Econ 301: Econometrics

Alpay Filiztekin

Sabanci University

Review of Probability

Definition of Probability

Basic Definitions

- The mutually exclusive potential results of a random process are called the **outcomes**.
- The set of all possible outcomes of a given experiment is called the **sample space**.
- An **event** is a subset of the sample space.

Basic Definitions

- The mutually exclusive potential results of a random process are called the **outcomes**.
- The set of all possible outcomes of a given experiment is called the **sample space**.
- An **event** is a subset of the sample space.

Basic Definitions

- The mutually exclusive potential results of a random process are called the **outcomes**.
- The set of all possible outcomes of a given experiment is called the **sample space**.
- An **event** is a subset of the sample space.

Classical Definition of Probability

Let S be the finite sample space for an experiment having $N(S)$ equally likely outcomes, and let $A \subset S$ be an event containing $N(A)$ elements. Then the probability of the event A, denoted by $P(A)$, is given by $P(A) = \frac{N(A)}{N(S)}$.

Relativist Definition of Probability

Let n be the number of times that an experiment is repeatedly performed under identical conditions. Let A be the event in the sample space S, and define n_A to be the number of times in n repetitions of the experiment that the event A occurs. Then the probability of the event A is given by the limit of the relative frequency n_A / n, as $P(A) = \lim_{n \to \infty} \frac{n_A}{n}$.

Subjective Probability

The subjective probability of an event A is a real number, $P(A)$, in the interval $[0, 1]$, chosen to express the degree of personal belief in the likelihood of occurrence or validity of event A, the number 1 being associated with certainty.

Axiomatic Definition of Probability

- The set of all events in the sample space S is called the event space.
- For any event $A \subset S$, $P(A) \geq 0$.
- $P(S) = 1$.
- $P(\emptyset) = 0$.
- The sum of probabilities of all disjoint events is one.

Axiomatic Definition of Probability

- The set of all events in the sample space S is called the event space.
- For any event $A \subset S$, $P(A) \geq 0$.
- $P(S) = 1$.
- $P(\emptyset) = 0$.
- The sum of probabilities of all disjoint events is one.

Axiomatic Definition of Probability

- The set of all events in the sample space S is called the event space.
- For any event $A \subset S$, $P(A) \geq 0$.
- $P(S) = 1$.
- $P(\emptyset) = 0$.
- The sum of probabilities of all disjoint events is one.

Axiomatic Definition of Probability

- The set of all events in the sample space S is called the event space.
- For any event $A \subset S$, $P(A) \geq 0$.
- $P(S) = 1$.
- $P(\emptyset) = 0$.
- The sum of probabilities of all disjoint events is one.

Axiomatic Definition of Probability

- The set of all events in the sample space S is called the event space.
- For any event $A \subset S$, $P(A) \geq 0$.
- $P(S) = 1$.
- $P(\emptyset) = 0$.
- The sum of probabilities of all disjoint events is one.

Random Variables and Distributions

Random Variables

A **random variable** is a numerical summary of a random outcome.

- A ***discrete*** random variable takes only discrete set of values;
- A ***continuous*** random variable takes on a continuum of possible values.

Random Variables

A **random variable** is a numerical summary of a random outcome.

- A ***discrete*** random variable takes only discrete set of values;
- A ***continuous*** random variable takes on a continuum of possible values.

Random Variables

A **random variable** is a numerical summary of a random outcome.

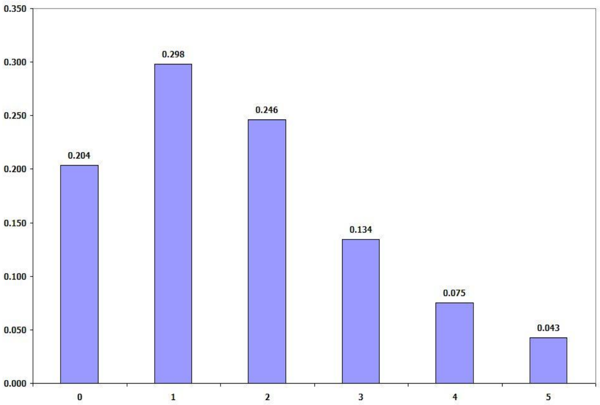
- A ***discrete*** random variable takes only discrete set of values;
- A ***continuous*** random variable takes on a continuum of possible values.

