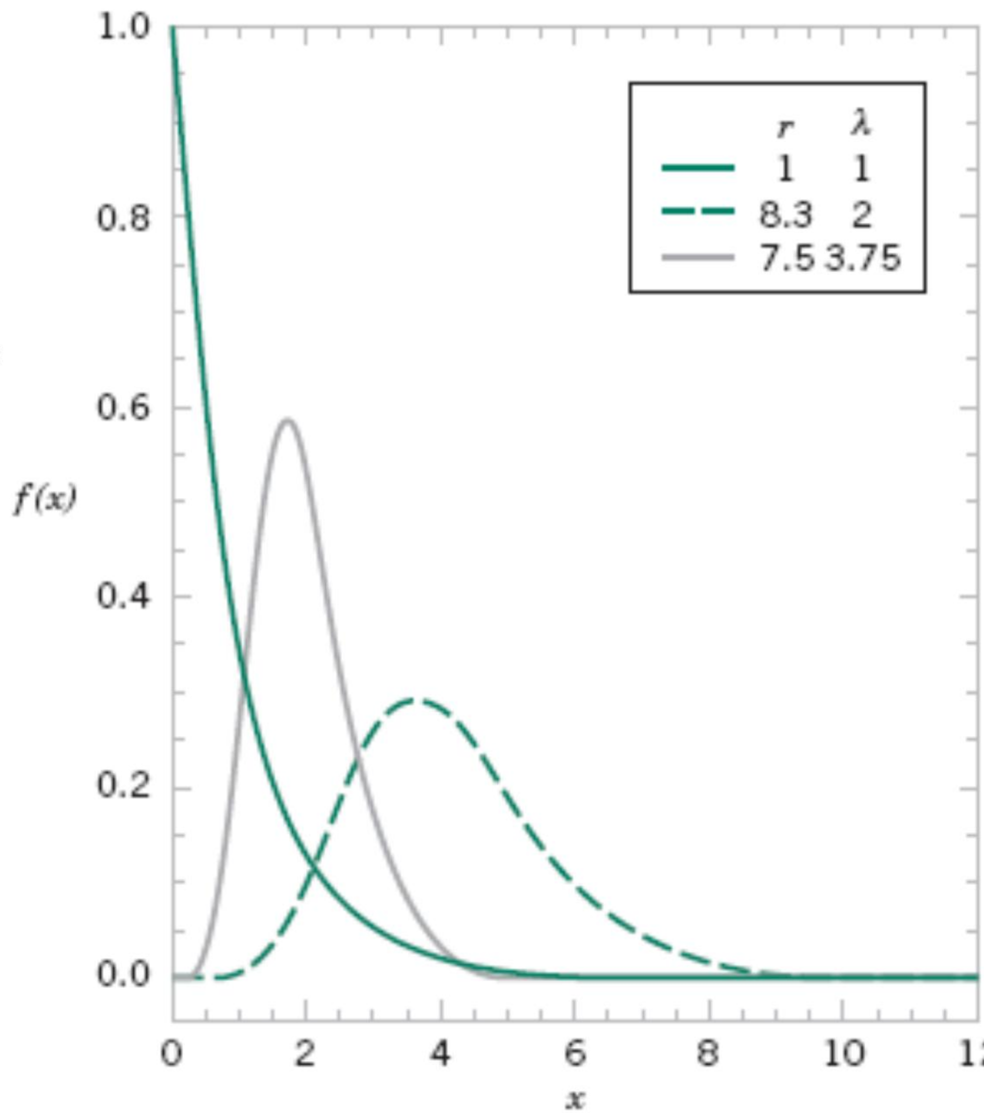
$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, \quad \text{for } x > 0 \qquad (4\text{-}18)$$

has a **gamma random variable** with parameters $\lambda > 0$ and $r > 0$. If r is an integer, X has an Erlang distribution.

Erlang and Gamma Distributions

Gamma Distribution

Figure Gamma probability dens functions for selected values of *r* and λ.



Erlang and Gamma Distributions

Gamma Distribution

If X is a **gamma random variable** with parameters λ and r,

$$\mu = E(X) = r/\lambda \quad \text{and} \quad \sigma^2 = V(X) = r/\lambda^2 \tag{4-19}$$

Lognormal Distribution

Let W have a normal distribution mean θ and variance ω^2; then $X = \exp(W)$ is a **lognormal random variable** with probability density function

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \theta)^2}{2\omega^2}\right] \qquad 0 < x < \infty$$

The mean and variance of X are

$$E(X) = e^{\theta + \omega^2/2} \qquad \text{and} \qquad V(X) = e^{2\theta + \omega^2}\left(e^{\omega^2} - 1\right) \qquad (4\text{-}22)$$

The parameters of a lognormal distribution are θ and ω^2, but care is needed to interpret that these are the mean and variance of the normal random variable W. The mean and variance of X are the functions of these parameters shown in (4-22). Figure 4-27 illustrates lognormal distributions for selected values of the parameters.

Vietnam Institute for Advanced Study in Mathematics

Lognormal Distribution

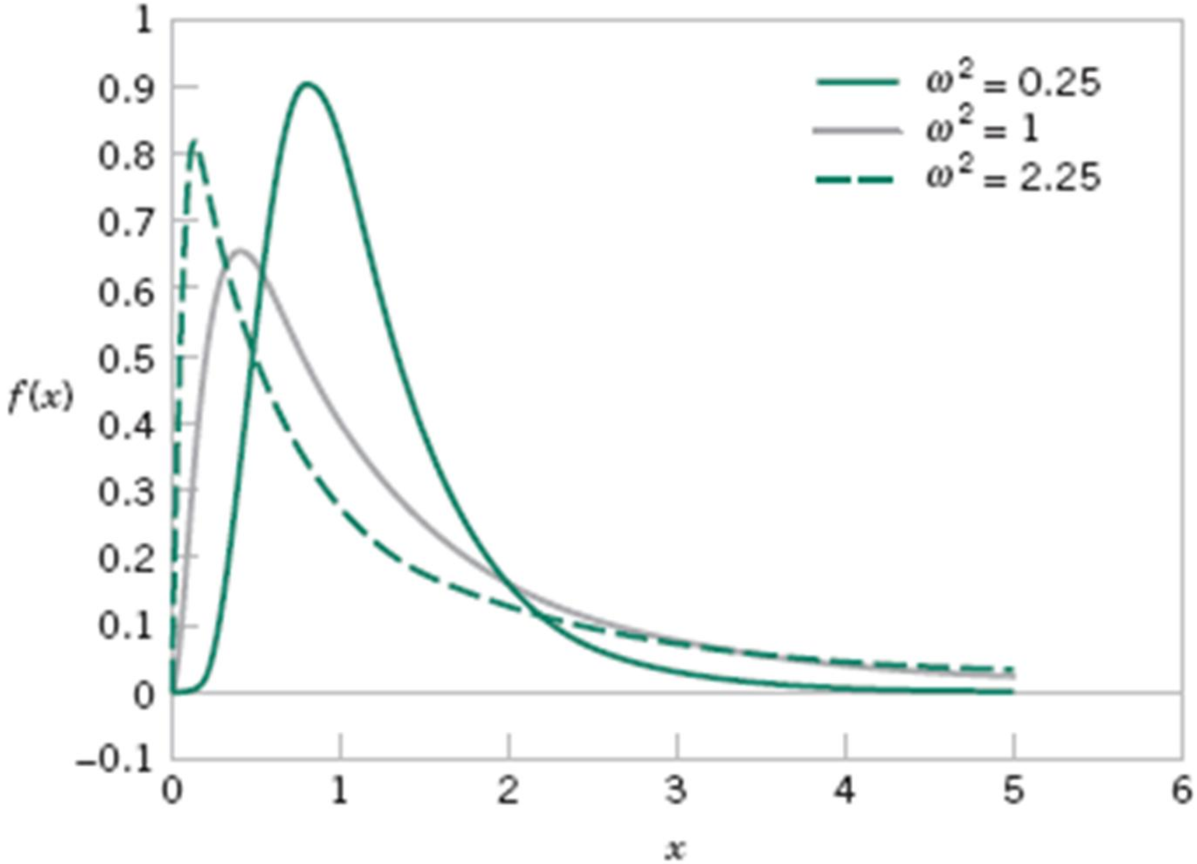


Figure Lognormal probability density functions with $\theta = 0$ for selected values of ω^2.

Joint Distribution

Two Discrete Random Variables

Joint Probability Distributions

The **joint probability mass function** of the discrete random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies

(1) $f_{XY}(x, y) \geq 0$

(2) $\sum_x \sum_y f_{XY}(x, y) = 1$

(3) $f_{XY}(x, y) = P(X = x, Y = y)$ (5-1)

Two Discrete Random Variables

Marginal Probability Distributions

• The individual probability distribution of a random variable is referred to as its **marginal probability distribution**.

• In general, the marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables. For example, to determine $P(X = x)$, we sum $P(X = x, Y = y)$ over all points in the range of (X, Y) for which $X = x$. Subscripts on the probability mass functions distinguish between the random variables.

Two Discrete Random Variables

	x = number of bars of signal strength				
y = number of times city name is stated	1	2	3		Marginal probability distribution of Y
4	0.15	0.1	0.05		0.3
3	0.02	0.1	0.05		0.17
2	0.02	0.03	0.2		0.25
1	0.01	0.02	0.25		0.28
	0.2	0.25	0.55		
	Marginal probability distribution of X				

Figure Marginal probability distributions of X and Y from Figure 5-1.

Two Discrete Random Variables

Definition: Marginal Probability Mass Functions

If X and Y are discrete random variables with joint probability mass function $f_{XY}(x, y)$, then the **marginal probability mass functions** of X and Y are

$$f_X(x) = P(X = x) = \sum_y f_{XY}(x, y) \quad \text{and} \quad f_Y(y) = P(Y = y) = \sum_x f_{XY}(x, y) \tag{5-2}$$

where the first sum is over all points in the range of (X, Y) for which $X = x$ and the second sum is over all points in the range of (X, Y) for which $Y = y$

Two Discrete Random Variables

Conditional Probability Distributions

Given discrete random variables X and Y with joint probability mass function $f_{XY}(x, y)$ the **conditional probability mass function** of Y given $X = x$ is

$$f_{Y|x}(y) = f_{XY}(x, y)/f_X(x) \qquad \text{for } f_X(x) > 0 \tag{5-3}$$

Two Discrete Random Variables

Conditional Probability Distributions

Because a conditional probability mass function $f_{Y|x}(y)$ is a probability mass function for all y in R_x, the following properties are satisfied:

(1) $f_{Y|x}(y) \geq 0$

(2) $\sum_y f_{Y|x}(y) = 1$

(3) $P(Y = y | X = x) = f_{Y|x}(y)$ (5-4)

Two Discrete Random Variables

Definition: Conditional Mean and Variance

The **conditional mean** of Y given $X = x$, denoted as $E(Y|x)$ or $\mu_{Y|x}$, is

$$E(Y|x) = \sum_y y f_{Y|x}(y) \tag{5-5}$$

and the **conditional variance** of Y given $X = x$, denoted as $V(Y|x)$ or $\sigma^2_{Y|x}$, is

$$V(Y|x) = \sum_y (y - \mu_{Y|x})^2 f_{Y|x}(y) = \sum_y y^2 f_{Y|x}(y) - \mu^2_{Y|x}$$

Example

For the joint probability distribution in Fig. 5-1, $f_{Y|x}(y)$ is found by dividing each $f_{XY}(x, y)$ by $f_X(x)$. Here, $f_X(x)$ is simply the sum of the probabilities in each column of Fig. 5-1. The function $f_{Y|x}(y)$ is shown in Fig. 5-3. In Fig. 5-3, each column sums to one because it is a probability distribution.

	x = number of bars of signal strength		
y = number of times city name is stated	1	2	3
4	0.750	0.400	0.091
3	0.100	0.400	0.091
2	0.100	0.120	0.364
1	0.050	0.080	0.454

Figure Conditional probability distributions of Y given X, $f_{Y|x}(y)$ in Example 5-6.

Two Discrete Random Variables

Independence

For discrete random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are **independent.**

(1) $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y

(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y, respectively. (5-6)

Two Discrete Random Variables

Multinomial Probability Distribution

Suppose a random experiment consists of a series of n trials. Assume that

(1) The result of each trial is classified into one of k classes.

(2) The probability of a trial generating a result in class 1, class 2, … , class k is constant over the trials and equal to $p_1, p_2, \ldots, p_k$, respectively.

(3) The trials are independent.

The random variables $X_1, X_2, \ldots, X_k$ that denote the number of trials that result in class 1, class 2, … , class k, respectively, have a **multinomial distribution** and the joint probability mass function is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \qquad (5\text{-}12)$$

for $x_1 + x_2 + \cdots + x_k = n$ and $p_1 + p_2 + \cdots + p_k = 1$.

Two Discrete Random Variables

Multinomial Probability Distribution

Each trial in a multinomial random experiment can be regarded as either generating or not generating a result in class i, for each $i = 1, 2, \ldots, k$. Because the random variable X_i is the number of trials that result in class i, X_i has a binomial distribution.

If $X_1, X_2, \ldots, X_k$ have a multinomial distribution, the marginal probability distribution of X_i is binomial with

$$E(X_i) = np_i \quad \text{and} \quad V(X_i) = np_i(1 - p_i) \tag{5-13}$$

Two Continuous Random Variables

Joint Probability Distribution

Definition

A **joint probability density function** for the continuous random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies the following properties:

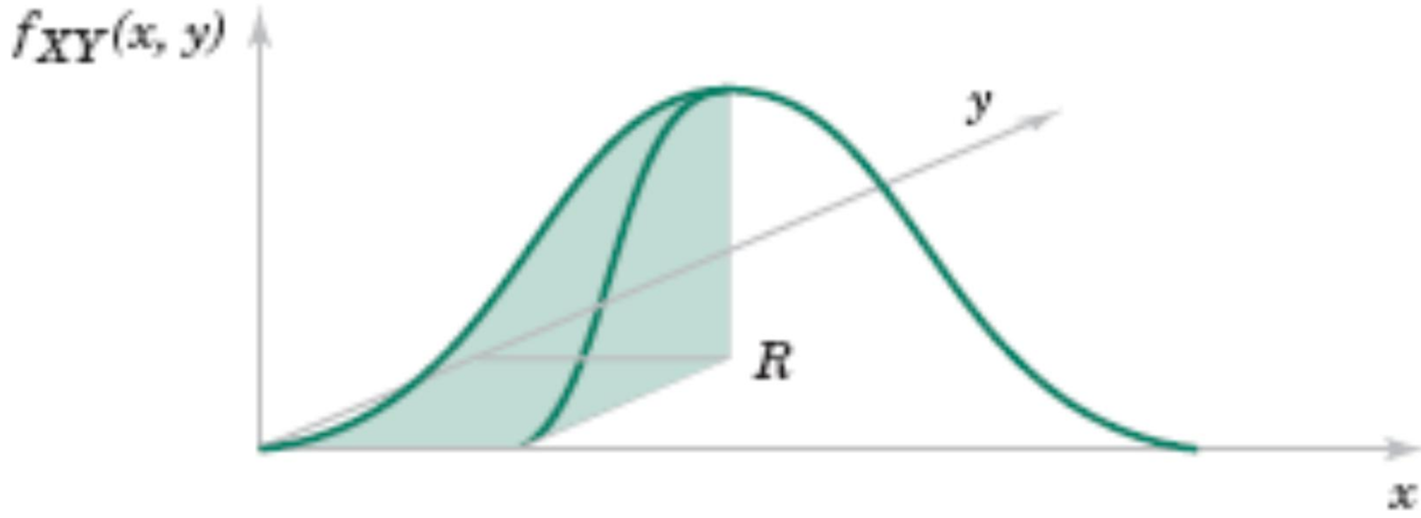
(1) $f_{XY}(x, y) \geq 0$ for all x, y

(2) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\, dy = 1$

(3) For any region R of two-dimensional space

$$P((X, Y) \in R) = \iint_R f_{XY}(x, y)\, dx\, dy \qquad (5\text{-}14)$$

Two Continuous Random Variables



Probability that (X, Y) is in the region R is determined by the volume of $f_{XY}(x, y)$ over the region R.

