Introduction to Econometrics

Recite 1 : Review of Probability Theory

Zishu Wang and Zhaopeng Qu

Business School,Nanjing University

March 1,2024

Outlines

Probabilities, the Sample Space and Random Variables

A Fundamental Axiom of Econometrics

1. Any economy can be seen as a **stochastic process** governed by a certain probability law.
2. Economic phenomena, often summarized in form of data, can be interpreted as a **realization** of this stochastic data generating process.

A Fundamental Axiom of Econometrics

1. Any economy can be seen as a **stochastic process** governed by a certain probability law.
2. Economic phenomena, often summarized in form of data, can be interpreted as a **realization** of this stochastic data generating process.

Probabilities and the Sample Space

- Random Phenomena, Outcomes and Probabilities
  - The mutually exclusive potential results of a random process are called the ***outcomes***.
  - The ***probability*** of an outcome is the proportion of the time that the outcome occurs in the long run.
- The Sample Space and Random Event
  - The set of all possible outcomes is called ***the sample space***.
  - An ***event*** is a subset of the sample space, that is, an event is a set of one or more outcomes.

Probabilities and the Sample Space

- Random Phenomena, Outcomes and Probabilities
  - The mutually exclusive potential results of a random process are called the ***outcomes***.
  - The ***probability*** of an outcome is the proportion of the time that the outcome occurs in the long run.
- The Sample Space and Random Event
  - The set of all possible outcomes is called ***the sample space***.
  - An ***event*** is a subset of the sample space, that is, an event is a set of one or more outcomes.

Probabilities and the Sample Space

- Random Phenomena, Outcomes and Probabilities
  - The mutually exclusive potential results of a random process are called the ***outcomes***.
  - The ***probability*** of an outcome is the proportion of the time that the outcome occurs in the long run.
- The Sample Space and Random Event
  - The set of all possible outcomes is called ***the sample space***.
  - An ***event*** is a subset of the sample space, that is, an event is a set of one or more outcomes.

Probabilities and the Sample Space

- Random Phenomena, Outcomes and Probabilities
  - The mutually exclusive potential results of a random process are called the ***outcomes***.
  - The ***probability*** of an outcome is the proportion of the time that the outcome occurs in the long run.
- The Sample Space and Random Event
  - The set of all possible outcomes is called ***the sample space***.
  - An ***event*** is a subset of the sample space, that is, an event is a set of one or more outcomes.

Probabilities and the Sample Space

- Random Phenomena, Outcomes and Probabilities
  - The mutually exclusive potential results of a random process are called the ***outcomes***.
  - The ***probability*** of an outcome is the proportion of the time that the outcome occurs in the long run.
- The Sample Space and Random Event
  - The set of all possible outcomes is called ***the sample space***.
  - An ***event*** is a subset of the sample space, that is, an event is a set of one or more outcomes.

Probabilities and the Sample Space

- Random Phenomena, Outcomes and Probabilities
  - The mutually exclusive potential results of a random process are called the ***outcomes***.
  - The ***probability*** of an outcome is the proportion of the time that the outcome occurs in the long run.
- The Sample Space and Random Event
  - The set of all possible outcomes is called ***the sample space***.
  - An ***event*** is a subset of the sample space, that is, an event is a set of one or more outcomes.

Random Variables

Random Variables(R.V.)

A random variable (r.v.) is a function that maps from the sample space of an experiment to the real line or $X : \Omega \rightarrow \mathbb{R}$

- A random variable is a numerical summary of a random outcome. They are numeric representation of uncertain events.(thus we can use math!)
- Notation: R.V.s are usually denoted by upper case letters (e.g. X), particular realizations are denoted by the corresponding lowercase letters (e.g. $x = 3$)

Example

Tossing a coin 5 times

- but not a random variable because it' s not numeric.
- $X(\omega)$ = number of heads in the five tosses. $X(HTHTT) = 2$

Random Variables

Random Variables(R.V.)

A random variable (r.v.) is a function that maps from the sample space of an experiment to the real line or $X : \Omega \to \mathbb{R}$

- A random variable is a numerical summary of a random outcome. They are numeric representation of uncertain events.(thus we can use math!)
- Notation: R.V.s are usually denoted by upper case letters (e.g. X), particular realizations are denoted by the corresponding lowercase letters (e.g. $x = 3$)

Example

Tossing a coin 5 times

- but not a random variable because it' s not numeric.
- $X(\omega)$ = number of heads in the five tosses. $X(HTHTT) = 2$

Random Variables

Random Variables(R.V.)

A random variable (r.v.) is a function that maps from the sample space of an experiment to the real line or $X : \Omega \rightarrow \mathbb{R}$

- A random variable is a numerical summary of a random outcome. They are numeric representation of uncertain events.(thus we can use math!)
- Notation: R.V.s are usually denoted by upper case letters (e.g. X), particular realizations are denoted by the corresponding lowercase letters (e.g. $x = 3$)

Example

Tossing a coin 5 times

- but not a random variable because it's not numeric.
- $X(\omega)$ = number of heads in the five tosses. $X(HTHTT) = 2$

Random Variables

Random Variables(R.V.)

A random variable (r.v.) is a function that maps from the sample space of an experiment to the real line or $X : \Omega \rightarrow \mathbb{R}$

- A random variable is a numerical summary of a random outcome. They are numeric representation of uncertain events.(thus we can use math!)
- Notation: R.V.s are usually denoted by upper case letters (e.g. X), particular realizations are denoted by the corresponding lowercase letters (e.g. $x = 3$)

Example

Tossing a coin 5 times

- but not a random variable because it’ s not numeric.
- $X(\omega) =$ number of heads in the five tosses. $X(HTHTT) = 2$

Random Variables

Random Variables(R.V.)

A random variable (r.v.) is a function that maps from the sample space of an experiment to the real line or $X : \Omega \rightarrow \mathbb{R}$

- A random variable is a numerical summary of a random outcome. They are numeric representation of uncertain events.(thus we can use math!)
- Notation: R.V.s are usually denoted by upper case letters (e.g. X), particular realizations are denoted by the corresponding lowercase letters (e.g. $x = 3$)

Example

Tossing a coin 5 times

- but not a random variable because it’ s not numeric.
- $X(\omega) =$ number of heads in the five tosses. $X(HTHTT) = 2$

Random Variables

Random Variables(R.V.)