Probability Distributions

- The **probability distribution** of a discrete random variable is the list of all possible values of the variable and the probability that each value will occur.
- The **cumulative probability distribution** is the probability that the random variable is less than or equal to a particular value.

Probability Distributions

The probability that a home team will score x number of goals:



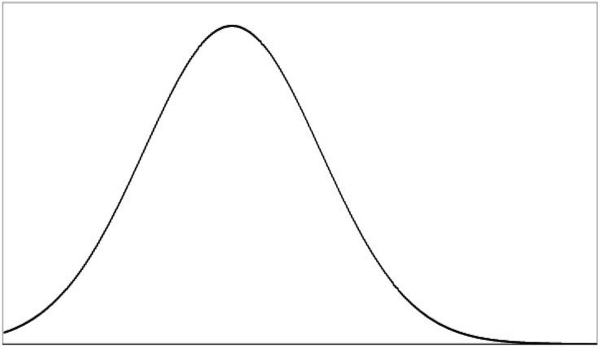
Probability Density Function

The **probability density function** (p.d.f) of a continuous random variable is the summary of probabilities of a continuum of events.

The area under the p.d.f. Between any two points is the probability that the random variable falls between those two points.

Probability Density Function

The probability of wages



Expected Value and Variance

Expected Value

The **expected value** of a random variable, Y, denoted $E(Y)$ or μ_Y, the probability-weighted average of the possible outcomes of the random variable.

- Discrete Case: $E(Y) = \sum_{i=1}^{n} y_i p_i$
- Continuous Case: $E(Y) = \int_a^b yf(y)dy$

Variance

The **variance** of a random variable, Y, denoted σ_Y^2, is $var(Y) = E[(Y - \mu_Y)^2]$.

- Discrete Case: $var(Y) = \sum_{i=1}^{n} (Y - \mu_Y)^2 p_i$
- Continuous Case: $E(Y) = \int_a^b (Y - \mu_Y)^2 f(y) dy$

The **standard deviation** is σ_Y.

Variance

H	Prob.	H*Prob	$(\mathbf{H} - \mu_{\mathbf{H}})^2$	$(\mathbf{H} - \mu_{\mathbf{H}})^2 * \mathbf{Prob}$
0	0.204	0.00	2.91	0.59
1	0.298	0.30	0.50	0.15
2	0.246	0.49	0.09	0.02
3	0.134	0.40	1.67	0.22
4	0.075	0.30	5.26	0.40
5	0.043	0.21	10.85	0.46
	μ_H	**1.71**	σ_H^2	**1.84**
			σ_H	**1.36**

Home team goals are denoted by H.

Multiple Random Variables

Joint and Marginal Distributions

The **joint probability distribution** of two discrete random variables, X and Y, is the probability that the random variables simultaneously take on certain values, x and y:

$$P(X = x, Y = y).$$

The **marginal probability distribution** of Y is probability distribution over all possible values of X:

$$P(Y = y) = \sum_{i=1}^{n_X} P(X = x_i, Y = y)$$

Joint and Marginal Distributions

	Away team goals (A)						
H	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
0	295	224	125	64	27	21	**756**
1	343	389	233	94	30	19	**1108**
2	295	335	159	75	29	22	**915**
3	163	188	88	32	14	13	**498**
4	95	98	61	19	4	2	**279**
5	59	52	36	8	2	1	**158**
Total	**1250**	**1286**	**702**	**292**	**106**	**78**	**3714**

Home team goals are denoted by H, away team goals are denoted by A.

Joint and Marginal Distributions

	Away team goals (A)						
H	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
0	7.9	6.0	3.4	1.7	0.7	0.6	**20.4**
1	9.2	10.5	6.3	2.5	0.8	0.5	**29.8**
2	7.9	9.0	4.3	2.0	0.8	0.6	**24.6**
3	4.4	5.1	2.4	0.9	0.4	0.4	**13.4**
4	2.6	2.6	1.6	0.5	0.1	0.1	**7.5**
5	1.6	1.4	1.0	0.2	0.1	0.0	**4.3**
Total	**33.7**	**34.6**	**18.9**	**7.9**	**2.9**	**2.1**	**100.0**

Home team goals are denoted by H, away team goals are denoted by A.