Joint probability density function for random variables *X* and *Y*.

Two Continuous Random Variables

Marginal Probability Distributions

Definition

If the joint probability density function of continuous random variables X and Y is $f_{XY}(x, y)$, the **marginal probability density functions** of X and Y are

$$f_X(x) = \int_y f_{XY}(x, y)\, dy \quad \text{and} \quad f_Y(y) = \int_x f_{XY}(x, y)\, dx \qquad (5\text{-}15)$$

where the first integral is over all points in the range of (X, Y) for which $X = x$ and the second integral is over all points in the range of (X, Y) for which $Y = y$

Two Continuous Random Variables

Conditional Probability Distributions

Definition

Given continuous random variables X and Y with joint probability density function $f_{XY}(x, y)$, the **conditional probability density function** of Y given $X = x$ is

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} \quad \text{for} \quad f_X(x) > 0 \qquad (5\text{-}16)$$

Vietnam Institute for Advanced Study in Mathematics

Two Continuous Random Variables

Conditional Probability Distributions

Because the conditional probability density function $f_{Y|x}(y)$ is a probability density function for all y in R_x, the following properties are satisfied:

(1) $f_{Y|x}(y) \geq 0$

(2) $\int_{R_x} f_{Y|x}(y)\, dy = 1$

(3) $P(Y \in B | X = x) = \int_B f_{Y|x}(y)\, dy$ for any set B in the range of Y

(5-17)

Two Continuous Random Variables

Definition: Conditional Mean and Variance

The **conditional mean** of Y given $X = x$, denoted as $E(Y|x)$ or $\mu_{Y|x}$, is

$$E(Y|x) = \int y f_{Y|x}(y)\, dy$$

and the **conditional variance** of Y given $X = x$, denoted as $V(Y|x)$ or $\sigma^2_{Y|x}$, is

$$V(Y|x) = \int_{-\infty} (y - \mu_{Y|x})^2 f_{Y|x}(y)\, dy = \int y^2 f_{Y|x}(y)\, dy - \mu^2_{Y|x} \qquad (5\text{-}18)$$

Two Continuous Random Variables

Independence

Definition

For continuous random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are said to be **independent.**

(1) $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y

(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y, respectively. (5-19)

Vietnam Institute for Advanced Study in Mathematics

Covariance and Correlation

Definition: Expected Value of a Function of Two Random Variables

$$E[h(X,Y)] = \begin{cases} \sum\sum_R h(x,y) f_{XY}(x,y) & X, Y \text{ discrete} \\ \iint_R h(x,y) f_{XY}(x,y)\,dx\,dy & X, Y \text{ continuous} \end{cases} \qquad (5\text{-}25)$$

Vietnam Institute for Advanced Study in Mathematics

Covariance and Correlation

Definition

The **covariance** between the random variables X and Y, denoted as $\text{cov}(X, Y)$ or σ_{XY}, is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y \qquad (5\text{-}26)$$

Covariance is a measure of **linear relationship** between the random variables. If the relationship between the random variables is nonlinear, the covariance might not be sensitive to the relationship. This is illustrated in Fig. 5-13(d). The only points with nonzero probability are the points on the circle. There is an identifiable relationship between the variables. Still, the covariance is zero.

Vietnam Institute for Advanced Study in Mathematics

Covariance and Correlation

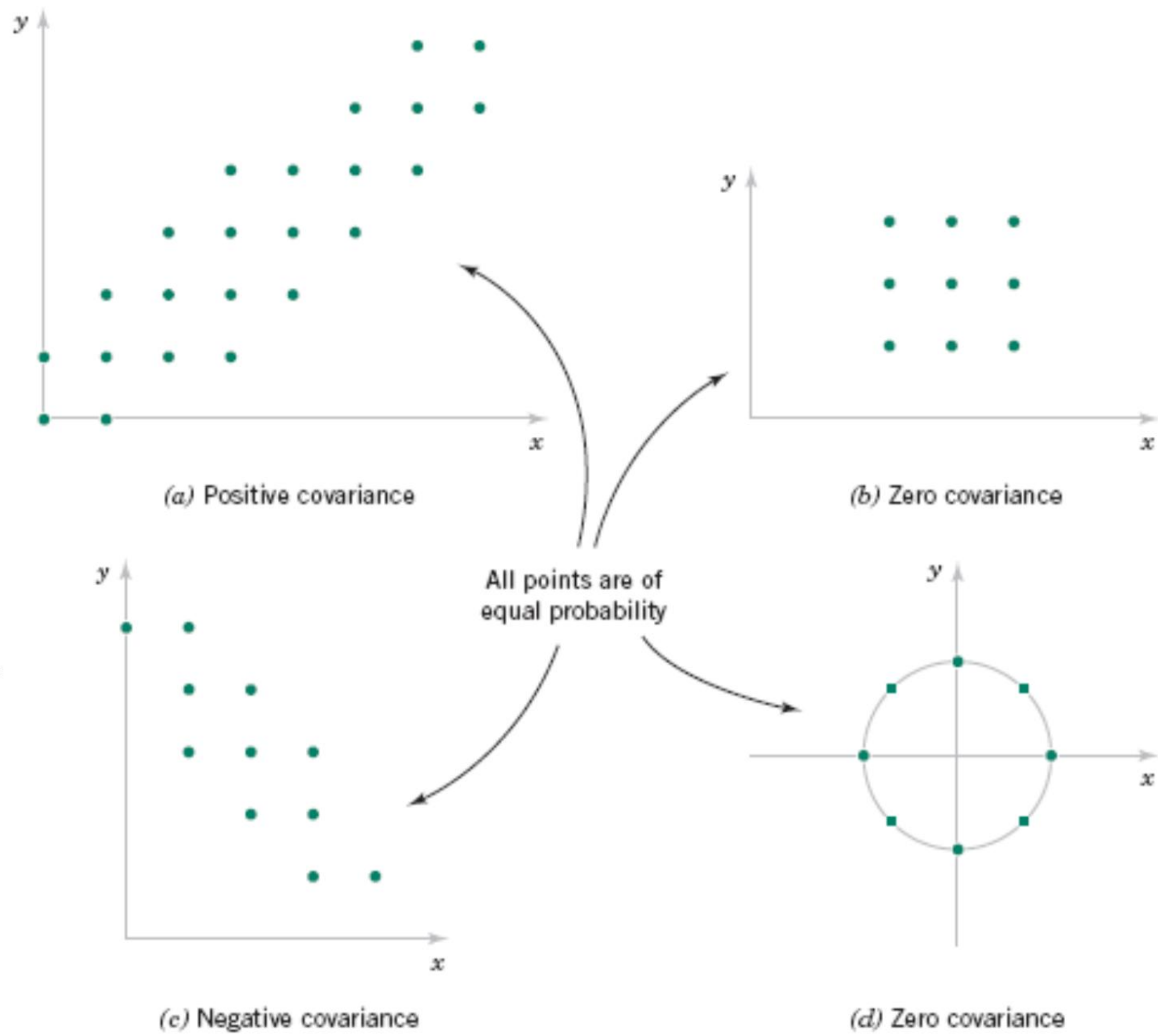


Figure Joint probability distributions and the sign of covariance between *X* and *Y*.

Covariance and Correlation

Definition

The **correlation** between random variables X and Y, denoted as ρ_{XY}, is

$$\rho_{XY} = \frac{\operatorname{cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{5-27}$$

For any two random variables X and Y

$$-1 \leq \rho_{XY} \leq +1 \tag{5-28}$$

Vietnam Institute for Advanced Study in Mathematics

Covariance and Correlation

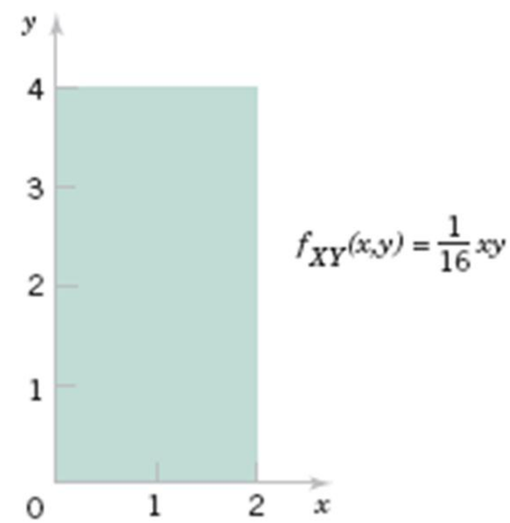
If X and Y are independent random variables,

$$\sigma_{XY} = \rho_{XY} = 0 \qquad (5\text{-}29)$$

Example

For the two random variables in Fig. 5-16, show that $\sigma_{XY} = 0$.

Figure 5-16 Random variables with zero covariance.



Bivariate Normal Distribution

Definition

The probability density function of a **bivariate normal distribution** is

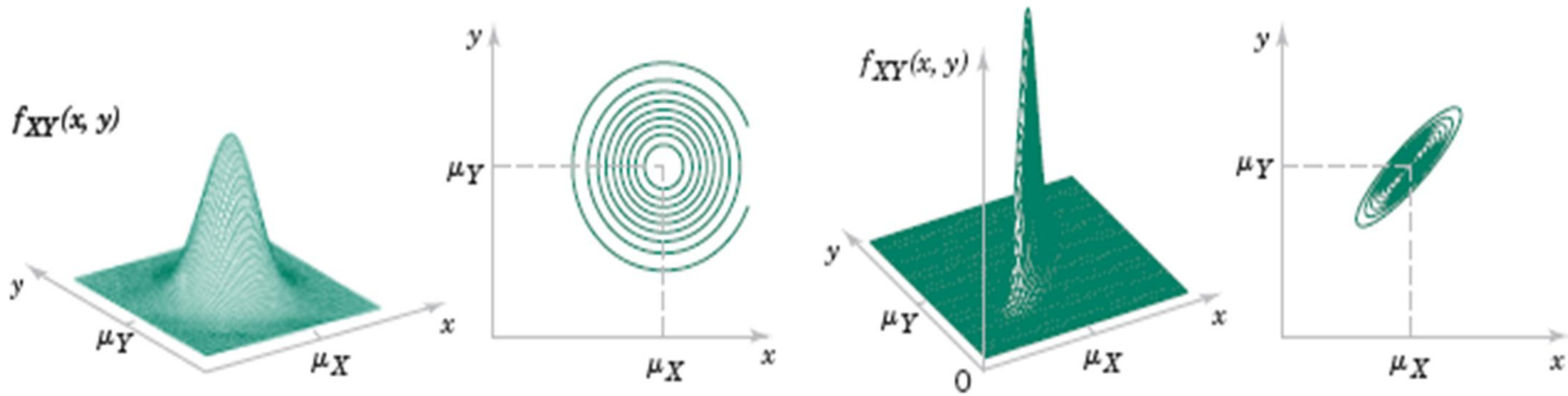
$$f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\} \quad (5\text{-}30)$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$, with parameters $\sigma_X > 0$, $\sigma_Y > 0$, $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, and $-1 < \rho < 1$.

Vietnam Institute for
Advanced Study in Mathematics

Bivariate Normal Distribution

Figure. Examples of bivariate normal distributions.

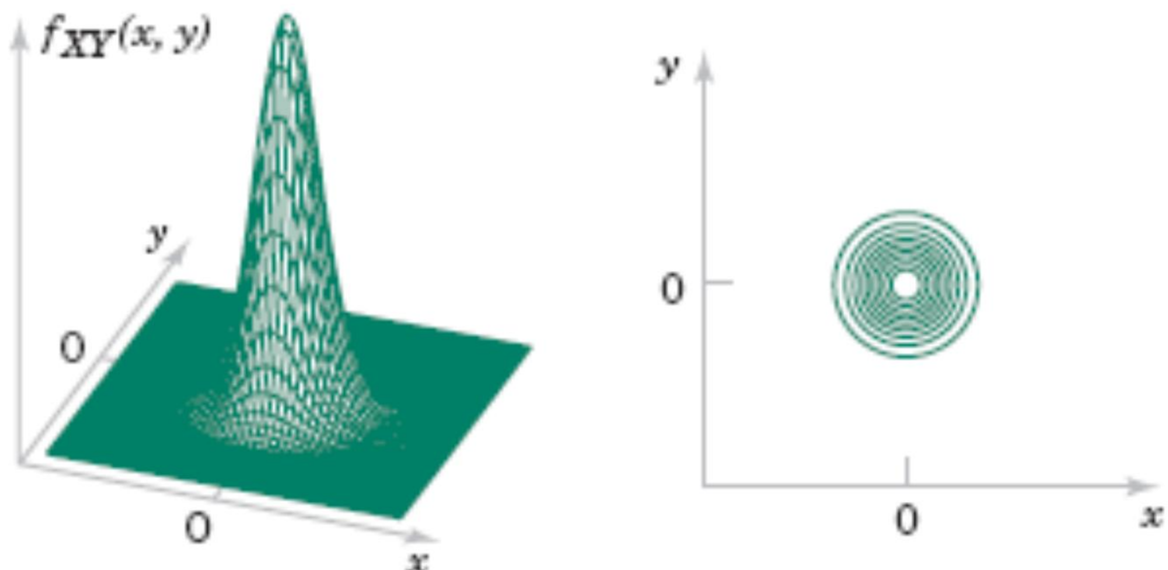


Bivariate Normal Distribution

Example

The joint probability density function $f_{XY}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-0.5(x^2+y^2)}$ is a special case of a bivariate normal distribution with $\sigma_X = 1$, $\sigma_Y = 1$, $\mu_X = 0$, $\mu_Y = 0$, and $\rho = 0$. This probability density function is illustrated in Fig. 5-18. Notice that the contour plot consists of concentric circles about the origin.