A random variable (r.v.) is a function that maps from the sample space of an experiment to the real line or $X : \Omega \to \mathbb{R}$

- A random variable is a numerical summary of a random outcome. They are numeric representation of uncertain events.(thus we can use math!)
- Notation: R.V.s are usually denoted by upper case letters (e.g. X), particular realizations are denoted by the corresponding lowercase letters (e.g. $x = 3$)

Example

Tossing a coin 5 times

- but not a random variable because it’ s not numeric.
- $X(\omega) =$ number of heads in the five tosses. $X(HTHTT) = 2$

Probability Distributions

- Uncertainty over Ω uncertainty over value of . We' ll use probability to formalize this uncertainty.
- The probability distribution of a r.v. gives the probability of all of the possible values of the r.v.

$$P_X(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

Example

Tossing two coins: let X be the number of heads.

ω	$P(\{\omega\})$	$X(\omega)$
TT	1/4	0
HT	1/4	1
TH	1/4	1
HH	1/4	2

x	$P(X = x)$
0	1/4
1	1/2
2	1/4

Probability Distributions

- Uncertainty over Ω uncertainty over value of . We' ll use probability to formalize this uncertainty.
- The probability distribution of a r.v. gives the probability of all of the possible values of the r.v.

$$P_X(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

Example

Tossing two coins: let X be the number of heads.

ω	$P(\{\omega\})$	$X(\omega)$
TT	1/4	0
HT	1/4	1
TH	1/4	1
HH	1/4	2

x	$P(X = x)$
0	1/4
1	1/2
2	1/4

Probability Distributions

- Uncertainty over Ω uncertainty over value of . We' ll use probability to formalize this uncertainty.
- The probability distribution of a r.v. gives the probability of all of the possible values of the r.v.

$$P_X(X = x) = P(\{\omega \in \Omega : X(\omega) = x\})$$

Example

Tossing two coins: let X be the number of heads.

ω	$P(\{\omega\})$	$X(\omega)$
TT	1/4	0
HT	1/4	1
TH	1/4	1
HH	1/4	2

x	$P(X = x)$
0	1/4
1	1/2
2	1/4

Probabilities Function and its Properties

- As to any event X, we have
  - $0 \leq P(x) \leq 1$
  - $P(\Omega) = 1$ and $P(\Phi) = 0$, where Ω is the universal set and Φ is the empty set
  - $P(X) = 1 - P(\overline{X})$, where $\overline{X}$ is the complementary set to X
  - if $X_1, X_2, ..., X_n...$ is mutual exclusion, then $P(\cup_{i=1}^{\infty} X_i) = \Sigma_{i=1}^{\infty} P(X_i)$

Probabilities Function and its Properties

- As to any event X, we have
  - $0 \leq P(x) \leq 1$
  - $P(\Omega) = 1$ and $P(\Phi) = 0$, where Ω is the universal set and Φ is the empty set
  - $P(X) = 1 - P(\overline{X})$, where $\overline{X}$ is the complementary set to X
  - if $X_1, X_{2,...}, X_n...$ is mutual exclusion, then $P(\cup_{i=1}^{\infty} X_i) = \Sigma_{i=1}^{\infty} P(X_i)$

Probabilities Function and its Properties

- As to any event X, we have
  - $0 \leq P(x) \leq 1$
  - $P(\Omega) = 1$ and $P(\Phi) = 0$, where Ω is the universal set and Φ is the empty set
  - $P(X) = 1 - P(\overline{X})$, where $\overline{X}$ is the complementary set to X
  - if $X_1, X_2, ..., X_n...$ is mutual exclusion, then $P(\cup_{i=1}^{\infty} X_i) = \Sigma_{i=1}^{\infty} P(X_i)$

Probabilities Function and its Properties

- As to any event X, we have
  - $0 \leq P(x) \leq 1$
  - $P(\Omega) = 1$ and $P(\Phi) = 0$, where Ω is the universal set and Φ is the empty set
  - $P(X) = 1 - P(\overline{X})$, where $\overline{X}$ is the complementary set to X
  - if $X_1, X_2, ..., X_n ...$ is mutual exclusion, then $P(\cup_{i=1}^{\infty} X_i) = \Sigma_{i=1}^{\infty} P(X_i)$

Probabilities Function and its Properties

- As to any event X, we have
  - $0 \leq P(x) \leq 1$
  - $P(\Omega) = 1$ and $P(\Phi) = 0$, where Ω is the universal set and Φ is the empty set
  - $P(X) = 1 - P(\overline{X})$, where $\overline{X}$ is the complementary set to X
  - if $X_1, X_2, ..., X_n ...$ is mutual exclusion, then $P(\cup_{i=1}^{\infty} X_i) = \Sigma_{i=1}^{\infty} P(X_i)$

The Total Probability Rule(全概法则)

- If $X_1, X_2, ..., X_n...$ are mutually exclusive and exhaustive to the sample space, then for any event Y,

$$P(Y) = \Sigma_i^{\infty} P(Y|X_i)P(X_i)$$

- 直观理解：若事件 Y 可以被划分为一个互斥子事件的集合，那么事件 Y 发生的概率，等于事件 Y 在每个子事件中发生的概率之和。

Example

猜球游戏。两个袋子：大袋装 20 个球，小袋装 10 个球。大袋中白球 15 个，黑球 5 个。小袋中白球 5 个，黑球 5 个。随机从袋子里抽到白球的概率是多少？

The Total Probability Rule(全概法则)

- If $X_1, X_2, ..., X_n...$ are mutually exclusive and exhaustive to the sample space, then for any event Y,

$$P(Y) = \Sigma_i^{\infty} P(Y|X_i) P(X_i)$$

- 直观理解：若事件 Y 可以被划分为一个互斥子事件的集合，那么事件 Y 发生的概率，等于事件 Y 在每个子事件中发生的概率之和。

Example

猜球游戏。两个袋子：大袋装 20 个球，小袋装 10 个球。大袋中白球 15 个，黑球 5 个。小袋中白球 5 个，黑球 5 个。随机从袋子里抽到白球的概率是多少？

The Total Probability Rule(全概法则)

- If $X_1, X_2, ..., X_n...$ are mutually exclusive and exhaustive to the sample space, then for any event Y,

$$P(Y) = \Sigma_i^{\infty} P(Y|X_i)P(X_i)$$

- 直观理解：若事件 Y 可以被划分为一个互斥子事件的集合，那么事件 Y 发生的概率，等于事件 Y 在每个子事件中发生的概率之和。

Example

猜球游戏。两个袋子：大袋装 20 个球，小袋装 10 个球。大袋中白球 15 个，黑球 5 个。小袋中白球 5 个，黑球 5 个。随机从袋子里抽到白球的概率是多少？