Conditional Distributions

The **conditional distribution** of a random variables is the distribution of that random variable conditional on the other taking on a specific value:

$$P(Y = y|X = x).$$

It can also be expressed as:

$$P(Y = y|X = x) = \frac{P(Y=y,X=x)}{P(X=x)}$$

Conditional Distributions

The **conditional expectation** of Y given X, is also called the conditional mean of Y given X, is the mean of the conditional distribution of Y given X:

$$E(Y|X = x) = \sum_{i=1}^{n_Y} y_i P(Y = y_i|X = x).$$

Conditional Distributions

What is the probability of away team scoring y goals given that home team scores x goals?

	Away team goals (A)						
H	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
0	39.0	29.6	16.5	8.5	3.6	2.8	**100.0**
1	31.0	35.1	21.0	8.5	2.7	1.7	**100.0**
2	32.2	36.6	17.4	8.2	3.2	2.4	**100.0**
3	32.7	37.8	17.7	6.4	2.8	2.6	**100.0**
4	34.1	35.1	21.9	6.8	1.4	0.7	**100.0**
5	37.3	32.9	22.8	5.1	1.3	0.6	**100.0**

Home team goals are denoted by H, away team goals are denoted by A.

Law of Iterated Expectations

The **law of iterated expectations** says:

$$E(Y) = \sum_{i=1}^{n_X} E(Y|X = x_i)P(X = x_i)$$

or

$$\text{E(Y)} = E_X[E_{Y|X}(Y|X)].$$

Independence

Two random variables X and Y are **independent**, if knowing the value of one of the variables provides no information about the other.

Specifically, X and Y are independent if the conditional distribution of Y given X equals the marginal distribution of Y.

Independence

X and Y are independent ***if and only if***

$$P(Y = y|X = x) = P(Y = y)$$

or

$$P(Y = y, X = x) = P(X = x) * P(Y = y).$$

Conditional Variance

The **conditional variance** is defined as:

$$var(Y|X) = \sum_{i=1}^{n_Y} [Y - E(Y|X = x)]^2 P(Y = y_i|X = x)$$

Conditional Variance

The **covariance** between X and Y, denoted σ_{XY}, is:

$$cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

$$= \sum_{j=1}^{n_X}\sum_{i=1}^{n_Y}(x_j-\mu_X)(y_i-\mu_Y)P(X=x_j, Y=y_i)$$

The **correlation** is $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$

Conditional Variance

Notice that

independence implies zero correlation

but

zero correlation does not imply independence.

Some Special Distributions

Normal Distribution

Normal Distribution: $N(\mu, \sigma^2)$

Standard Normal Distribution: $N(0, 1)$.

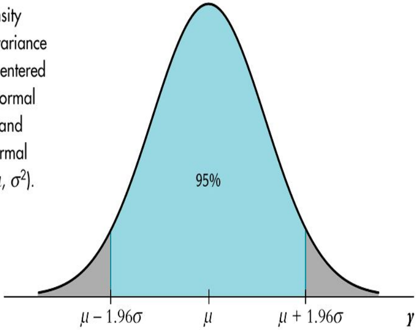
If $Y \sim N(\mu, \sigma^2)$, then $X \sim N(0, 1)$ where $X = (Y - \mu)/\sigma$.

Multivariate Normal Distribution: $N(\mu, \Sigma)$

Normal Distribution

FIGURE 2.3 The Normal Probability Density

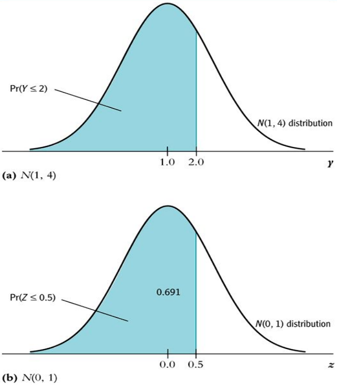
The normal probability density function with mean μ and variance σ^2 is a bell-shaped curve, centered at μ. The area under the normal p.d.f. between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 0.95. The normal distribution is denoted $N(\mu, \sigma^2)$.