Figure 5-18



Bivariate Normal Distribution

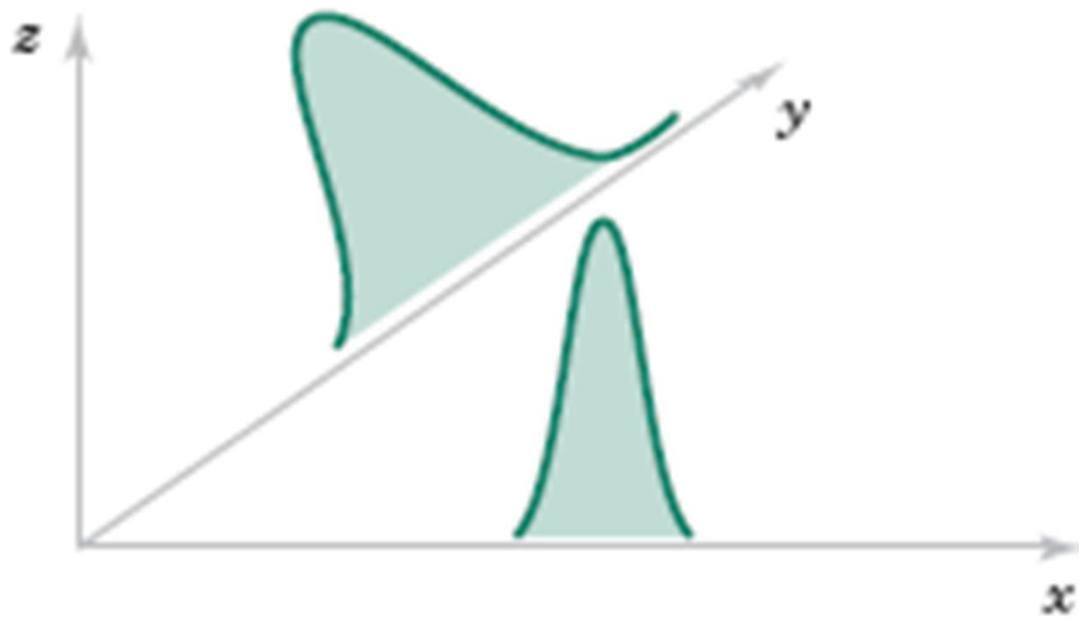
Marginal Distributions of Bivariate Normal Random Variables

If X and Y have a bivariate normal distribution with joint probability density $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$, the **marginal probability distributions** of X and Y are normal with means μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively. (5-31)

Bivariate Normal Distribution

Figure 5-19 illustrates that the marginal probability distributions of X and Y are normal. Furthermore, as the notation suggests, ρ represents the correlation between X and Y. The following result is left as an exercise.

Figure Marginal probability density functions of a bivariate normal distributions.



Bivariate Normal Distribution

If X and Y have a bivariate normal distribution with joint probability density function $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$, the correlation between X and Y is ρ. (5-32)

If X and Y have a bivariate normal distribution with $\rho = 0$, X and Y are independent. (5-33)

Linear Combinations of Random Variables

Definition

Given random variables $X_1, X_2, \ldots, X_p$ and constants $c_1, c_2, \ldots, c_p$,

$$Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p \tag{5-34}$$

is a **linear combination** of $X_1, X_2, \ldots, X_p$.

Mean of a Linear Combination

If $Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$,

$$E(Y) = c_1E(X_1) + c_2E(X_2) + \cdots + c_pE(X_p) \tag{5-35}$$

Linear Combinations of Random Variables

Variance of a Linear Combination

If $X_1, X_2, \ldots, X_p$ are random variables, and $Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$, then in general

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) + 2\sum_{i<j}\sum c_ic_j \operatorname{cov}(X_i, X_j) \quad (5\text{-}36)$$

If $X_1, X_2, \ldots, X_p$ are **independent,**

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) \quad (5\text{-}37)$$

Linear Combinations of Random Variables

Mean and Variance of an Average

If $\overline{X} = (X_1 + X_2 + \cdots + X_p)/p$ with $E(X_i) = \mu$ for $i = 1, 2, \ldots, p$

$$E(\overline{X}) = \mu \tag{5-38a}$$

if $X_1, X_2, \ldots, X_p$ are also independent with $V(X_i) = \sigma^2$ for $i = 1, 2, \ldots, p$,

$$V(\overline{X}) = \frac{\sigma^2}{p} \tag{5-38b}$$

Linear Combinations of Random Variables

Reproductive Property of the Normal Distribution

If $X_1, X_2, \ldots, X_p$ are independent, normal random variables with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$, for $i = 1, 2, \ldots, p$,

$$Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$$

is a normal random variable with

$$E(Y) = c_1\mu_1 + c_2\mu_2 + \cdots + c_p\mu_p$$

and

$$V(Y) = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_p^2\sigma_p^2 \qquad (5\text{-}39)$$

General Functions of Random Variables

A Discrete Random Variable

Suppose that X is a **discrete** random variable with probability distribution $f_X(x)$. Let $Y = h(X)$ define a one-to-one transformation between the values of X and Y so that the equation $y = h(x)$ can be solved uniquely for x in terms of y. Let this solution be $x = u(y)$. Then the probability mass function of the random variable Y is

$$f_Y(y) = f_X[u(y)] \tag{5-40}$$

Vietnam Institute for Advanced Study in Mathematics

General Functions of Random Variables

A Continuous Random Variable

Suppose that X is a **continuous** random variable with probability distribution $f_X(x)$. The function $Y = h(X)$ is a one-to-one transformation between the values of Y and X so that the equation $y = h(x)$ can be uniquely solved for x in terms of y. Let this solution be $x = u(y)$. The probability distribution of Y is

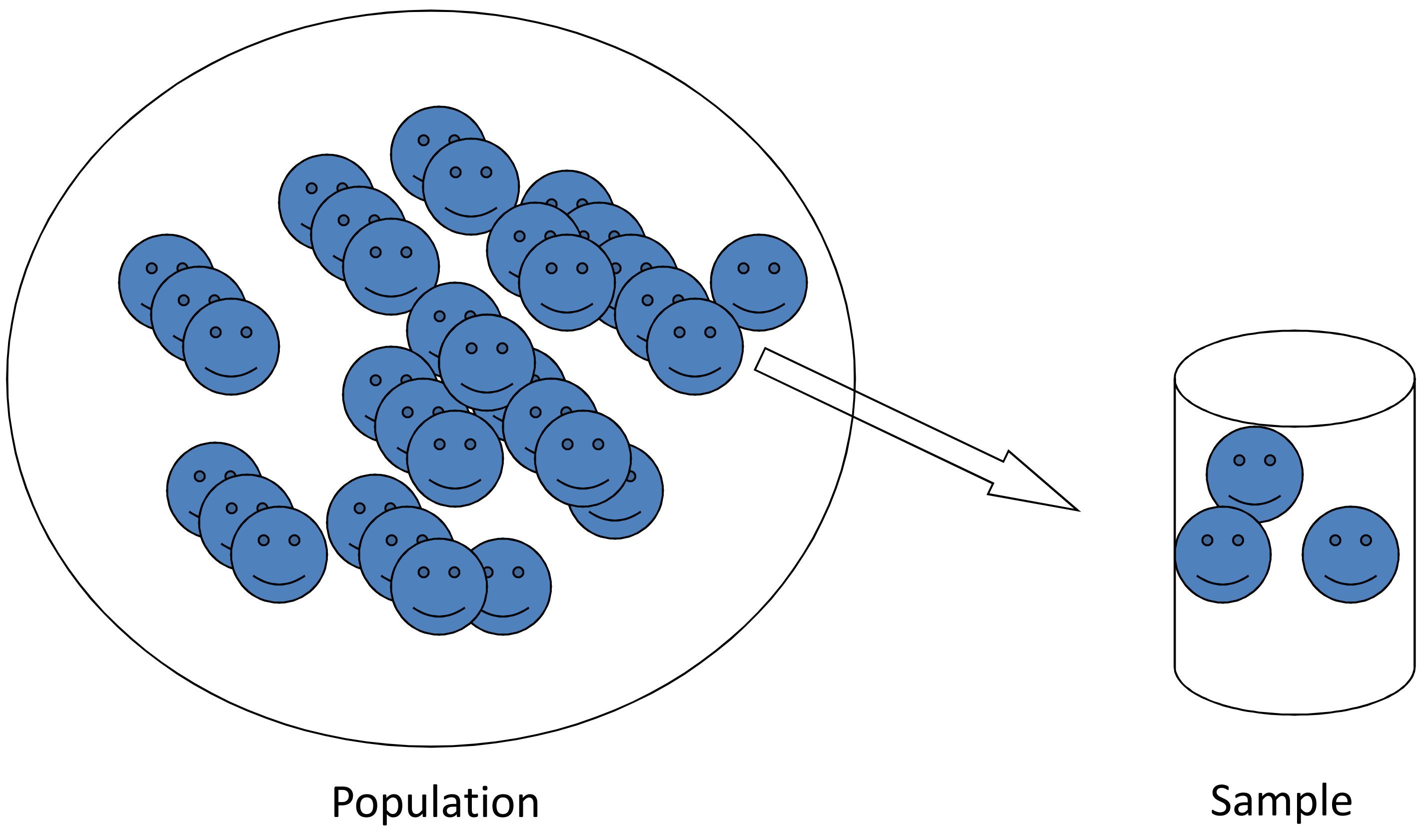
$$f_Y(y) = f_X[u(y)]\,|J| \tag{5-41}$$

where $J = u'(y)$ is called the **Jacobian** of the transformation and the absolute value of J is used.

Random sampling and Data description

Vietnam Institute for
Advanced Study in Mathematics

Sample vs. Population



Descriptive Statistics

An Illustration:

Which Group is Smarter?

Class A--IQs of 13 Students		Class B--IQs of 13 Students	
102	115	127	162
128	109	131	103
131	89	96	111
98	106	80	109
140	119	93	87
93	97	120	105
110		109	

Each individual may be different. If you try to understand a group by remembering the qualities of each member, you become overwhelmed and fail to understand the group.

Descriptive **Statistics**

Which group is smarter now?

Class A--Average IQ	Class B--Average IQ
110.54	110.23

They're roughly the same!

With a summary descriptive statistic, it is much easier to answer our question.

Vietnam Institute for Advanced Study in Mathematics

Descriptive Statistics

Types of descriptive statistics:

- Organize Data
  - Tables
  - Graphs

- Summarize Data
  - Central Tendency
  - Variation

Descriptive Statistics

Types of descriptive statistics:

- Organize Data
  - Tables
    - Frequency Distributions
    - Relative Frequency Distributions
  - Graphs
    - Bar Chart or Histogram
    - Stem and Leaf Plot
    - Frequency Polygon

Descriptive Statistics

Summarizing Data:

- Central Tendency (or Groups' "Middle Values")
  - Mean
  - Median
  - Mode

- Variation (or Summary of Differences Within Groups)
  - Range
  - Interquartile Range
  - Variance
  - Standard Deviation

Mean

Most commonly called the "average."

Add up the values for each case and divide by the total number of cases.

$$\text{Y-bar} = \frac{(Y1 + Y2 + \ldots + Yn)}{n}$$

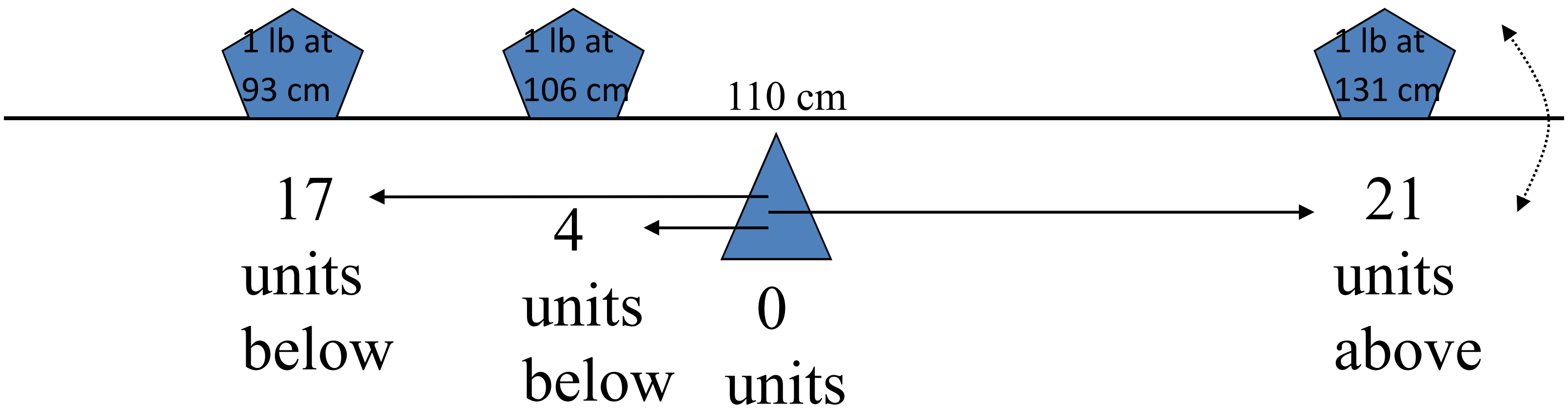
$$\text{Y-bar} = \frac{\Sigma\, Y\textit{i}}{n}$$

Vietnam Institute for Advanced Study in Mathematics

Mean

The mean is the "balance point."

Each person's score is like 1 pound placed at the score's position on a see-saw. Below, on a 200 cm see-saw, the mean equals 110, the place on the see-saw where a fulcrum finds balance:

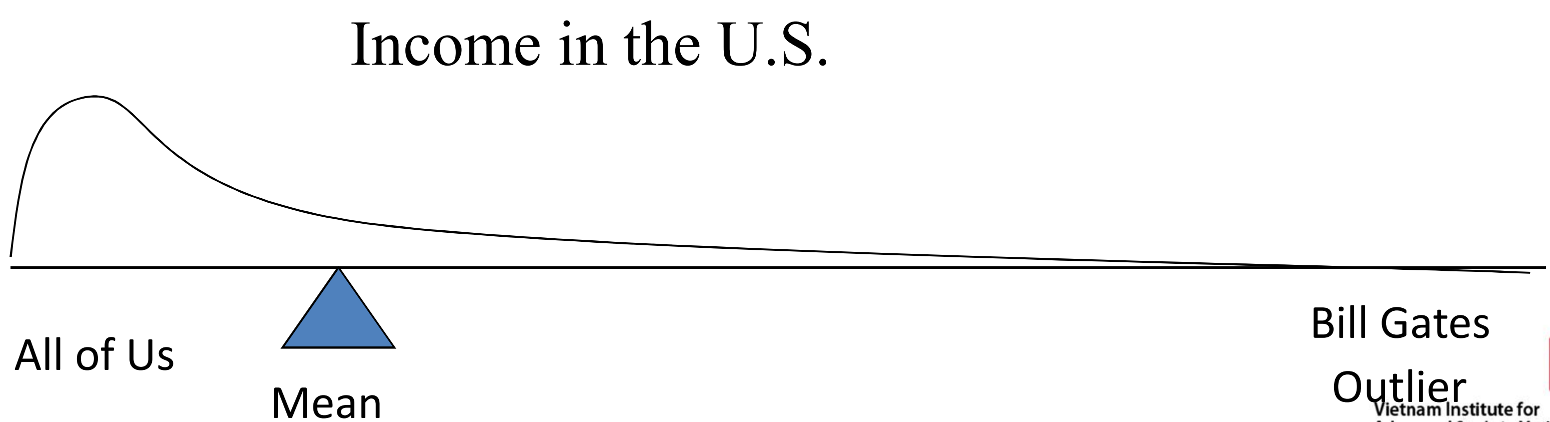


The scale is balanced because…

$$17 + 4 \text{ on the left} \quad = \quad 21 \text{ on the right}$$

Mean

1. Means can be badly affected by outliers (data points with extreme values unlike the rest)
2. Outliers can make the mean a bad measure of central tendency or common experience



Vietnam Institute for Advanced Study in Mathematics

Median

The middle value when a variable's values are ranked in order; the point that divides a distribution into two equal halves.

When data are listed in order, the median is the point at which 50% of the cases are above and 50% below it.