Bayse's Rule(贝叶斯法则)

- The conditional probability of Y given X is the conditional probability of X given Y times the relative marginal probabilities of Y and X

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(X)}$$

Example

仍然是上一个猜球游戏设定。已知抽到的了白球，请问来自小袋子的概率是多少？

- 直观理解：加入新信息可以修正初始概率。

Bayse's Rule(贝叶斯法则)

- The conditional probability of Y given X is the conditional probability of X given Y times the relative marginal probabilities of Y and X

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(X)}$$

Example

仍然是上一个猜球游戏设定。已知抽到的了白球，请问来自小袋子的概率是多少？

- 直观理解：加入新信息可以修正初始概率。

Bayse's Rule(贝叶斯法则)

- The conditional probability of Y given X is the conditional probability of X given Y times the relative marginal probabilities of Y and X

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(X)}$$

Example

仍然是上一个猜球游戏设定。已知抽到的了白球，请问来自小袋子的概率是多少？

- 直观理解：加入新信息可以修正初始概率。

Distributional Functions of R.V.

- It is cumbersome to derive the probabilities of X each time we need them, so it is helpful to have a function that can give us the probability of values or sets of values of X.

Definition

The **cumulative distribution function** or **c.d.f** of a r.v. X, denoted $F_X(x)$, is defined by

$$F_X(x) \equiv P_X(X \leq x)$$

- The c.d.f tells us the probability of a r.v. being less than some given value.

Distributional Functions of R.V.

- It is cumbersome to derive the probabilities of X each time we need them, so it is helpful to have a function that can give us the probability of values or sets of values of X.

Definition

The **cumulative distribution function** or **c.d.f** of a r.v. X, denoted $F_X(x)$, is defined by

$$F_X(x) \equiv P_X(X \leq x)$$

- The c.d.f tells us the probability of a r.v. being less than some given value.

Distributional Functions of R.V.

- It is cumbersome to derive the probabilities of X each time we need them, so it is helpful to have a function that can give us the probability of values or sets of values of X.

Definition

The **cumulative distribution function** or **c.d.f** of a r.v. X, denoted $F_X(x)$, is defined by

$$F_X(x) \equiv P_X(X \leq x)$$

- The c.d.f tells us the probability of a r.v. being less than some given value.

Distribution Functions of R.V.

- We have two kinds of r.v.s

Definition

A r.v. X, is **discrete** if its range(the set of values it can take) is finite ($X \in \{x_1, x_2, ...x_k\}$) or countably infinite($X \in \{x_1, x_2, ...\}$)

- eg: the number of computer crashes before deadline

Definition

A r.v. X, is **continuous** if it can contain all real numbers in a interval. There are an uncountably infinite number of possible realizations.

- eg: commuting times from home to school

Distribution Functions of R.V.

- We have two kinds of r.v.s

Definition

A r.v. X, is **discrete** if its range(the set of values it can take) is finite ($X \in \{x_1, x_2, ...x_k\}$) or countably infinite($X \in \{x_1, x_2, ...\}$)

- eg: the number of computer crashes before deadline

Definition

A r.v. X, is **continuous** if it can contain all real numbers in a interval. There are an uncountably infinite number of possible realizations.

- eg: commuting times from home to school

Distribution Functions of R.V.

- We have two kinds of r.v.s

Definition

A r.v. X, is **discrete** if its range(the set of values it can take) is finite $(X \in \{x_1, x_2, ...x_k\})$ or countably infinite$(X \in \{x_1, x_2, ...\})$

- eg: the number of computer crashes before deadline

Definition

A r.v. X, is **continuous** if it can contain all real numbers in a interval. There are an uncountably infinite number of possible realizations.

- eg: commuting times from home to school

Distribution Functions of R.V.

- We have two kinds of r.v.s

Definition

A r.v. X, is **discrete** if its range(the set of values it can take) is finite $(X \in \{x_1, x_2, ...x_k\})$ or countably infinite$(X \in \{x_1, x_2, ...\})$

- eg: the number of computer crashes before deadline

Definition

A r.v. X, is **continuous** if it can contain all real numbers in a interval. There are an uncountably infinite number of possible realizations.

- eg: commuting times from home to school

Distribution Functions of R.V.

- We have two kinds of r.v.s

Definition

A r.v. X, is **discrete** if its range(the set of values it can take) is finite ($X \in \{x_1, x_2, ...x_k\}$) or countably infinite($X \in \{x_1, x_2, ...\}$)

- eg: the number of computer crashes before deadline

Definition

A r.v. X, is **continuous** if it can contain all real numbers in a interval. There are an uncountably infinite number of possible realizations.

- eg: commuting times from home to school

Probability Distribution of a ***Discrete*** R.V.

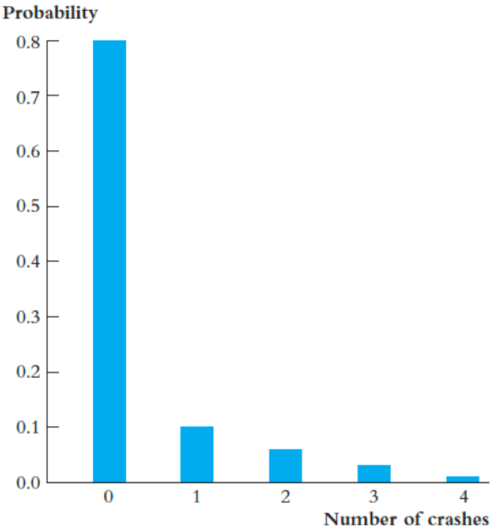
Probability mass function

Probability mass function (p.m.f.) describes the distribution of r.v. when it is discrete:

$$f_X(x_k) = P(X = x_k) = p_x,\ k = 1, 2, ..., n$$

FIGURE 2.1 Probability Distribution of the Number of Computer Crashes

The height of each bar is the probability that the computer crashes the indicated number of times. The height of the first bar is 0.8, so the probability of 0 computer crashes is 80%. The height of the second bar is 0.1, so the probability of 1 computer crash is 10%, and so forth for the other bars.