Normal Distribution

FIGURE 2.4 **Calculating the Probability that $Y \leq 2$ When Y is Distributed $N(1, 4)$**

To calculate $\Pr(Y \leq 2)$, standardize Y, then use the standard normal distribution table. Y is standardized by subtracting its mean ($\mu = 1$) and dividing by its standard deviation ($\sigma_Y = 2$). The probability that $Y \leq 2$ is shown in Figure 2.4a, and the corresponding probability after standardizing Y is shown in Figure 2.4b. Because the standardized random variable, $\frac{Y-1}{2}$, is a standard normal (Z) random variable, $\Pr(Y \leq 2) = \Pr\left(\frac{Y-1}{2} \leq \frac{2-1}{2}\right) = \Pr(Z \leq 0.5)$. From Appendix Table 1, $\Pr(Z \leq 0.5) = 0.691$.



Normal Distribution

Some Properties of Normal Distributions

- Linear functions of normals are normal.
- Marginal distributions of multivariate normals are normal.
- Conditional distributions of normals are normal.

Other Special Distribution

- Chi-squared Distribution, χ^2:
  - It is the distribution of the sum of m squared independent standard normal variables.
- F-Distribution, $F_{m,\infty}$:
  - It is the distribution of a random variable with chi-squared distribution with m degrees of freedom, divided by m. Or it is the distribution of the average of m squared standard normal random variables.

Other Special Distribution

- F-Distribution: $F_{m,n}$:
  - Let W_1 and W_2 be two independent random variables with chi-squared distributions with respective degrees of freedom m and n. The the random variable $F = (W_1/m)/(W_2/n)$ has an F-distribution with degrees of freedom of m and n.
- Student t-distribution: , t_m:
  - It is defined the be the ratio of a standard normal random variable, divided by the square root of an independently distributed chi-squared random variable with m degrees of freedom.

Review of Statistics

Random Sampling

Random Sample

- The **population** or the **statistical universe** is the totality of elements about which some information is desired.
- A **sample** is a set of observations drawn from a probability distribution.
- A **random sample** of n elements is a sample that has the property that every combination of n elements has an equal chance of being in the selected sample.

Random Sample

- The **population** or the **statistical universe** is the totality of elements about which some information is desired.
- A **sample** is a set of observations drawn from a probability distribution.
- A **random sample** of n elements is a sample that has the property that every combination of n elements has an equal chance of being in the selected sample.

Random Sample

- The **population** or the **statistical universe** is the totality of elements about which some information is desired.
- A **sample** is a set of observations drawn from a probability distribution.
- A **random sample** of n elements is a sample that has the property that every combination of n elements has an equal chance of being in the selected sample.

Sampling Distribution

- The notion of **sampling distribution** arises from the fact that observations on a random variable are themselves random variables.
- A **sampling distribution** is the joint distribution of random variables comprising a sample, or random variables that are functions of such observations.

Sampling Distribution

- The notion of **sampling distribution** arises from the fact that observations on a random variable are themselves random variables.
- A **sampling distribution** is the joint distribution of random variables comprising a sample, or random variables that are functions of such observations.

Random Sample

- Let $Y_1, ..., Y_n$ denote observations that are randomly chosen from a population.
- Then the values of observations $Y_1, ..., Y_n$ are themselves random.

Random Sample

- Let $Y_1, ..., Y_n$ denote observations that are randomly chosen from a population.
- Then the values of observations $Y_1, ..., Y_n$ are themselves random.

Random Sample

- When Y_i has the same marginal distribution for $i = 1, ..., n$ then the observations $Y_1, ..., Y_n$ are said to be ***identically distributed***.
- When Y_i are independent, then observations $Y_1, ..., Y_n$ are said to be ***independently distributed***.
- In a random sample observations are ***independently and identically distributed (iid)***.

Random Sample

- When Y_i has the same marginal distribution for $i = 1, ..., n$ then the observations $Y_1, ..., Y_n$ are said to be ***identically distributed***.
- When Y_i are independent, then observations $Y_1, ..., Y_n$ are said to be ***independently distributed***.
- In a random sample observations are ***independently and identically distributed (iid)***.