The 50^{th} percentile.

Median

Class A--IQs of 13 Students

89

93

97

98

102

106

109

110

115

119

128

131 140

Median = 109

(six cases above, six below)

Vietnam Institute for Advanced Study in Mathematics

Median

If the first student were to drop out of Class A, there would be a new median:

89
93
97
98
102
106
109
110
115
119
128
131
140

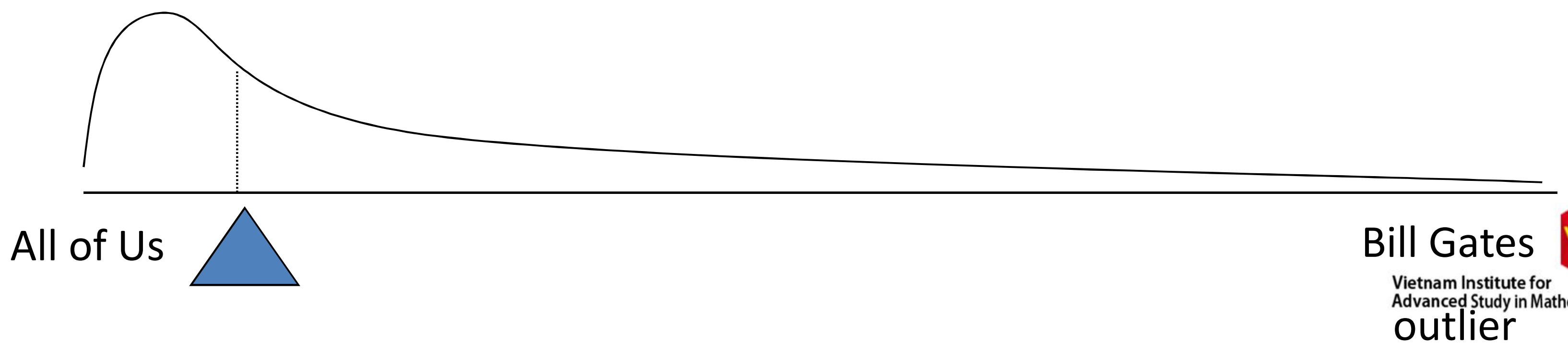
Median = 109.5

109 + 110 = 219/2 = 109.5

(six cases above, six below)

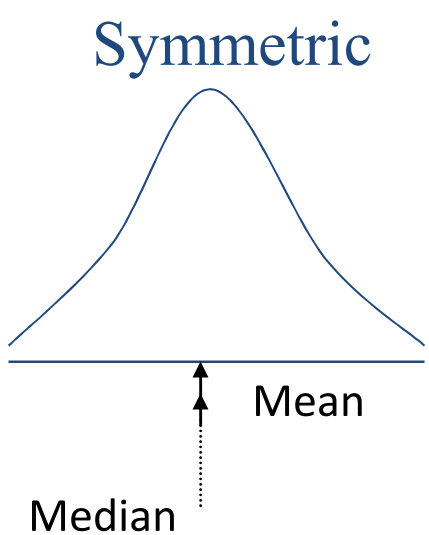
Median

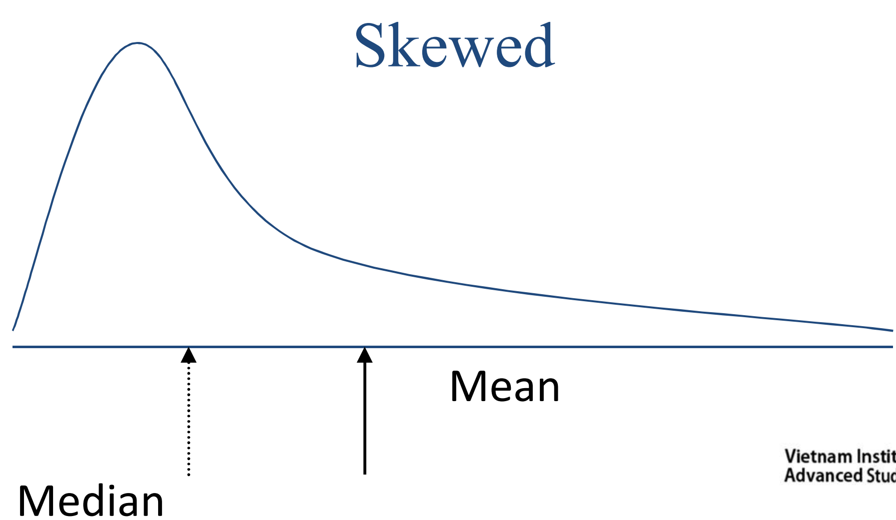
1. The median is unaffected by outliers, making it a better measure of central tendency, better describing the "typical person" than the mean when data are skewed.



Median

2. If the recorded values for a variable form a symmetric distribution, the median and mean are identical.
3. In skewed data, the mean lies further toward the skew than the median.





Vietnam Institute for Advanced Study in Mathematics

Median

The middle score or measurement in a set of ranked scores or measurements; the point that divides a distribution into two equal halves.

Data are listed in order—the median is the point at which 50% of the cases are above and 50% below.

The 50^{th} percentile.

Mode

The most common data point is called the mode.

The combined IQ scores for Classes A & B:

80 87 89 93 93 96 97 98 102 103 105 106 109 109 109 110 111 115 119 120 127 128 131 131 140 162

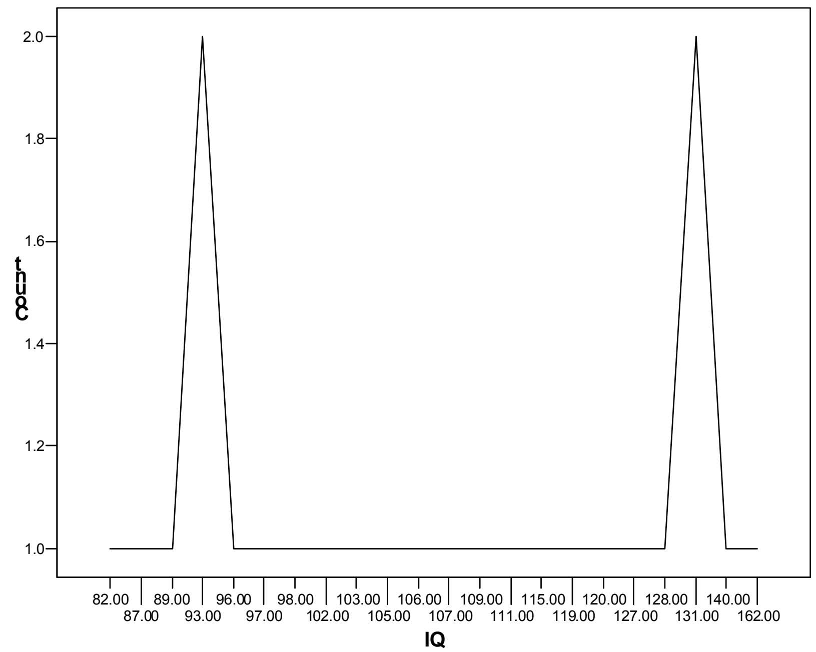
A la mode!!

BTW, It is possible to have more than one mode!

Mode

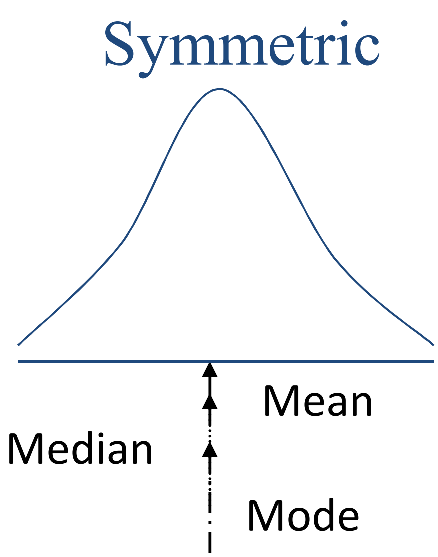
It may mot be at the center of a distribution.

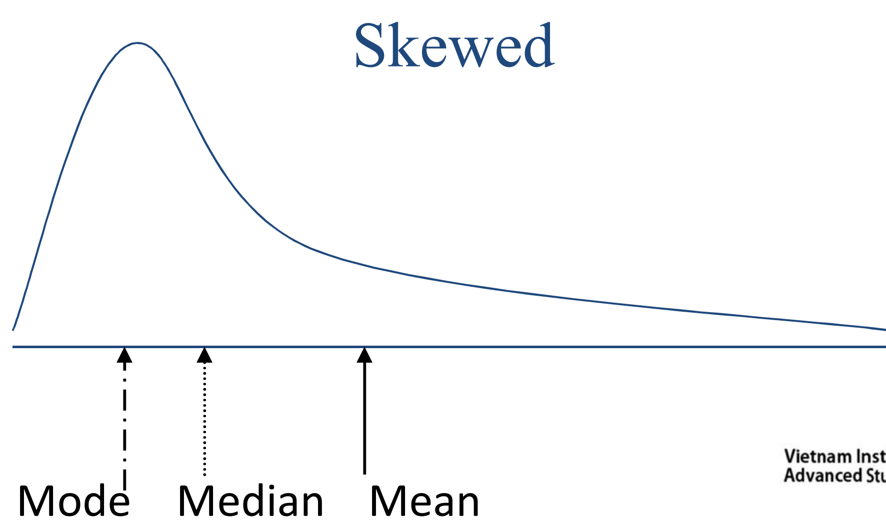
Data distribution on the right is “bimodal” (even statistics can be open-minded)



Mode

1. It may give you the most likely experience rather than the "typical" or "central" experience.
2. In symmetric distributions, the mean, median, and mode are the same.
3. In skewed data, the mean and median lie further toward the skew than the mode.





Descriptive Statistics

Summarizing Data:

- ✓ Central Tendency (or Groups' "Middle Values")
  - ✓Mean
  - ✓Median
  - ✓Mode

- ❑ Variation (or Summary of Differences Within Groups)
  - Range
  - Interquartile Range
  - Variance
  - Standard Deviation

Range

The spread, or the distance, between the lowest and highest values of a variable.

To get the range for a variable, you subtract its lowest value from its highest value.

Class A--IQs of 13 Students

102	115
128	109
131	89
98	106
140	119
93	97
110	

Class A Range = 140 - 89 = 51

Class B--IQs of 13 Students

127	162
131	103
96	111
80	109
93	87
120	105
109	

Class B Range = 162 - 80 = 82

Interquartile Range

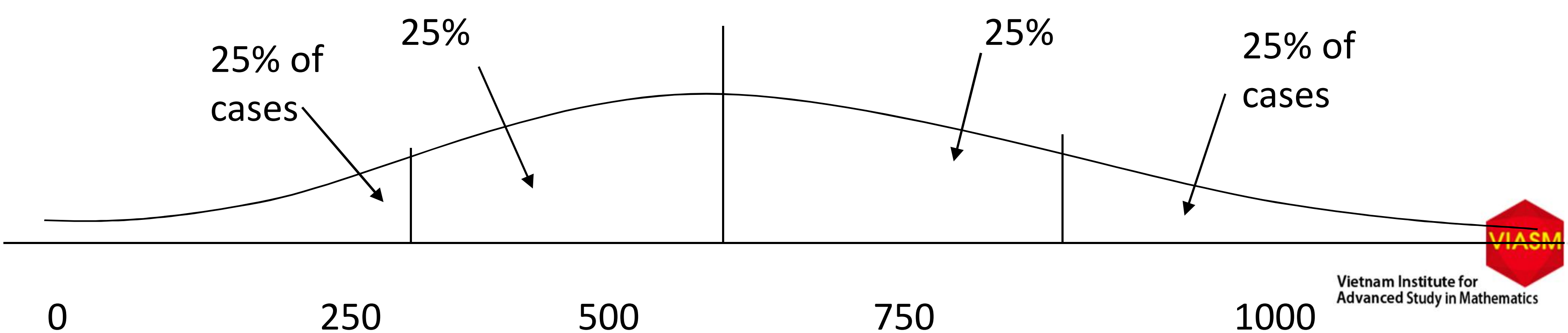
A quartile is the value that marks one of the divisions that breaks a series of values into four equal parts.

The median is a quartile and divides the cases in half.

25th percentile is a quartile that divides the first ¼ of cases from the latter ¾.

75th percentile is a quartile that divides the first ¾ of cases from the latter ¼.

The interquartile range is the distance or range between the 25th percentile and the 75th percentile. Below, what is the interquartile range?

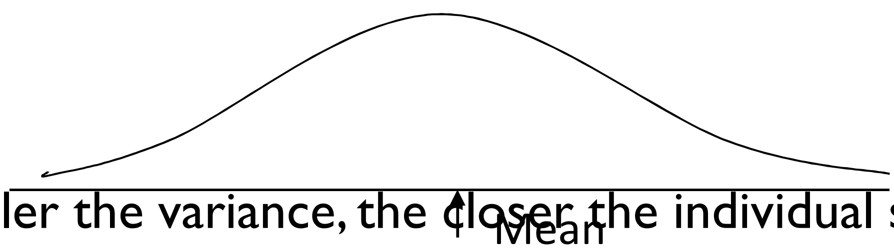


Vietnam Institute for Advanced Study in Mathematics

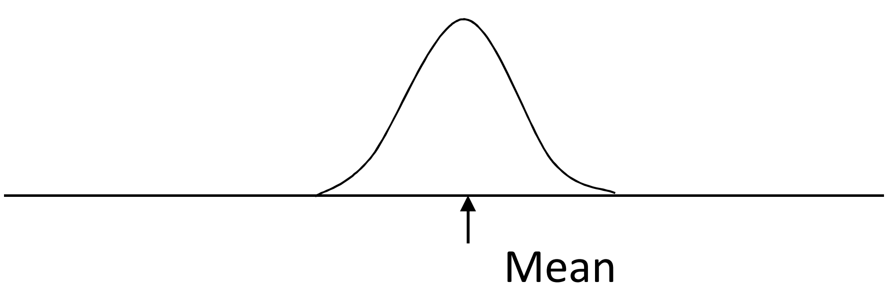
Variance

A measure of the spread of the recorded values on a variable. A measure of dispersion.

The larger the variance, the further the individual cases are from the mean.



The smaller the variance, the closer the individual scores are to the mean.



Variance

Variance is a number that at first seems complex to calculate.

Calculating variance starts with a "deviation."

A deviation is the distance away from the mean of a case's score.

Y_i – Y-bar

If the average person's car costs $20,000, my deviation from the mean is - $14,000!

6K - 20K = -14K

Variance

- We want to add these to get total deviations, but if we were to do that, we would get zero every time. Why?
- We need a way to eliminate negative signs.

Squaring the deviations will eliminate negative signs...

A Deviation Squared: $(Y_i - \bar{Y})^2$

Back to the IQ example,

A deviation squared for 102 is: of 115:

$(102 - 110.54)^2 = (-8.54)^2 = 72.93$ $(115 - 110.54)^2 = (4.46)^2 = 19.89$

Vietnam Institute for Advanced Study in Mathematics

Variance

If you were to add all the squared deviations together, you'd get what we call the "Sum of Squares."

Sum of Squares (SS) = $\Sigma\,(Y\mathit{i} - \text{Y-bar})^2$

$SS = (Y1 - \text{Y-bar})^2 + (Y2 - \text{Y-bar})^2 + \ldots + (Yn - \text{Y-bar})^2$

Variance

The last step…

The approximate average sum of squares is the variance.