Probability Distribution of a ***Discrete*** R.V.

c.d.f of a discrete r.v

the c.d.f of a discrete r.v. is denoted as

$$F_X(x) = P(X \leq x) = \sum_{X_k \leq x} f_X(x_k)$$

TABLE 2.1 Probability of Your Computer Crashing *M* Times

	Outcome (number of crashes)				
	0	1	2	3	4
Probability distribution	0.80	0.10	0.06	0.03	0.01
Cumulative probability distribution	0.80	0.90	0.96	0.99	1.00

Probability Distribution of a ***Continuous*** R.V.

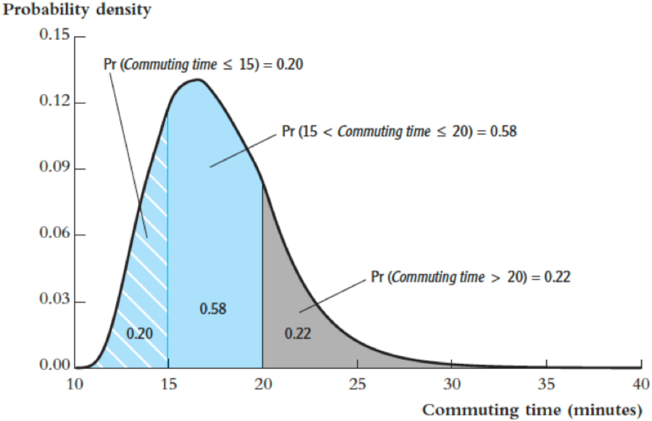
Probability density function

The probability density function or p.d.f., for a continuous random variable X is the function that satisfies for any interval, B

$$P(X \in B) = \int_B f_X(x)dx$$

Probability Distribution of a ***Continuous*** R.V.

- Specifically, for a subset of the real line(a, b): $P(a < X < b) = \int_a^b f_X(x)dx$, thus the probability of a region is the area under the p.d.f. for that region.



(b) Probability density function of commuting time

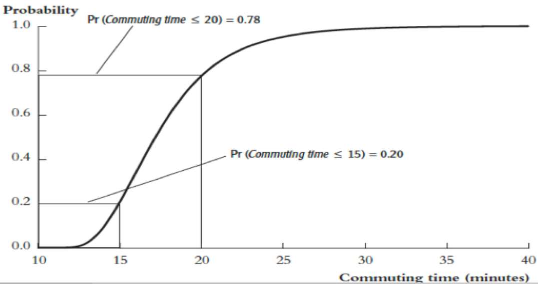
Probability Distribution of a *Continuous* R.V.

Cumulative probability distribution

just as it is for a discrete random variable, except using p.d.f to calculate the probability of x,

$$F(X) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)dx$$

FIGURE 2.2 Cumulative Distribution and Probability Density Functions of Commuting Time



Expected Values, Mean, and Variance

Properties of Distributions

- Probability distributions describe the uncertainty about r.v.s. The cdf/pmf/pdf give us all the information about the distribution of some r.v., but we are quite often interested in some feature of the distribution rather than the entire distribution.
  - What is the difference between these two density curves? How might we summarize this difference?
- There are two simple indictors:
  1. **Central tendency:** where the center of the distribution is.
  2. **Spread:** how spread out the distribution is around the center.

Properties of Distributions

- Probability distributions describe the uncertainty about r.v.s. The cdf/pmf/pdf give us all the information about the distribution of some r.v., but we are quite often interested in some feature of the distribution rather than the entire distribution.
  - What is the difference between these two density curves? How might we summarize this difference?
- There are two simple indictors:
  1. **Central tendency**: where the center of the distribution is.
  2. **Spread**: how spread out the distribution is around the center.

Properties of Distributions

- Probability distributions describe the uncertainty about r.v.s. The cdf/pmf/pdf give us all the information about the distribution of some r.v., but we are quite often interested in some feature of the distribution rather than the entire distribution.
  - What is the difference between these two density curves? How might we summarize this difference?
- There are two simple indictors:
  1. **Central tendency**: where the center of the distribution is.
     - Mean/expectation（均值或期望）
  2. **Spread**: how spread out the distribution is around the center.
     - Variance/standard deviation（方差或标准差）

Properties of Distributions

- Probability distributions describe the uncertainty about r.v.s. The cdf/pmf/pdf give us all the information about the distribution of some r.v., but we are quite often interested in some feature of the distribution rather than the entire distribution.
  - What is the difference between these two density curves? How might we summarize this difference?
- There are two simple indictors:
  1. **Central tendency**: where the center of the distribution is.
     - Mean/expectation (均值或期望)
  2. **Spread**: how spread out the distribution is around the center.
     - Variance/standard deviation (方差或标准差)

Properties of Distributions

- Probability distributions describe the uncertainty about r.v.s. The cdf/pmf/pdf give us all the information about the distribution of some r.v., but we are quite often interested in some feature of the distribution rather than the entire distribution.
  - What is the difference between these two density curves? How might we summarize this difference?
- There are two simple indictors:
  1. **Central tendency**: where the center of the distribution is.
     - Mean/expectation (均值或期望)
  2. **Spread**: how spread out the distribution is around the center.
     - Variance/standard deviation (方差或标准差)

Properties of Distributions

- Probability distributions describe the uncertainty about r.v.s. The cdf/pmf/pdf give us all the information about the distribution of some r.v., but we are quite often interested in some feature of the distribution rather than the entire distribution.
  - What is the difference between these two density curves? How might we summarize this difference?
- There are two simple indictors:
  1. **Central tendency**: where the center of the distribution is.
     - Mean/expectation (均值或期望)
  2. **Spread**: how spread out the distribution is around the center.
     - Variance/standard deviation（方差或标准差）

Properties of Distributions

- Probability distributions describe the uncertainty about r.v.s. The cdf/pmf/pdf give us all the information about the distribution of some r.v., but we are quite often interested in some feature of the distribution rather than the entire distribution.
  - What is the difference between these two density curves? How might we summarize this difference?
- There are two simple indictors:
  1. **Central tendency**: where the center of the distribution is.
     - Mean/expectation (均值或期望)
  2. **Spread**: how spread out the distribution is around the center.
     - Variance/standard deviation（方差或标准差）

The Expected Value of a Random Variable

- The expected value of a random variable X, denoted $E(X)$ or μ_x, is the long-run average value of the random variable over many repeated trials or occurrences. it is a natural measure of central tendency.
- For a *discrete* r.v., $X \in \{x_1, x_2, ..., x_k\}$

$$\mu_X = E[X] = \sum_{j=1}^{k} x_j p_j$$

  it is computed as a *weighted average* of the value of r.v., where the weights are the probability of each value occurring.
- For a *continuous* r.v., X, use the integral

$$\mu_X = E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$$

The Expected Value of a Random Variable

- The expected value of a random variable X, denoted $E(X)$ or μ_x, is the long-run average value of the random variable over many repeated trials or occurrences. it is a natural measure of central tendency.
- For a *discrete* r.v., $X \in \{x_1, x_2, ..., x_k\}$

$$\mu_X = E[X] = \sum_{j=1}^{k} x_j p_j$$

it is computed as a *weighted average* of the value of r.v., where the weights are the probability of each value occurring.