Random Sample

- When Y_i has the same marginal distribution for $i = 1, ..., n$ then the observations $Y_1, ..., Y_n$ are said to be ***identically distributed***.
- When Y_i are independent, then observations $Y_1, ..., Y_n$ are said to be ***independently distributed***.
- In a random sample observations are ***independently and identically distributed (iid)***.

Random Sample

An Example: The Sample Mean

$$E(\overline{X}) = \mu_X$$

$$var(\overline{X}) = \sigma_X^2 / n$$

Large-Sample Approximations to Sampling Distribution

- **The finite-sample distribution:** The ***exact*** approach entails deriving a formula for the sampling distribution that holds exactly for any value of n.
- **The asymptotic distribution:** The ***approximate*** approach uses approximations to the sampling distribution that rely on the sample size being large.

Large-Sample Approximations to Sampling Distribution

- **The finite-sample distribution:** The ***exact*** approach entails deriving a formula for the sampling distribution that holds exactly for any value of n.
- **The asymptotic distribution:** The ***approximate*** approach uses approximations to the sampling distribution that rely on the sample size being large.

Large-Sample Approximations to Sampling Distribution

Consistency:
For any positive number ε, however small, if

$$\lim_{n \to \infty} Pr(|\gamma - c_n|) \leq \varepsilon) = 1$$

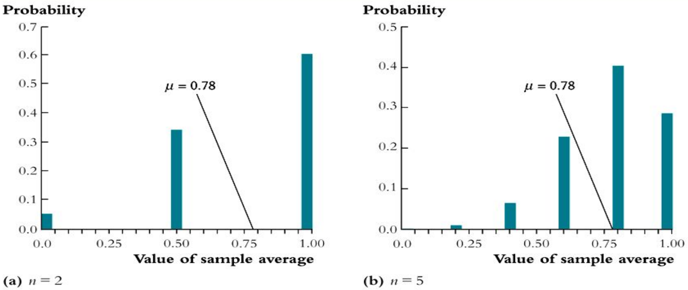
then c_n is a consistent estimator of γ.

Large-Sample Approximations to Sampling Distribution

The **Law of Large Numbers** states that, under general conditions, the sample average will be near the population mean with very high probability when n is large.

Law of Large Numbers

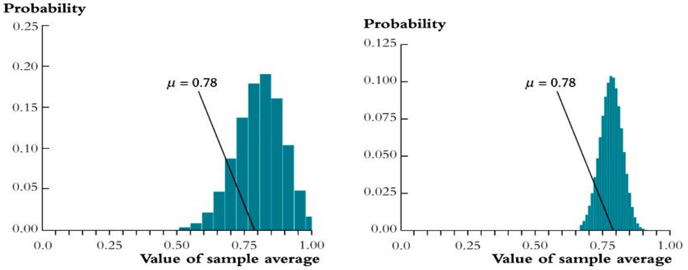
FIGURE 2.6 Sampling Distribution of the Sample Average of *n* Bernoulli Random Variables



The distributions are the sampling distributions of $\overline{Y}$, the sample average of n independent Bernoulli random variables with $p = \Pr(Y_i = 1) = 0.78$ (the probability of a fast commute is 78%). The variance of the sampling distribution of $\overline{Y}$ decreases as n gets larger, so the sampling distribution becomes more tightly concentrated around its mean $\mu = 0.78$ as the sample size n increases.

Law of Large Numbers

FIGURE 2.6 **Sampling Distribution of the Sample Average of *n* Bernoulli Random Variables**



(c) $n = 25$ **(d)** $n = 100$

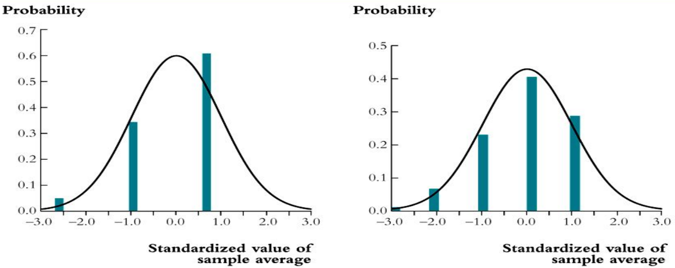
The distributions are the sampling distributions of $\overline{Y}$, the sample average of n independent Bernoulli random variables with $p = \Pr(Y_i = 1) = 0.78$ (the probability of a fast commute is 78%). The variance of the sampling distribution of $\overline{Y}$ decreases as n gets larger, so the sampling distribution becomes more tightly concentrated around its mean $\mu = 0.78$ as the sample size n increases.