SS/N = Variance for a population.

SS/n-1 = Variance for a sample.

Variance = $\Sigma(Y_i - \bar{Y})^2 / n - 1$

Variance

For Class A, Variance = 2825.39 / n - 1

= 2825.39 / 12 = 235.45

How helpful is that???

Vietnam Institute for Advanced Study in Mathematics

Standard Deviation

To convert variance into something of meaning, let's create standard deviation.

The square root of the variance reveals the average deviation of the observations from the mean.

$$\text{s.d.} = \sqrt{\frac{\Sigma(Y_i - \text{Y-bar})^2}{n - 1}}$$

VIASM
Vietnam Institute for
Advanced Study in Mathematics

Standard Deviation

For Class A, the standard deviation is:

$$\sqrt{235.45} = 15.34$$

The average of persons' deviation from the mean IQ of 110.54 is 15.34 IQ points.

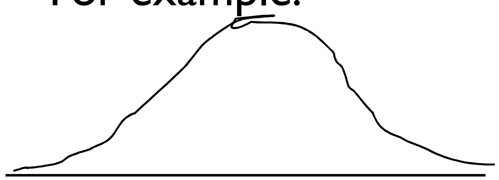
Review:

1. Deviation
2. Deviation squared
3. Sum of squares
4. Variance
5. Standard deviation

Standard Deviation

1. Larger s.d. = greater amounts of variation around the mean.

   For example:

   19 25 31

   $\bar{Y} = 25$

   s.d. = 3

   13 25 37

   $\bar{Y} = 25$

   s.d. = 6

2. s.d. = 0 only when all values are the same (only when you have a constant and not a "variable")
3. If you were to "rescale" a variable, the s.d. would change by the same magnitude—if we changed units above so the mean equaled 250, the s.d. on the left would be 30, and on the right, 60
4. Like the mean, the s.d. will be inflated by an outlier case value.

Descriptive Statistics

Summarizing Data:

- ✓ Central Tendency (or Groups' "Middle Values")
  - ✓ Mean
  - ✓ Median
  - ✓ Mode
- ✓ Variation (or Summary of Differences Within Groups)
  - ✓ Range
  - ✓ Interquartile Range
  - ✓ Variance
  - ✓ Standard Deviation
- ❑ …Wait! There's more

Vietnam Institute for Advanced Study in Mathematics

Box-Plots

A way to graphically portray almost all the descriptive statistics at once is the box-plot.

A box-plot shows: Upper and lower quartiles

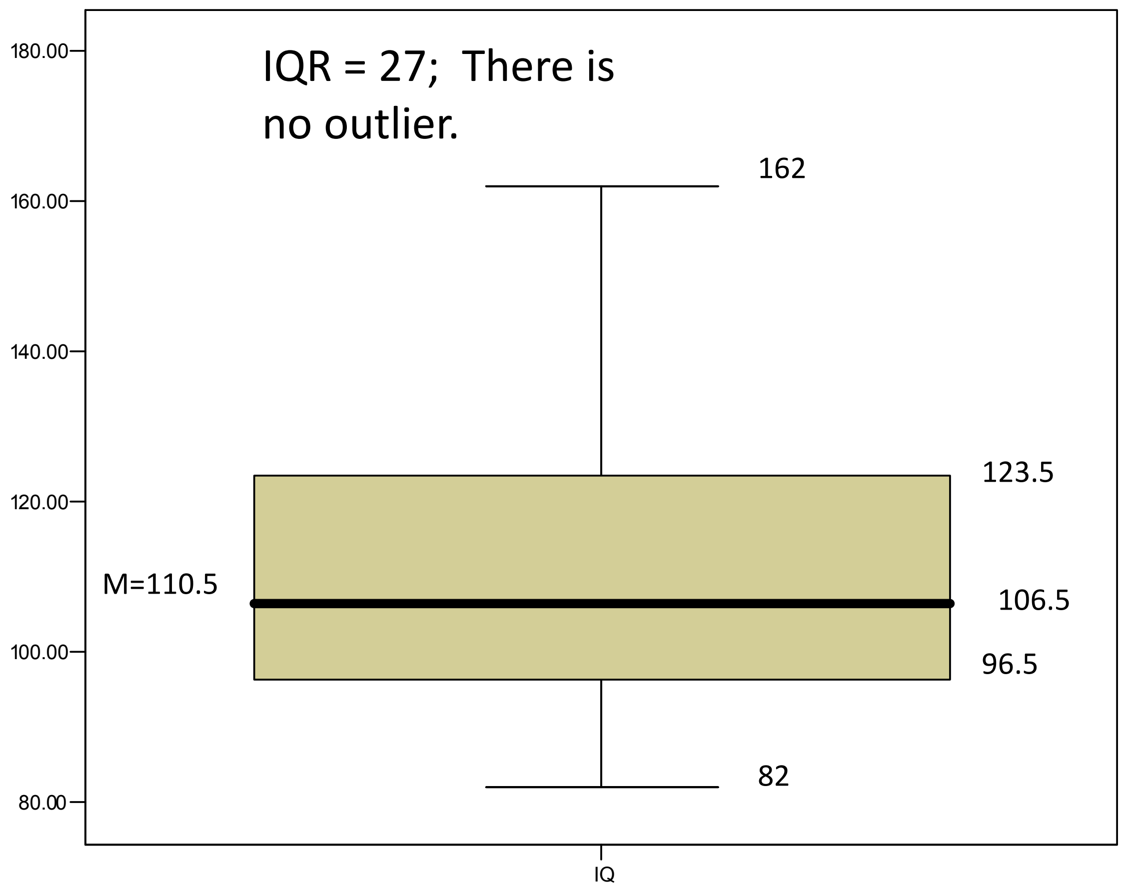
Mean

Median

Range

Outliers (1.5 IQR)

Box-Plots


IQR = 27; There is
no outlier.
162
123.5
M=110.5
106.5
96.5
82
180.00
160.00
140.00
120.00
100.00
80.00
IQ

Point Estimation

A point estimate of some population parameter θ is a single numerical value $\hat{\theta}$ of a statstistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the point estimator.

Vietnam Institute for Advanced Study in Mathematics

Sampling Distributions and the Central Limit Theorem

Statistical inference is concerned with making decisions about a population based on the information contained in a random sample from that population.

Definitions:

The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if (a) the X_i's are independent random variables, and (b) every X_i has the same probability distribution.

A **statistic** is any function of the observations in a random sample.

The probability distribution of a statistic is called a **sampling distribution.**

Sampling Distributions and the Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

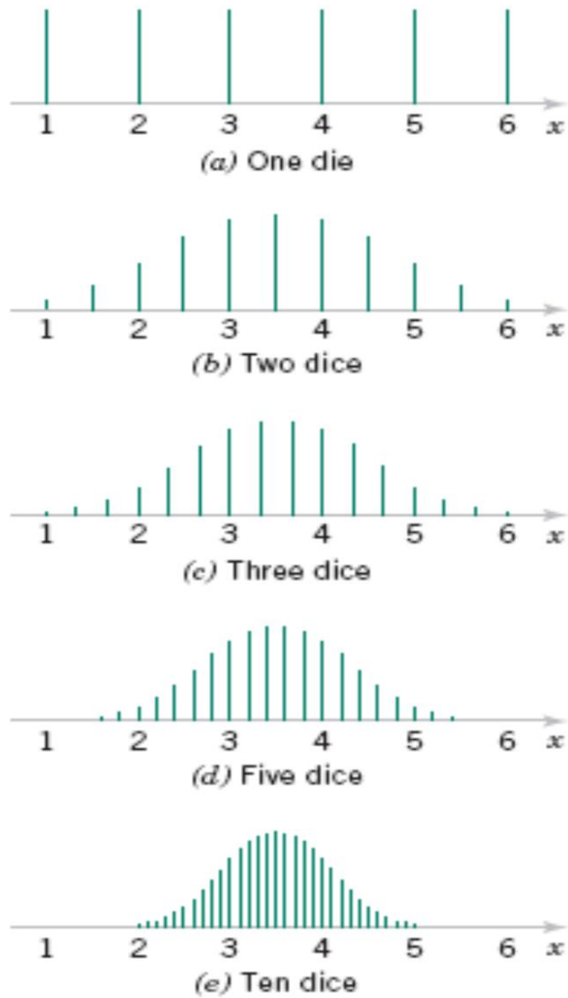
$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \qquad (7\text{-}1)$$

as $n \to \infty$, is the standard normal distribution.

Vietnam Institute for Advanced Study in Mathematics

Sampling Distributions and the Central Limit Theorem

Figure Distributions of average scores from throwing dice. [Adapted with permission from Box, Hunter, and Hunter (1978).]



Vietnam Institute for Advanced Study in Mathematics

Sampling Distributions and the Central Limit Theorem

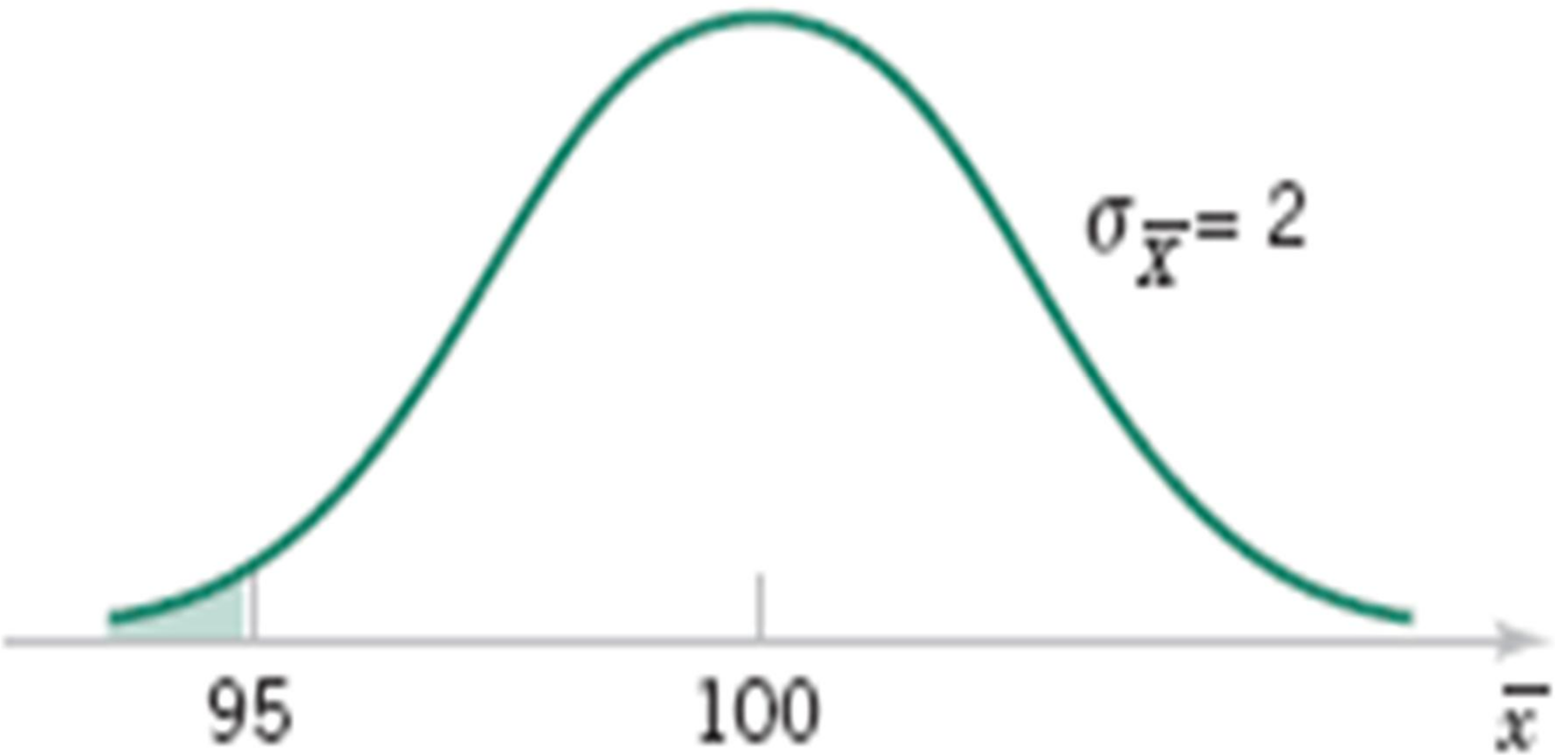


Figure 7-2 Probability for Example 7-1

Sampling Distributions and the Central Limit Theorem

Approximate Sampling Distribution of a Difference in Sample Means

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if $\overline{X}_1$ and $\overline{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \qquad (7\text{-}4)$$

is approximately standard normal, if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

General Concepts of Point Estimation

Unbiased Estimators

Definition

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$$E(\hat{\Theta}) = \theta \qquad (7\text{-}5)$$

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta}) - \theta \qquad (7\text{-}6)$$

is called the **bias** of the estimator $\hat{\Theta}$.

General Concepts of Point Estimation

Example

Suppose that X is a random variable with mean μ and variance σ^2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the population represented by X. Show that the sample mean $\overline{X}$ and sample variance S^2 are unbiased estimators of μ and σ^2, respectively.

First consider the sample mean. In Equation 5.40a in Chapter 5, we showed that $E(\overline{X}) = \mu$. Therefore, the sample mean $\overline{X}$ is an unbiased estimator of the population mean μ.

Now consider the sample variance. We have

$$E(S^2) = E\left[\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}\right] = \frac{1}{n-1} E \sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$= \frac{1}{n-1} E \sum_{i=1}^{n}(X_i^2 + \overline{X}^2 - 2\overline{X}X_i) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right]$$

General Concepts of Point Estimation

Example (continued)

The last equality follows from Equation 5-37 in Chapter 5. However, since $E(X_i^2) = \mu^2 + \sigma^2$ and $E(\overline{X}^2) = \mu^2 + \sigma^2/n$, we have

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2+\sigma^2) - n(\mu^2+\sigma^2/n)\right]$$

$$= \frac{1}{n-1}(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2)$$

$$= \sigma^2$$

Therefore, the sample variance S^2 is an unbiased estimator of the population variance σ^2.

Vietnam Institute for Advanced Study in Mathematics

General Concepts of Point Estimation

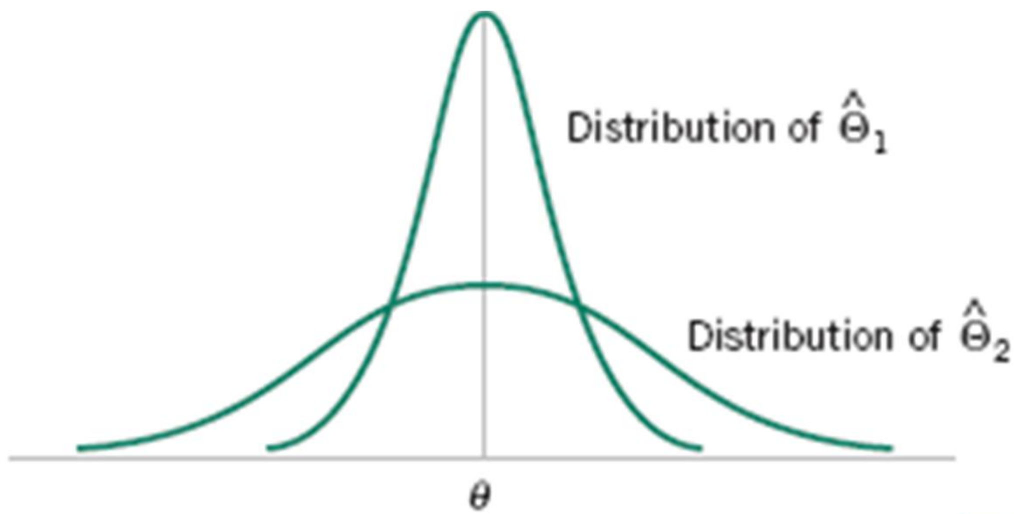
Variance of a Point Estimator

Definition

If we consider all unbiased estimators of θ, the one with the smallest variance is called the **minimum variance unbiased estimator** (MVUE).