- For a *continuous* r.v., X, use the integral

$$\mu_X = E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$$

The Expected Value of a Random Variable

- The expected value of a random variable X, denoted $E(X)$ or μ_x, is the long-run average value of the random variable over many repeated trials or occurrences. it is a natural measure of central tendency.
- For a *discrete* r.v., $X \in \{x_1, x_2, ..., x_k\}$

$$\mu_X = E[X] = \sum_{j=1}^{k} x_j p_j$$

it is computed as a *weighted average* of the value of r.v., where the weights are the probability of each value occurring.

- For a *continuous* r.v., X, use the integral

$$\mu_X = E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Properties of Expectation

1. **Additivity**: expectation of sums are sums of expectations

$$E[X+Y]=E[X]+E[Y]$$

2. **Homogeneity**: Suppose that a and b are constants. Then

$$E[aX+b]=aE[X]+b$$

3. **Law of the Unconscious Statistician**, or LOTUS, if g(x) is a function of a discrete random variable, then

$$E[g(X)]=\begin{cases}\sum_x g(x)f_X(x) & when\ x\ is\ discrete\\ \int g(x)f_X(x)dx & when\ x\ is\ continuous\end{cases}$$

Properties of Expectation

1. **Additivity**: expectation of sums are sums of expectations

$$E[X+Y]=E[X]+E[Y]$$

2. **Homogeneity**: Suppose that a and b are constants. Then

$$E[aX+b]=aE[X]+b$$

3. **Law of the Unconscious Statistician**, or LOTUS, if g(x) is a function of a discrete random variable, then

$$E[g(X)]=\begin{cases}\sum_x g(x)f_X(x) & \textit{when x is discrete}\\ \int g(x)f_X(x)dx & \textit{when x is continuous}\end{cases}$$

Properties of Expectation

1. **Additivity**: expectation of sums are sums of expectations

$$E[X+Y]=E[X]+E[Y]$$

2. **Homogeneity**: Suppose that a and b are constants. Then

$$E[aX+b]=aE[X]+b$$

3. **Law of the Unconscious Statistician**, or LOTUS, if g(x) is a function of a discrete random variable, then

$$E\left[g(X)\right]=\begin{cases}\sum_x g(x)f_X(x) & when\, x\, is\, discrete\\ \int g(x)f_X(x)dx & when\, x\, is\, continuous\end{cases}$$

The Variance of a Random Variable

- Besides some sense of where the middle of the distribution is, we also want to know how spread out the distribution is around that middle.

Definition

The **Variance** of a random variable X, denoted $var(X)$ or σ_X^2

$$\sigma_X^2 = Var(X) = E[(X - \mu_X)^2]$$

The **Standard Deviation** of X, denoted σ_X, is just the square root of the variance.

$$\sigma_X = \sqrt{Var(X)}$$

The Variance of a Random Variable

- Besides some sense of where the middle of the distribution is, we also want to know how spread out the distribution is around that middle.

Definition

The**Variance** of a random variable X, denoted $var(X)$or σ_X^2

$$\sigma_X^2 = Var(X) = E[(X - \mu_X)^2]$$

The **Standard Deviation** of X, denoted σ_X, is just the square root of the variance.

$$\sigma_X = \sqrt{Var(X)}$$

Properties of Variance

- If a and b are constants, then we have the following properties:
  1. $V(b) = 0$
  2. $V(aX + b) = a^2 V(X)$
  3. $V(X) = E[X^2] - (E[X])^2$

Example

Bernoulli Distribution:

Properties of Variance

- If a and b are constants, then we have the following properties:
  1. $V(b) = 0$
  2. $V(aX + b) = a^2 V(X)$
  3. $V(X) = E[X^2] - (E[X])^2$

Example

Bernoulli Distribution:

Properties of Variance

- If a and b are constants, then we have the following properties:
  1. $V(b) = 0$
  2. $V(aX + b) = a^2 V(X)$
  3. $V(X) = E[X^2] - (E[X])^2$

Example

Bernoulli Distribution:

Properties of Variance

- If a and b are constants, then we have the following properties:
  1. $V(b) = 0$
  2. $V(aX + b) = a^2 V(X)$
  3. $V(X) = E[X^2] - (E[X])^2$

Example

Bernoulli Distribution:

Properties of Variance

- If a and b are constants, then we have the following properties:
  1. $V(b) = 0$
  2. $V(aX + b) = a^2 V(X)$
  3. $V(X) = E[X^2] - (E[X])^2$

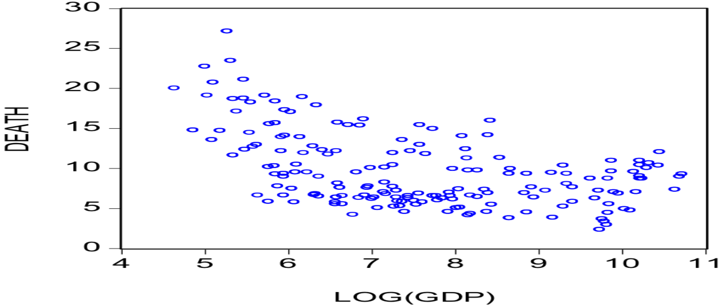
Example

Bernoulli Distribution:

Multiple Random Variables

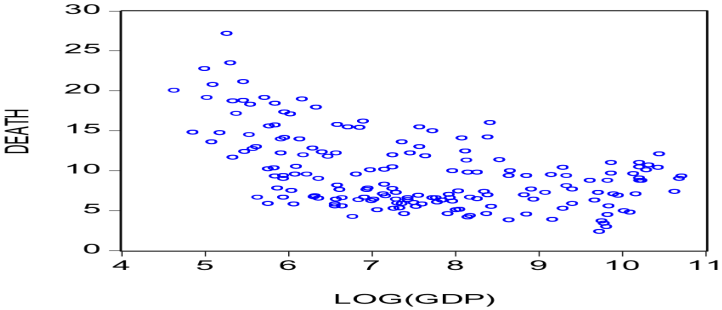
Why multiple random variables?