Large-Sample Approximations to Sampling Distribution

The **Central Limit Theorem** states that, under general conditions, the distribution of the sample average is well approximated by a normal distribution when n is large.

Central Limit Theorem

FIGURE 2.7 **Distribution of the Standardized Sample Average of *n* Bernoulli Random Variables with $p = .78$**

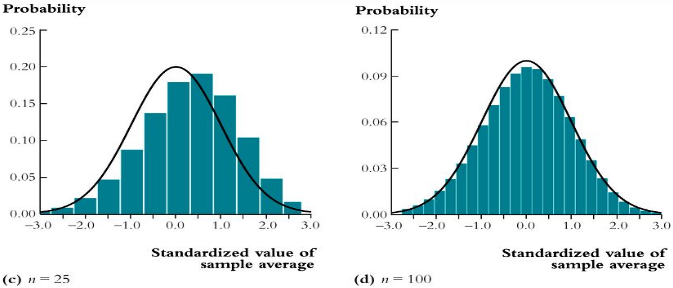


(a) $n = 2$ **(b)** $n = 5$

The sampling distribution of $\overline{Y}$ in Figure 2.6 is plotted here after standardizing $\overline{Y}$. This centers the distributions in Figure 2.6 and magnifies the scale on the horizontal axis by a factor of $\sqrt{n}$. When the sample size is large, the sampling distributions are increasingly well approximated by the normal distribution (the solid line), as predicted by the central limit theorem.

Central Limit Theorem

FIGURE 2.7 **Distribution of the Standardized Sample Average of *n* Bernoulli Random Variables with *p* = .78**



The sampling distribution of $\overline{Y}$ in Figure 2.6 is plotted here after standardizing $\overline{Y}$. This centers the distributions in Figure 2.6 and magnifies the scale on the horizontal axis by a factor of $\sqrt{n}$. When the sample size is large, the sampling distributions are increasingly well approximated by the normal distribution (the solid line), as predicted by the central limit theorem.

Estimation

Estimators

- A random variable is a **statistic**, if and only if it can be expressed as a function, not involving any unknown quantities, of a sample.
- An **estimator** is a statistic, whose outcome is used to estimate the value of an unknown parameter.
    - A *point estimate* is a single number that represents our best guess as the value of the unknown parameter.
    - An *interval estimate* is a range of numbers generated by a procedure having a high a priori probability of including the unknown parameter.

Estimators

- A random variable is a **statistic**, if and only if it can be expressed as a function, not involving any unknown quantities, of a sample.
- An **estimator** is a statistic, whose outcome is used to estimate the value of an unknown parameter.
    - A ***point estimate*** is a single number that represents our best guess as the value of the unknown parameter.
    - An ***interval estimate*** is a range of numbers generated by a procedure having a high a priori probability of including the unknown parameter.

Estimators

- A random variable is a **statistic**, if and only if it can be expressed as a function, not involving any unknown quantities, of a sample.
- An **estimator** is a statistic, whose outcome is used to estimate the value of an unknown parameter.
  - A ***point estimate*** is a single number that represents our best guess as the value of the unknown parameter.
  - An ***interval estimate*** is a range of numbers generated by a procedure having a high a priori probability of including the unknown parameter.

Estimators

- A random variable is a **statistic**, if and only if it can be expressed as a function, not involving any unknown quantities, of a sample.
- An **estimator** is a statistic, whose outcome is used to estimate the value of an unknown parameter.
  - A ***point estimate*** is a single number that represents our best guess as the value of the unknown parameter.
  - An ***interval estimate*** is a range of numbers generated by a procedure having a high a priori probability of including the unknown parameter.

Properties of Good Estimators

- **Unbiasedness:**

  Let c denote some estimator of γ. The estimator c is unbiased if $E[c] = \gamma$.

- **Consistency:**

  If the probability that the estimator c is within small interval of true value of γ approaches 1 as the sample size increases, then c is a consistent estimator of γ.