Figure The sampling distributions of two unbiased estimators

$\hat{\Theta}_1$ and $\hat{\Theta}_2$.



General Concepts of Point Estimation

Variance of a Point Estimator

If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, the sample mean $\overline{X}$ is the MVUE for μ.

General Concepts of Point Estimation

Standard Error: Reporting a Point Estimate

Definition

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation, given by $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$. If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error**, denoted by $\hat{\sigma}_{\hat{\Theta}}$.

General Concepts of Point Estimation

Standard Error: Reporting a Point Estimate

Suppose we are sampling from a normal distribution with mean μ and variance σ^2. Now the distribution of $\overline{X}$ is normal with mean μ and variance σ^2/n, so the standard error of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

If we did not know σ but substituted the sample standard deviation S into the above equation, the estimated standard error of $\overline{X}$ would be

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$$

Vietnam Institute for Advanced Study in Mathematics

General Concepts of Point Estimation

Mean Square Error of an Estimator

Definition

The **mean squared error** of an estimator $\hat{\Theta}$ of the parameter θ is defined as

$$\text{MSE}(\hat{\Theta}) = E(\hat{\Theta} - \theta)^2 \qquad (7\text{-}7)$$

General Concepts of Point Estimation

Mean Square Error of an Estimator

The mean squared error is an important criterion for comparing two estimators. Let $\hat{\Theta}_1$ and $\hat{\Theta}_2$ be two estimators of the parameter θ, and let MSE $(\hat{\Theta}_1)$ and MSE $(\hat{\Theta}_2)$ be the mean squared errors of $\hat{\Theta}_1$ and $\hat{\Theta}_2$. Then the **relative efficiency** of $\hat{\Theta}_2$ to $\hat{\Theta}_1$ is defined as

$$\frac{\mathrm{MSE}(\hat{\Theta}_1)}{\mathrm{MSE}(\hat{\Theta}_2)} \tag{7-8}$$

If this relative efficiency is less than 1, we would conclude that $\hat{\Theta}_1$ is a more efficient estimator of θ than $\hat{\Theta}_2$, in the sense that it has a smaller mean square error.

Vietnam Institute for Advanced Study in Mathematics

General Concepts of Point Estimation

Mean Square Error of an Estimator

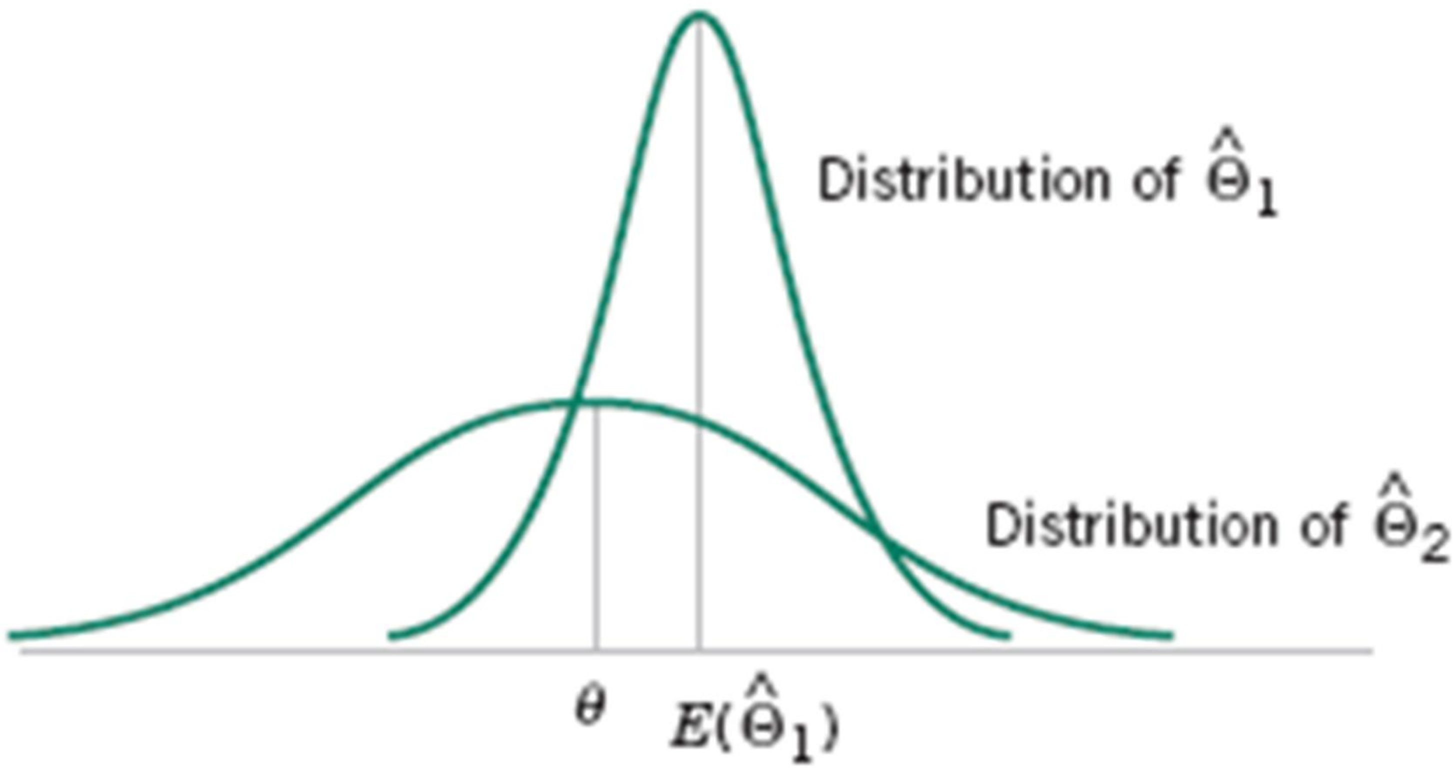


Figure A biased estimator $\hat{\Theta}_1$ that has smaller variance than the unbiased estimator $\hat{\Theta}_2$.

Vietnam Institute for Advanced Study in Mathematics

Methods of Point Estimation

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability distribution $f(x)$, where $f(x)$ can be a discrete probability mass function or a continuous probability density function. The kth **population moment** (or **distribution moment**) is $E(X^k)$, $k = 1, 2, \ldots$. The corresponding kth **sample moment** is $(1/n)\sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots$.

Definition

Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability mass function or probability density function with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The **moment estimators** $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters.

Methods of Point Estimation

Example

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with parameters μ and σ^2. For the normal distribution $E(X) = \mu$ and $E(X^2) = \mu^2 + \sigma^2$. Equating $E(X)$ to $\overline{X}$ and $E(X^2)$ to $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ gives

$$\mu = \overline{X}, \qquad \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Solving these equations gives the moment estimators

$$\hat{\mu} = \overline{X}, \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)^2}{n} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}$$

Notice that the moment estimator of σ^2 is not an unbiased estimator.

Methods of Point Estimation

Method of Maximum Likelihood

Definition

Suppose that X is a random variable with probability distribution $f(x; \theta)$, where θ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size n. Then the **likelihood function** of the sample is

$$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \cdots \cdot f(x_n; \theta) \qquad (7\text{-}9)$$

Note that the likelihood function is now a function of only the unknown parameter θ. The **maximum likelihood estimator (MLE)** of θ is the value of θ that maximizes the likelihood function $L(\theta)$.

Methods of Point Estimation

Example

Let X be a Bernoulli random variable. The probability mass function is

$$f(x; p) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

where p is the parameter to be estimated. The likelihood function of a random sample of size n is

$$\begin{aligned} L(p) &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

Methods of Point Estimation

Example (continued)

We observe that if $\hat{p}$ maximizes $L(p)$, $\hat{p}$ also maximizes $\ln L(p)$. Therefore,

$$\ln L(p) = \left(\sum_{i=1}^{n} x_i \right) \ln p + \left(n - \sum_{i=1}^{n} x_i \right) \ln (1 - p)$$

Now

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i \right)}{1 - p}$$

Equating this to zero and solving for p yields $\hat{p} = (1/n) \sum_{i=1}^{n} x_i$. Therefore, the maximum likelihood estimator of p is

$$\hat{P} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Methods of Point Estimation

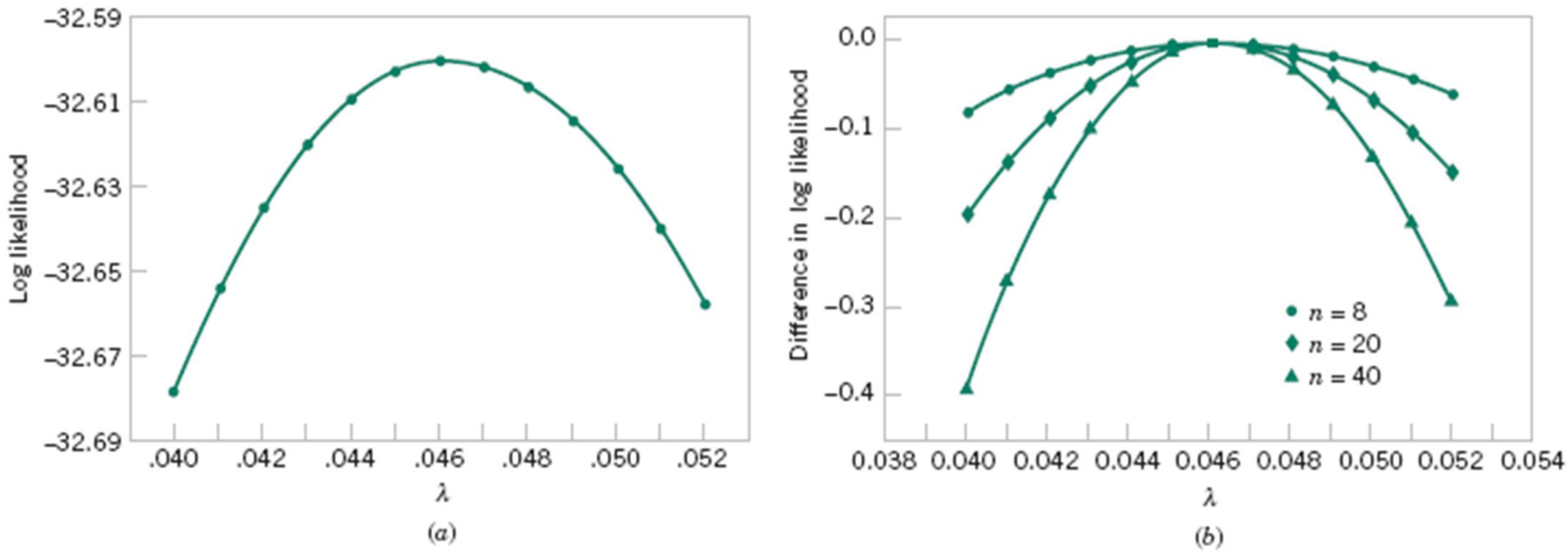


Figure Log likelihood for the exponential distribution, using the failure time data. (a) Log likelihood with n = 8 (original data). (b) Log likelihood if n = 8, 20, and 40.

Methods of Point Estimation

Example

Let X be normally distributed with mean μ and variance σ^2, where both μ and σ^2 are unknown. The likelihood function for a random sample of size n is

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma^2)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i-\mu)^2}$$

and

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Vietnam Institute for Advanced Study in Mathematics

Methods of Point Estimation

Example (continued)

Now

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial (\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

The solutions to the above equation yield the maximum likelihood estimators

$$\hat{\mu} = \overline{X} \qquad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Once again, the maximum likelihood estimators are equal to the moment estimators.

Methods of Point Estimation

Properties of the Maximum Likelihood Estimator

Under very general and not restrictive conditions, when the sample size n is large and if $\hat{\Theta}$ is the maximum likelihood estimator of the parameter θ,

(1) $\hat{\Theta}$ is an approximately unbiased estimator for θ $[E(\hat{\Theta}) \simeq \theta]$,

(2) the variance of $\hat{\Theta}$ is nearly as small as the variance that could be obtained with any other estimator, and

(3) $\hat{\Theta}$ has an approximate normal distribution.

Methods of Point Estimation

The Invariance Property

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$ be the maximum likelihood estimators of the parameters $\theta_1, \theta_2, \ldots, \theta_k$. Then the maximum likelihood estimator of any function $h(\theta_1, \theta_2, \ldots, \theta_k)$ of these parameters is the same function $h(\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k)$ of the estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$.

Methods of Point Estimation

Example

In the normal distribution case, the maximum likelihood estimators of μ and σ^2 were $\hat{\mu} = \overline{X}$ and $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/n$. To obtain the maximum likelihood estimator of the function $h(\mu, \sigma^2) = \sqrt{\sigma^2} = \sigma$, substitute the estimators $\hat{\mu}$ and $\hat{\sigma}^2$ into the function h, which yields

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]^{1/2}$$

Thus, the maximum likelihood estimator of the standard deviation σ is *not* the sample standard deviation S.

Vietnam Institute for Advanced Study in Mathematics

Methods of Point Estimation

Complications in Using Maximum Likelihood Estimation

- It is not always easy to maximize the likelihood function because the equation(s) obtained from $dL(\theta)/d\theta = 0$ may be difficult to solve.

- It may not always be possible to use calculus methods directly to determine the maximum of $L(\theta)$.

Methods of Point Estimation

Bayesian Estimation of Parameters

and the marginal distribution of $X_1, X_2, \ldots, X_n$ is

$$f(x_1, x_2, \ldots, x_n) = \begin{cases} \sum_{\theta} f(x_1, x_2, \ldots, x_n, \theta), & \theta \text{ discrete} \\ \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n, \theta)\, d\theta, & \theta \text{ continuous} \end{cases}$$

Therefore, the desired distribution is

$$f(\theta \mid x_1, x_2, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n, \theta)}{f(x_1, x_2, \ldots, x_n)}$$

We define the **Bayes estimator** of θ as the value $\tilde{\theta}$ that corresponds to the mean of the posterior distribution $f(\theta \mid x_1, x_2, \ldots, x_n)$.