- We are going to want to know what the relationships are between variables. “The objective of science is the discovery of the relations” —Lord Kelvin
- In most cases, we often want to explore the relationship between two variables in one study.
  - eg. Mortality and GDP growth



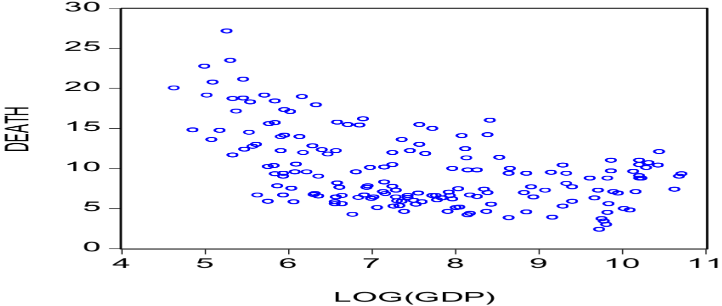
Why multiple random variables?

- We are going to want to know what the relationships are between variables. "The objective of science is the discovery of the relations" —Lord Kelvin
- In most cases, we often want to explore the relationship between two variables in one study.
    - eg. Mortality and GDP growth



Why multiple random variables?

- We are going to want to know what the relationships are between variables. "The objective of science is the discovery of the relations" —Lord Kelvin
- In most cases, we often want to explore the relationship between two variables in one study.
  - eg. Mortality and GDP growth



Joint Probability Distribution

- Consider two *discrete* random variables X and Y with a joint probability distribution, then the joint probability mass function of (X, Y) describes the probability of any pair of values:

$$f_{X,Y}(x, y) = P(X = x, Y = y) = p_{xy}$$

TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times

	Rain ($X = 0$)	No Rain ($X = 1$)	Total
Long commute ($Y = 0$)	0.15	0.07	0.22
Short commute ($Y = 1$)	0.15	0.63	0.78
Total	0.30	0.70	1.00

Marginal Probability Distribution

- The marginal distribution: often need to figure out the distribution of just one of the r.v.s.

$$f_Y(y) = P(Y = y) = \sum_x f_{X,Y}(x, y)$$

- Intuition: sum over the probability that $Y = y$ for all possible values of x.

TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times

	Rain ($X = 0$)	No Rain ($X = 1$)	Total
Long commute ($Y = 0$)	0.15	0.07	0.22
Short commute ($Y = 1$)	0.15	0.63	0.78
Total	0.30	0.70	1.00

Marginal Probability Distribution

- The marginal distribution: often need to figure out the distribution of just one of the r.v.s.

$$f_Y(y) = P(Y = y) = \sum_x f_{X,Y}(x, y)$$

- Intuition: sum over the probability that $Y = y$ for all possible values of x.

TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times

	Rain ($X = 0$)	No Rain ($X = 1$)	Total
Long commute ($Y = 0$)	0.15	0.07	0.22
Short commute ($Y = 1$)	0.15	0.63	0.78
Total	0.30	0.70	1.00

Joint Probability Density Function

- Consider two *continuous* random variables X and Y with a joint probability distribution, then the **joint probability density function** of (X, Y) is a function, denoted as $f_{X,Y}(x, y)$ such that:
  1. $f_{X,Y}(x, y) \geq 0$
  2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dxdy = 1$
  3. $P(a < X < b, c < Y < d) = \int_c^d \int_a^b f_{X,Y}(x, y)\, dxdy$, thus the probability in the $\{a, b, c, d\}$area.

Joint Probability Density Function

- Consider two *continuous* random variables X and Y with a joint probability distribution, then the **joint probability density function** of (X, Y) is a function, denoted as $f_{X,Y}(x,y)$ such that:
  1. $f_{X,Y}(x,y) \geq 0$
  2. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dxdy = 1$
  3. $P(a < X < b, c < Y < d) = \int_c^d \int_a^b f_{X,Y}(x,y)\,dxdy$, thus the probability in the $\{a,b,c,d\}$area.

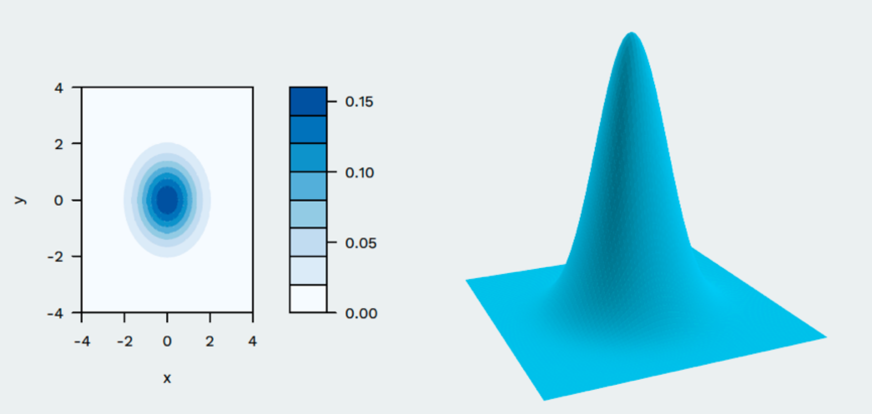
Joint Probability Density Function

- Consider two *continuous* random variables X and Y with a joint probability distribution, then the **joint probability density function** of (X, Y) is a function, denoted as $f_{X,Y}(x, y)$ such that:
  1. $f_{X,Y}(x, y) \geq 0$
  2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dxdy = 1$
  3. $P(a < X < b, c < Y < d) = \int_c^d \int_a^b f_{X,Y}(x, y)\, dxdy$, thus the probability in the $\{a, b, c, d\}$area.

Joint Probability Density Function

- Consider two *continuous* random variables X and Y with a joint probability distribution, then the **joint probability density function** of (X, Y) is a function, denoted as $f_{X,Y}(x, y)$ such that:
  1. $f_{X,Y}(x, y) \geq 0$
  2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dxdy = 1$
  3. $P(a < X < b, c < Y < d) = \int_c^d \int_a^b f_{X,Y}(x, y)\, dxdy$, thus the probability in the $\{a, b, c, d\}$area.