- **(Relative) Efficiency:**

  Let $\tilde{c}$ is another estimator of γ, and suppose that both c and $\tilde{c}$ are unbiased. Then c is said to be more efficient if $var(c) < var(\tilde{c})$

Properties of Good Estimators

- **Unbiasedness:**

  Let c denote some estimator of γ. The estimator c is unbiased if $E[c] = \gamma$.

- **Consistency:**

  If the probability that the estimator c is within small interval of true value of γ approaches 1 as the sample size increases, then c is a consistent estimator of γ.

- **(Relative) Efficiency:**

  Let $\tilde{c}$ is another estimator of γ, and suppose that both c and $\tilde{c}$ are unbiased. Then c is said to be more efficient if $var(c) < var(\tilde{c})$

Properties of Good Estimators

- **Unbiasedness:**

  Let c denote some estimator of γ. The estimator c is unbiased if $E[c] = \gamma$.

- **Consistency:**

  If the probability that the estimator c is within small interval of true value of γ approaches 1 as the sample size increases, then c is a consistent estimator of γ.

- **(Relative) Efficiency:**

  Let $\tilde{c}$ is another estimator of γ, and suppose that both c and $\tilde{c}$ are unbiased. Then c is said to be more efficient if $var(c) < var(\tilde{c})$

Estimation Techniques

- **Method of Moments:**
  Based on matching sample moments with population moments.
- **Maximum Likelihood Estimation:**
  It is the technique that attempts to find the most likely population from which the sample is drawn.
- **Least Squares Estimation:**
  Best fitting model.

Estimation Techniques

- **Method of Moments:**

  Based on matching sample moments with population moments.

- **Maximum Likelihood Estimation:**

  It is the technique that attempts to find the most likely population from which the sample is drawn.

- **Least Squares Estimation:**

  Best fitting model.

Estimation Techniques

- **Method of Moments:**

  Based on matching sample moments with population moments.

- **Maximum Likelihood Estimation:**

  It is the technique that attempts to find the most likely population from which the sample is drawn.

- **Least Squares Estimation:**

  Best fitting model.

Hypothesis Testing

Hypothesis Testing

- The starting point of statistical hypotheses is specifying the hypothesis to be tested, called the ***null hypothesis, H0***.
- Hypothesis testing entails using data to compare the null hypothesis to a second hypothesis, called the ***alternative hypothesis, H1***.

Hypothesis Testing

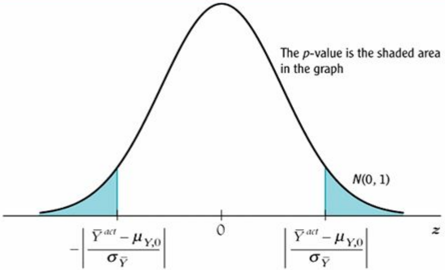
- The starting point of statistical hypotheses is specifying the hypothesis to be tested, called the ***null hypothesis, H0***.
- Hypothesis testing entails using data to compare the null hypothesis to a second hypothesis, called the ***alternative hypothesis, H1***.

Hypothesis Testing

The **p-value**, or **the significance probability**, is the probability drawing a statistic at least as adverse to the null hypothesis as the one calculated using the sample, assuming the null hypothesis is correct.

FIGURE 3.1 Calculating a *p*-value

The *p*-value is the probability of drawing a value of $\bar{Y}$ that differs from $\mu_{Y,0}$ by at least as much as $\bar{Y}^{act}$. In large samples, $\bar{Y}$ is distributed $N(\mu_{Y,0}, \sigma^2_{\bar{Y}})$ under the null hypothesis, so $(\bar{Y} - \mu_{Y,0})/\sigma_{\bar{Y}}$ is distributed $N(0, 1)$. Thus the *p*-value is the shaded standard normal tail probability outside $\pm|(\bar{Y}^{act} - \mu_{Y,0})/\sigma_{\bar{Y}}|$.



Hypothesis Testing

Errors in Hypothesis Testing

Rejecting a true null hypothesis is called a **type I error.**
Failure to reject a false null hypothesis is called a **type II error.**

		True State of Nature	
		H0 is true	**H0 is false**
The	**Accept H0**	Correct Decision	Type II error
Decision	**Reject H0**	Type I error	Correct Decision