Methods of Point Estimation

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution with mean μ and variance σ^2, where μ is unknown and σ^2 is known. Assume that the prior distribution for μ is normal with mean μ_0 and variance σ_0^2; that is

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(\mu-\mu_0)^2/(2\sigma_0^2)} = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(\mu^2-2\mu_0\mu+\mu_0^2)/(2\sigma_0^2)}$$

The joint probability distribution of the sample is

$$f(x_1, x_2, \ldots, x_n \mid \mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)(\Sigma x_i^2 - 2\mu\Sigma x_i + n\mu^2)}$$

Methods of Point Estimation

Example (Continued)

Thus, the joint probability distribution of the sample and μ is

$$f(x_1, x_2, \ldots, x_n, \mu) = \frac{1}{(2\pi\sigma^2)^{n/2}\sqrt{2\pi\sigma_0}} e^{-(1/2)[(1/\sigma_0^2 + n/\sigma^2)\mu^2 - (2\mu_0/\sigma_0^2 + 2\Sigma x_i/\sigma^2)\mu + \Sigma x_i^2/\sigma^2 + \mu_0^2/\sigma_0^2]}$$

$$= e^{-(1/2)\left[\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\mu^2 - 2\left(\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}}{\sigma^2/n}\right)\mu\right]} h_1(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

Upon completing the square in the exponent

$$f(x_1, x_2, \ldots, x_n, \mu) = e^{-(1/2)\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\left[\mu^2 - \left(\frac{(\sigma^2/n)\mu_0}{\sigma_0^2 + \sigma^2/n} + \frac{\bar{x}\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\right)\right]^2} h_2(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

where $h_i(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$ is a function of the observed values, σ^2, μ_0, and σ_0^2.

Vietnam Institute for Advanced Study in Mathematics

Methods of Point Estimation

Example (Continued)

Now, because $f(x_1, \ldots, x_n)$ does not depend on μ,

$$f(\mu \mid x_1, \ldots, x_n) = e^{-(1/2)\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}\right)\left[\mu^2-\left(\frac{(\sigma^2/n)\mu_0+\sigma_0^2\bar{x}}{\sigma_0^2+\sigma^2/n}\right)\right]} h_3(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

This is recognized as a normal probability density function with posterior mean

$$\frac{(\sigma^2/n)\mu_0 + \sigma_0^2\bar{x}}{\sigma_0^2 + \sigma^2/n}$$

and posterior variance

$$\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)^{-1} = \frac{\sigma_0^2(\sigma^2/n)}{\sigma_0^2 + \sigma^2/n}$$

Vietnam Institute for Advanced Study in Mathematics

Methods of Point Estimation

Example (Continued)

Consequently, the Bayes estimate of μ is a weighted average of μ_0 and $\bar{x}$. For purposes of comparison, note that the maximum likelihood estimate of μ is $\hat{\mu} = \bar{x}$.

To illustrate, suppose that we have a sample of size $n = 10$ from a normal distribution with unknown mean μ and variance $\sigma^2 = 4$. Assume that the prior distribution for μ is normal with mean $\mu_0 = 0$ and variance $\sigma_0^2 = 1$. If the sample mean is 0.75, the Bayes estimate of μ is

$$\frac{(4/10)0 + 1(0.75)}{1 + (4/10)} = \frac{0.75}{1.4} = 0.536$$

Note that the maximum likelihood estimate of μ is $\bar{x} = 0.75$.

Interval Estimation

Interval Estimation

- A point estimator cannot be expected to provide the exact value of the population parameter
- An interval estimate can be computed by adding and subtracting a margin of error to the point estimate.
- The purpose of an interval estimate is to provide information about how close the point estimate is to the value of the parameter.

Interval estimation

- X is a r.v whose distribution depends on unknown parameter θ
- Based on a random sample of X, find L and U s.t

$$P(L \leq \theta \leq U) = 1 - \alpha$$

- (L, U) is referred to as a C.I. (or two-sided C.I.) for θ
- L=- ∞ or U= ∞ we have one-sided C.I.s.

Interval Estimation

Interpreting a Confidence Interval

- The confidence interval is a random interval

- The appropriate interpretation of a confidence interval for a parameter is: The observed interval [*l*, *u*] brackets the true value of the parameter, with confidence $100(1-\alpha)$.

Vietnam Institute for Advanced Study in Mathematics

Confidence Interval

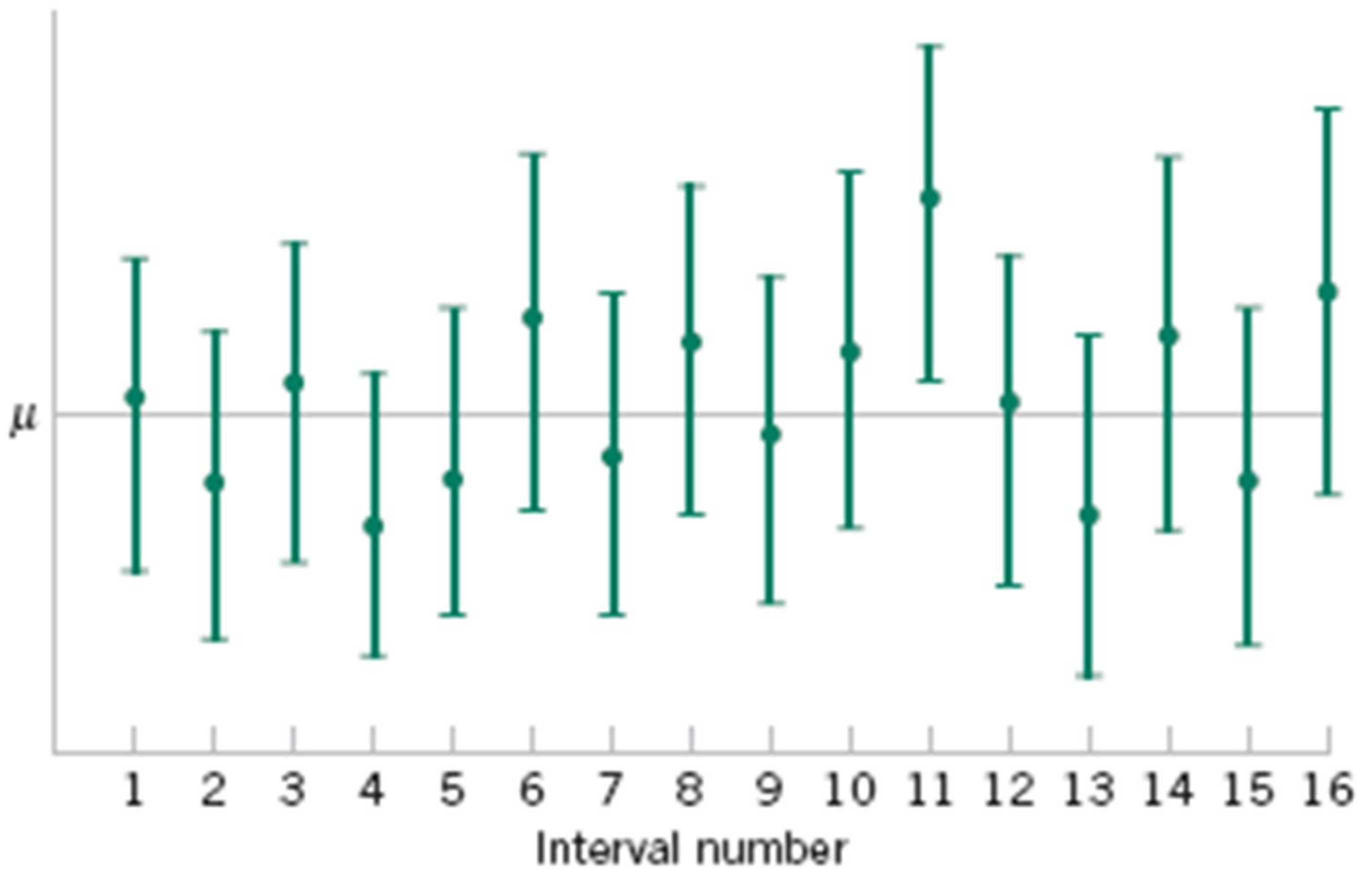


Figure Repeated construction of a confidence interval for μ.

Vietnam Institute for Advanced Study in Mathematics

Hypothesis Testing

Hypothesis Testing

Statistical Hypotheses

Statistical hypothesis testing and confidence interval estimation of parameters are the fundamental methods used at the data analysis stage of a **comparative experiment,** in which the engineer is interested, for example, in comparing the mean of a population to a specified value.

Definition

A **statistical hypothesis** is a statement about the parameters of one or more populations.

Hypothesis Testing

Statistical Hypotheses

For example, suppose that we are interested in the burning rate of a solid propellant used to power aircrew escape systems.

- Now burning rate is a random variable that can be described by a probability distribution.
- Suppose that our interest focuses on the **mean** burning rate (a parameter of this distribution).
- Specifically, we are interested in deciding whether or not the mean burning rate is 50 centimeters per second.

Hypothesis Testing

Statistical Hypotheses

Two-sided Alternative Hypothesis

H_0: $\mu = 50$ centimeters per second — null hypothesis

H_1: $\mu \neq 50$ centimeters per second — alternative hypothesis

One-sided Alternative Hypotheses

H_0: $\mu = 50$ centimeters per second

H_1: $\mu < 50$ centimeters per second

or

H_0: $\mu = 50$ centimeters per second

H_1: $\mu > 50$ centimeters per second

Hypothesis Testing

Statistical Hypotheses

Test of a Hypothesis

- A procedure leading to a decision about a particular hypothesis

- Hypothesis-testing procedures rely on using the information in a **random sample from the population of interest**.

- If this information is *consistent* with the hypothesis, then we will conclude that the hypothesis is **true**; if this information is *inconsistent* with the hypothesis, we will conclude that the hypothesis is **false**.

Hypothesis Testing

Tests of Statistical Hypotheses

$$H_0\colon \mu = 50 \text{ centimeters per second}$$
$$H_1\colon \mu \neq 50 \text{ centimeters per second}$$

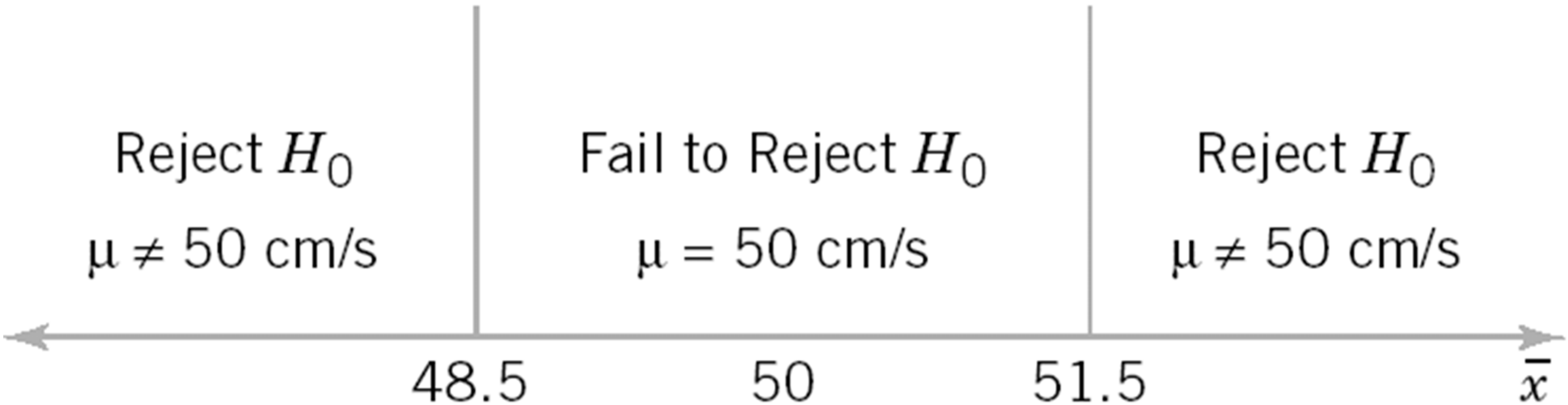


Figure 9-1 Decision criteria for testing H_0:μ = 50 centimeters per second versus H_1:μ $\neq$ 50 centimeters per second.

Hypothesis Testing

Tests of Statistical Hypotheses

Definitions

Rejecting the null hypothesis H_0 when it is true is defined as a **type I error.**

Failing to reject the null hypothesis when it is false is defined as a **type II error.**

VIASM
Vietnam Institute for
Advanced Study in Mathematics

Hypothesis Testing

Tests of Statistical Hypotheses

Table 9-1 Decisions in Hypothesis Testing

Decision	H_0 Is True	H_0 Is False
Fail to reject H_0	no error	type II error
Reject H_0	type I error	no error

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

Sometimes the type I error probability is called the **significance level**, or the **α-error**, or the **size** of the test.

Hypothesis Testing

Tests of Statistical Hypotheses

$$\alpha = P(\bar{X} < 48.5 \text{ when } \mu = 50) + P(\bar{X} > 51.5 \text{ when } \mu = 50)$$

The z-values that correspond to the critical values 48.5 and 51.5 are

$$z_1 = \frac{48.5 - 50}{0.79} = -1.90 \qquad \text{and} \qquad z_2 = \frac{51.5 - 50}{0.79} = 1.90$$

Therefore

$$\alpha = P(Z < -1.90) + P(Z > 1.90) = 0.028717 + 0.028717 = 0.057434$$

Hypothesis Testing

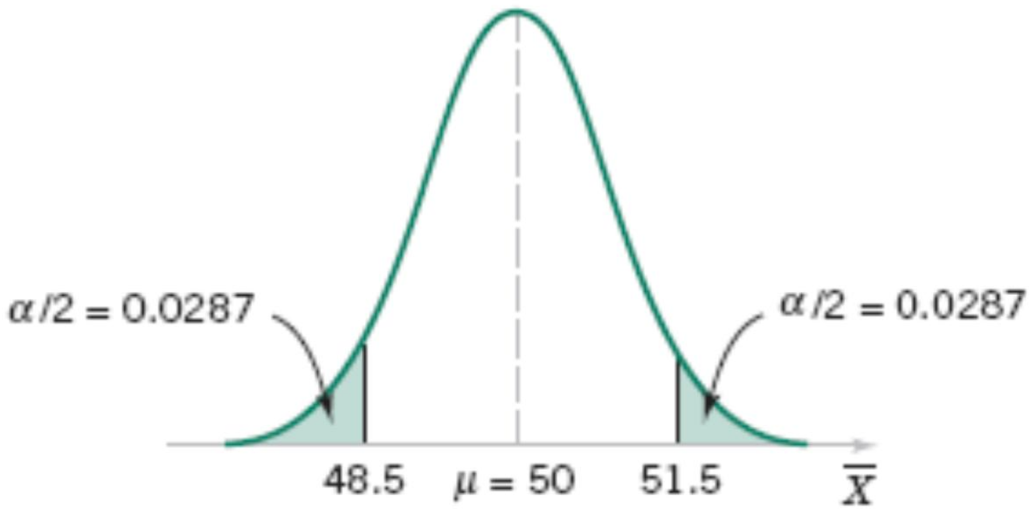


Figure 9-2 The critical region for H_0: $\mu = 50$ versus H_1: $\mu \neq 50$ and $n = 10$.