Joint Probability Density Function

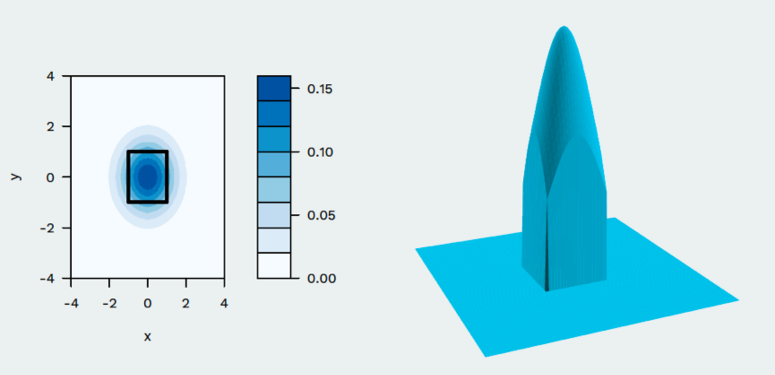


- Y and X axes denote on the "floor", height is the value of $f_{XY}(x, y)$

Joint Probability Density Function

- The probability equals to volume above a specific region

$$P(X, Y) \in A) = \int_{(x,y)\in A} f_{X,Y}(x, y)dxdy$$



Continuous Marginal Distribution

- the marginal p.d.f of Y by integrating over the distribution of X:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx$$

- the marginal p.d.f of X by integrating over the distribution of Y:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$

Continuous Marginal Distribution

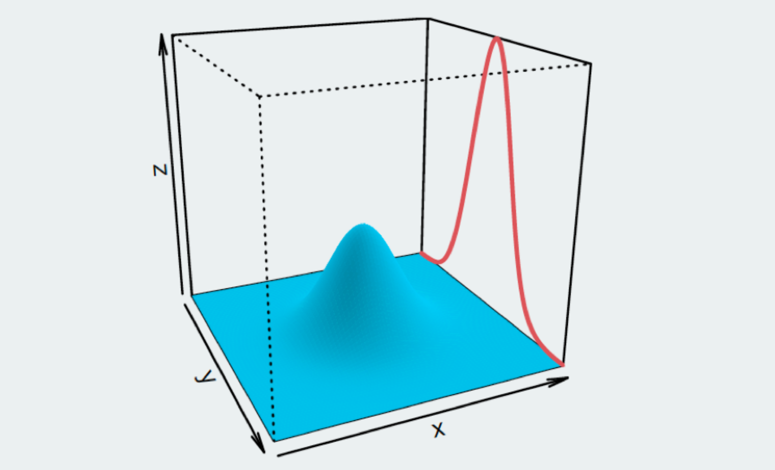
- the marginal p.d.f of Y by integrating over the distribution of X:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dx$$

- the marginal p.d.f of X by integrating over the distribution of Y:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y)dy$$

Continuous Marginal Distribution



- Pile up all of the joint density onto a single dimension

Joint Cumulative Distribution Function

- The **joint cumulative distribution function** of (X, Y) is

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v)\, dudv$$

- Transform joint c.d.f and joint p.d.f

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial y}$$

Joint Cumulative Distribution Function

- The **joint cumulative distribution function** of (X, Y) is

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v)\, dudv$$

- Transform joint c.d.f and joint p.d.f

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial y}$$

Properties of Joint Distributions

Expectations over multiple r.v.s

- Expectations over multiple r.v.s

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y) f_{X,Y}(x,y) & if \\ \int_x \int_y g(x,y) f_{X,Y}(x,y) dxdy & if \end{cases}$$

- Marginal expectation

$$E[Y] = \begin{cases} \sum_x \sum_y y f_{X,Y}(x,y) & if \\ \int_x \int_y y f(x,y) dxdy & if \end{cases}$$

Expectations over multiple r.v.s

- Expectations over multiple r.v.s

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y) f_{X,Y}(x,y) & \textit{if} \\ \int_x \int_y g(x,y) f_{X,Y}(x,y) dxdy & \textit{if} \end{cases}$$

- Marginal expectation

$$E[Y] = \begin{cases} \sum_x \sum_y y f_{X,Y}(x,y) & \textit{if} \\ \int_x \int_y y f(x,y) dxdy & \textit{if} \end{cases}$$

Independence

Independence

Two r.v.s X and Y are *independent, which we denote it as* $X \perp Y$*, if for all sets* A *and* B

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

- Intuition: knowing the value of X gives us no information about the value of Y.
- IfX and Y are *independent*, then
  - Joint p.d.f is the product of marginal p.d.f, thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - Joint c.d.f is the product of marginal c.d.f, thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - functions of independent r.v.s are independent, thus $h(X) \perp g(Y)$ for any functions $h(\cdot)$ and $g(\cdot)$.

Independence

Independence

Two r.v.s X and Y are *independent, which we denote it as* $X \perp Y$*, if for all sets* A *and* B

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

- Intuition: knowing the value of X gives us no information about the value of Y.
- If X and Y are *independent*, then
  - Joint p.d.f is the product of marginal p.d.f, thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - Joint c.d.f is the product of marginal c.d.f, thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - functions of independent r.v.s are independent, thus $h(X) \perp g(Y)$ for any functions $h(\cdot)$ and $g(\cdot)$.

Independence

Independence

Two r.v.s X and Y are *independent*, *which we denote it as* $X \perp Y$*, if for all sets* A *and* B

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

- Intuition: knowing the value of X gives us no information about the value of Y.
- If X and Y are *independent*, then
  - Joint p.d.f is the product of marginal p.d.f, thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - Joint c.d.f is the product of marginal c.d.f, thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - functions of independent r.v.s are independent, thus $h(X) \perp g(Y)$ for any functions $h(\cdot)$ and $g(\cdot)$.

Independence

Independence

Two r.v.s X and Y are *independent, which we denote it as* $X \perp Y$*, if for all sets* A *and* B

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

- Intuition: knowing the value of X gives us no information about the value of Y.
- If X and Y are *independent*, then
  - Joint p.d.f is the product of marginal p.d.f, thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - Joint c.d.f is the product of marginal c.d.f, thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - functions of independent r.v.s are independent, thus $h(X) \perp g(Y)$ for any functions $h(\cdot)$ and $g(\cdot)$.

Independence

Independence

Two r.v.s X and Y are *independent, which we denote it as* $X \perp Y$*, if for all sets* A *and* B

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

- Intuition: knowing the value of X gives us no information about the value of Y.
- IfX and Y are *independent*, then
  - Joint p.d.f is the product of marginal p.d.f, thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - Joint c.d.f is the product of marginal c.d.f, thus $f_{X,Y}(x,y) = f_X(x)f_Y(y)$
  - functions of independent r.v.s are independent, thus $h(X) \perp g(Y)$ for any functions $h(\cdot)$ and $g(\cdot)$.