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \qquad (9\text{-}3)$$

Hypothesis Testing

$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false}) \qquad (9\text{-}4)$$

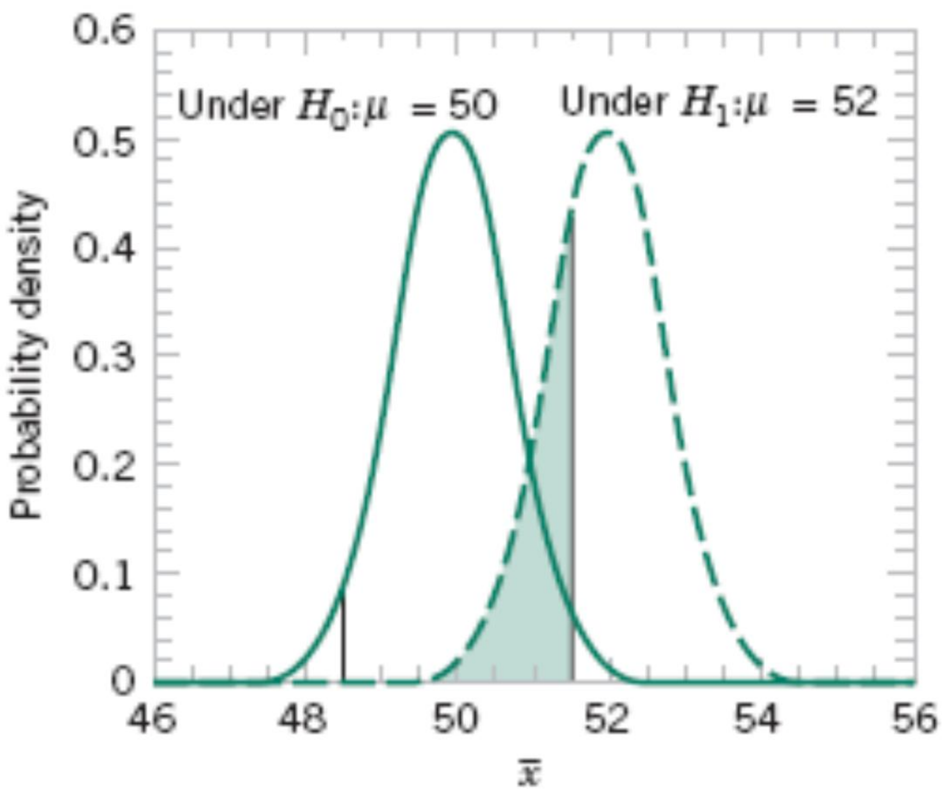


Figure 9-3 The probability of type II error when μ = 52 and n = 10.

Hypothesis Testing

$$\beta = P(48.5 \leq \overline{X} \leq 51.5 \text{ when } \mu = 52)$$

The z-values corresponding to 48.5 and 51.5 when $\mu = 52$ are

$$z_1 = \frac{48.5 - 52}{0.79} = -4.43 \qquad \text{and} \qquad z_2 = \frac{51.5 - 52}{0.79} = -0.63$$

Therefore

$$\begin{aligned} \beta &= P(-4.43 \leq Z \leq -0.63) = P(Z \leq -0.63) - P(Z \leq -4.43) \\ &= 0.2643 - 0.0000 = 0.2643 \end{aligned}$$

Vietnam Institute for Advanced Study in Mathematics

Hypothesis Testing

$$\beta = P(48.5 \leq \overline{X} \leq 51.5 \text{ when } \mu = 50.5)$$

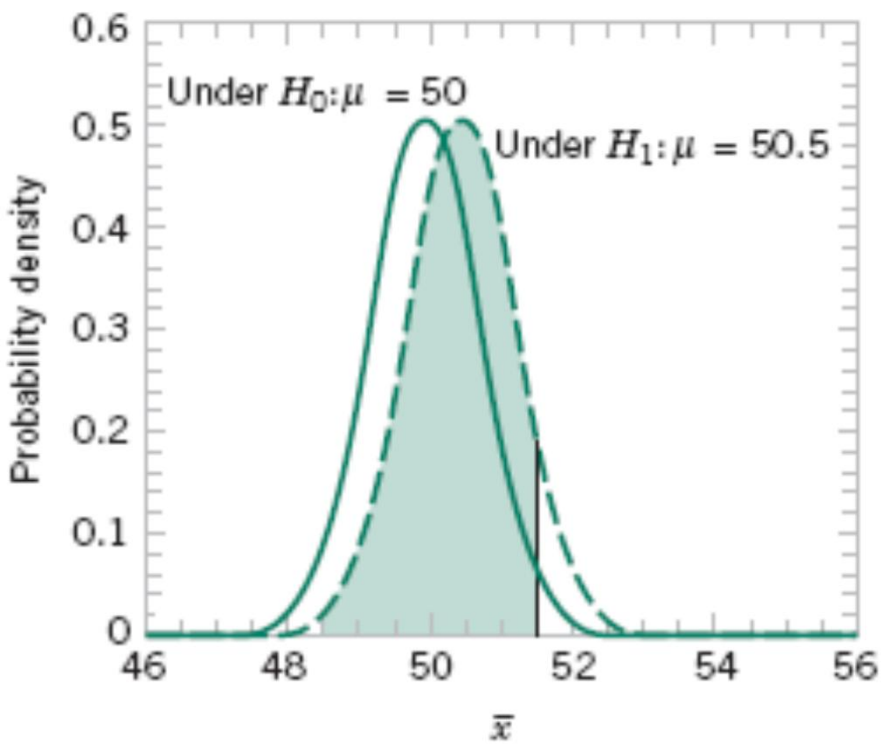


Figure 9-4 The probability of type II error when μ = 50.5 and n = 10.

Vietnam Institute for Advanced Study in Mathematics

Hypothesis Testing

$$\beta = P(48.5 \le \overline{X} \le 51.5 \text{ when } \mu = 50.5)$$

As shown in Fig. 9-4, the z-values corresponding to 48.5 and 51.5 when $\mu = 50.5$ are

$$z_1 = \frac{48.5 - 50.5}{0.79} = -2.53 \qquad \text{and} \qquad z_2 = \frac{51.5 - 50.5}{0.79} = 1.27$$

Therefore

$$\begin{aligned}\beta &= P(-2.53 \le Z \le 1.27) = P(Z \le 1.27) - P(Z \le -2.53) \\ &= 0.8980 - 0.0057 = 0.8923\end{aligned}$$

Vietnam Institute for Advanced Study in Mathematics

Hypothesis Testing

$$\beta = P(48.5 \le \overline{X} \le 51.5 \text{ when } \mu = 52)$$

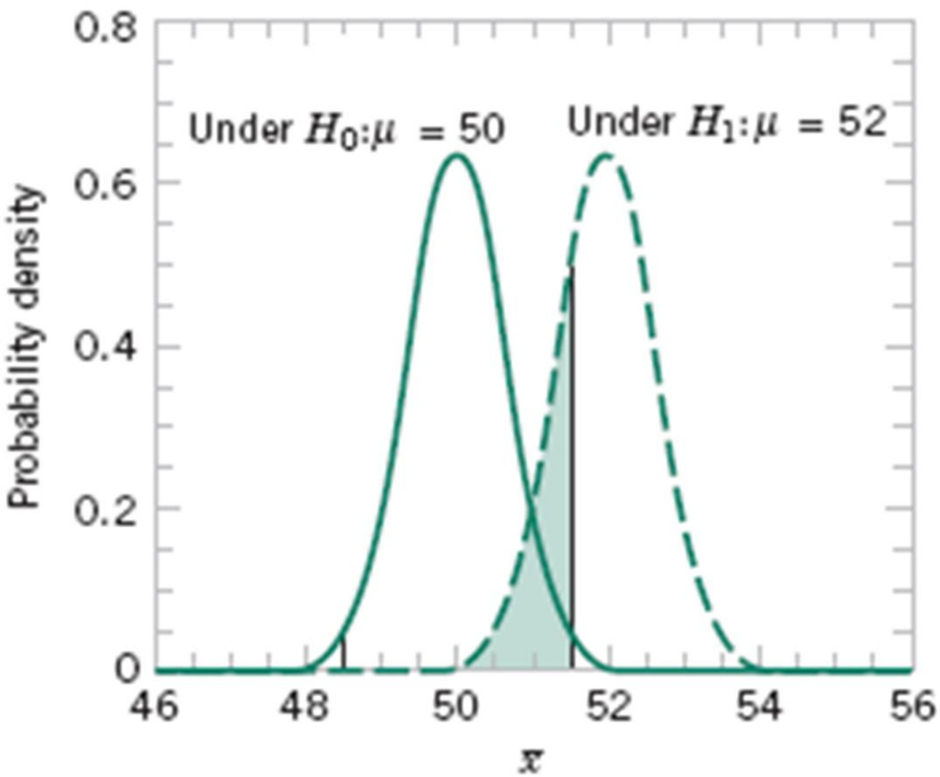


Figure The probability of type II error when μ = 2 and n = 16.

Hypothesis Testing

$$\beta = P(48.5 \leq \overline{X} \leq 51.5 \text{ when } \mu = 52)$$

When $n = 16$, the standard deviation of $\overline{X}$ is $\sigma/\sqrt{n} = 2.5/\sqrt{16} = 0.625$, and the z-values corresponding to 48.5 and 51.5 when $\mu = 52$ are

$$z_1 = \frac{48.5 - 52}{0.625} = -5.60 \quad \text{and} \quad z_2 = \frac{51.5 - 52}{0.625} = -0.80$$

Therefore

$$\begin{aligned}\beta &= P(-5.60 \leq Z \leq -0.80) = P(Z \leq -0.80) - P(Z \leq -5.60)\\ &= 0.2119 - 0.0000 = 0.2119\end{aligned}$$

Hypothesis Testing

Acceptance Region	Sample Size	α	β at $\mu = 52$	β at $\mu = 50.5$
$48.5 < \bar{x} < 51.5$	10	0.0576	0.2643	0.8923
$48 < \bar{x} < 52$	10	0.0114	$\boxed{0.5000}$	$\boxed{0.9705}$
$48.5 < \bar{x} < 51.5$	16	0.0164	0.2119	$\boxed{0.9445}$
$48 < \bar{x} < 52$	16	$\boxed{0.0014}$	$\boxed{0.5000}$	$\boxed{0.9918}$

Hypothesis Testing

Definition

The **power** of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true.

- The power is computed as 1 - β, and power can be interpreted as *the probability of correctly rejecting a false null hypothesis.* We often compare statistical tests by comparing their power properties.

- For example, consider the propellant burning rate problem when we are testing $H_0 : \mu = 50$ centimeters per second against $H_1 : \mu$ not equal 50 centimeters per second . Suppose that the true value of the mean is $\mu = 52$. When $n = 10$, we found that $\beta = 0.2643$, so the power of this test is $1 - \beta = 1 - 0.2643 = 0.7357$ when $\mu = 52$.

Hypothesis Testing

One-Sided and Two-Sided Hypotheses

Two-Sided Test:

$$H_0\colon \mu = \mu_0$$
$$H_1\colon \mu \neq \mu_0$$

One-Sided Tests:

$$H_0\colon \mu = \mu_0 \qquad\qquad\qquad H_0\colon \mu = \mu_0$$
$$\text{or}$$
$$H_1\colon \mu > \mu_0 \qquad\qquad\qquad H_1\colon \mu < \mu_0$$

Hypothesis Testing

P-Values in Hypothesis Tests

Definition

The ***P*-value** is the smallest level of significance that would lead to rejection of the null hypothesis H_0 with the given data.

Hypothesis Testing

P-Values in Hypothesis Tests

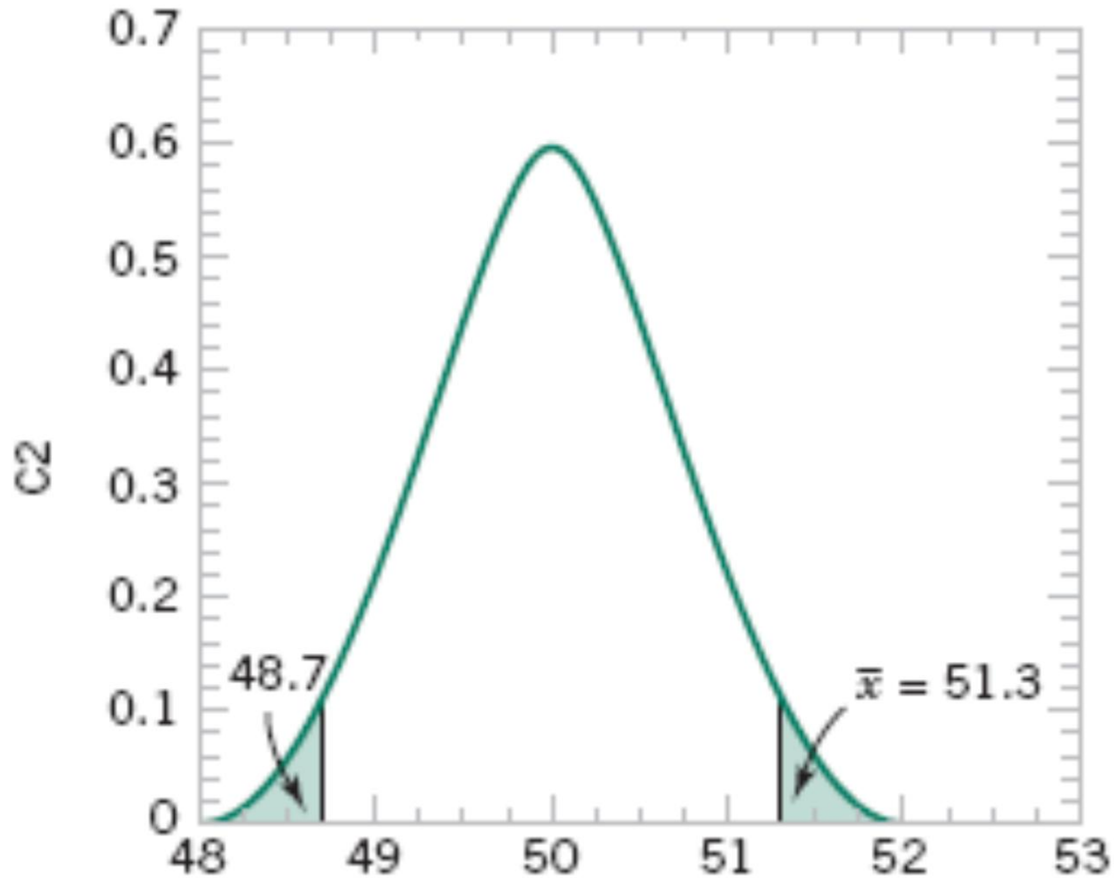


Figure 9-6 *P*-value is area of shaded region when $\bar{x} = 51.3$.

Hypothesis Testing

Connection between Hypothesis Tests and Confidence Intervals

There is a close relationship between the test of a hypothesis about any parameter, say θ, and the confidence interval for θ. If $[l, u]$ is a $100(1 - \alpha)\%$ confidence interval for the parameter θ, the test of size α of the hypothesis

$$H_0: \theta = \theta_0$$
$$H_1: \theta \neq \theta_0$$

will lead to rejection of H_0 if and only if θ_0 is **not** in the $100(1 - \alpha)\%$ CI $[l, u]$. As an illustration, consider the escape system propellant problem with $\bar{x} = 51.3$, $\sigma = 2.5$, and $n = 16$. The null hypothesis H_0: $\mu = 50$ was rejected, using $\alpha = 0.05$. The 95% two-sided CI on μ can be calculated using Equation 8-7. This CI is $51.3 \pm 1.96(2.5/\sqrt{16})$ and this is $50.075 \leq \mu \leq 52.525$. Because the value $\mu_0 = 50$ is not included in this interval, the null hypothesis H_0: $\mu = 50$ is rejected.

Vietnam Institute for Advanced Study in Mathematics