Independence

Independence

Two r.v.s X and Y are *independent, which we denote it as* $X \perp Y$, *if for all sets* A *and* B

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

- Intuition: knowing the value of X gives us no information about the value of Y.
- IfX and Y are *independent*, then
  - Joint p.d.f is the product of marginal p.d.f, thus $f_{X,Y}(x, y) = f_X(x)f_Y(y)$
  - Joint c.d.f is the product of marginal c.d.f, thus $f_{X,Y}(x, y) = f_X(x)f_Y(y)$
  - functions of independent r.v.s are independent, thus $h(X) \perp g(Y)$ for any functions $h(\cdot)$ and $g(\cdot)$.

Independence

Theorem (Independence)

if X and Y are independent r.v.s, then

$$E[XY] = E[X]E[Y]$$

Proof.

Skip. you could finish it by yourself. □

Independence

Theorem (Independence)

if X and Y are independent r.v.s, then

$$E[XY] = E[X]E[Y]$$

Proof.

Skip. you could finish it by yourself. □

Covariance

- If two variables are not independent, we could still measure the strength of their dependence by the definition of covariance.

Covariance

the covariance between X and Y is defined as

$$Cov[X, Y] = E[(X - E[X])(Y - E[Y])]$$

- Properties of covariances:
  - $Cov[X, Y] = E[XY] - E[X]E[Y]$
  - If $X \perp Y$, $Cov[X, Y] = 0$

Covariance

- If two variables are not independent, we could still measure the strength of their dependence by the definition of covariance.

Covariance

the covariance between X and Y is defined as

$$Cov[X,Y] = E\left[(X - E[X])\,(Y - E[Y])\right]$$

- Properties of covariances:
  - $Cov[X,Y] = E[XY] - E[X]E[Y]$
  - If $X \perp Y$, $Cov[X,Y] = 0$

Covariance

- If two variables are not independent, we could still measure the strength of their dependence by the definition of covariance.

Covariance

the covariance between X and Y is defined as

$$Cov[X,Y] = E\left[(X - E[X])\,(Y - E[Y])\right]$$

- Properties of covariances:
  - $Cov[X,Y] = E[XY] - E[X]E[Y]$
  - If $X \perp Y$, $Cov[X,Y] = 0$

Covariance

- If two variables are not independent, we could still measure the strength of their dependence by the definition of covariance.

Covariance

the covariance between X and Y is defined as

$$Cov[X, Y] = E\left[(X - E[X])\,(Y - E[Y])\right]$$

- Properties of covariances:
  - $Cov[X, Y] = E[XY] - E[X]E[Y]$
  - If $X \perp Y$, $Cov[X, Y] = 0$

Covariance

- If two variables are not independent, we could still measure the strength of their dependence by the definition of covariance.

Covariance

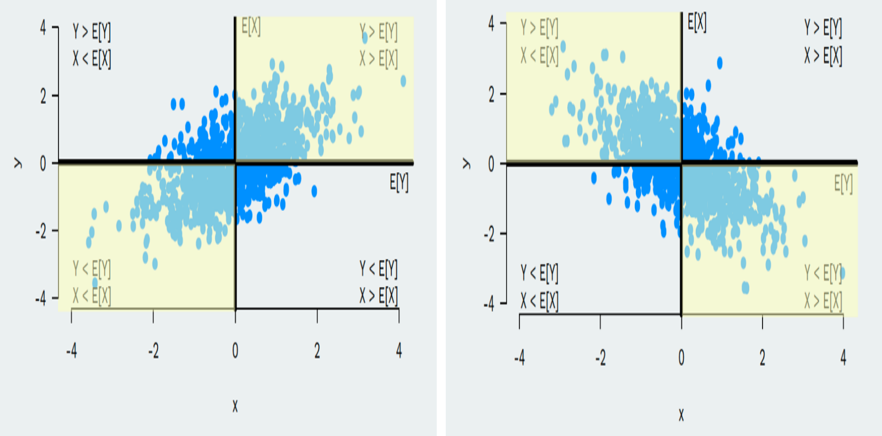
the covariance between X and Y is defined as

$$Cov[X, Y] = E\left[\left(X - E[X]\right)\left(Y - E[Y]\right)\right]$$

- Properties of covariances:
  - $Cov[X, Y] = E[XY] - E[X]E[Y]$
  - If $X \perp Y$, $Cov[X, Y] = 0$

Intuition of Covariance

- The conditional probability mass function(conditional p.m.f) of Y conditional of X is



Intuition of Covariance

- The conditional probability mass function(conditional p.m.f) of Y conditional of X is

Covariance and Independence

- Properties of covariances:
  - $Cov[X, Y] = E[XY] - E[X]E[Y]$
  - $Cov[aX + b, cY + d] = acCoV[XY]$
  - $Cov[X, X] = Var[X]$
- Covariance and Independence
  - If $X \perp Y$, then $Cov[X, Y] = 0$, thus independence $\Rightarrow Cov[X, Y] = 0$.
  - If $Cov[X, Y] = 0$, then $X \perp Y$? **NO!** $Cov[X, Y] = 0 \nRightarrow$ independence.

Covariance and Independence

- Properties of covariances:
  - $Cov[X, Y] = E[XY] - E[X]E[Y]$
  - $Cov[aX + b, cY + d] = acCoV[XY]$
  - $Cov[X, X] = Var[X]$
- Covariance and Independence
  - If $X \perp Y$, then $Cov[X, Y] = 0$. thus independence $\Rightarrow Cov[X, Y] = 0$.
  - If $Cov[X, Y] = 0$, then $X \perp Y$? **NO!** $Cov[X, Y] = 0 \nRightarrow$ independence.

Covariance and Independence

- Properties of covariances:
  - $Cov[X, Y] = E[XY] - E[X]E[Y]$
  - $Cov[aX + b, cY + d] = acCoV[XY]$
  - $Cov[X, X] = Var[X]$
- Covariance and Independence
  - If $X \perp Y$, then $Cov[X, Y] = 0$, thus independence $\Rightarrow Cov[X, Y] = 0$.
  - If $Cov[X, Y] = 0$, then $X \perp Y$? **NO!** $Cov[X, Y] = 0 \not\Rightarrow$ independence.

Covariance and Independence

- Properties of covariances:
  - $Cov[X, Y] = E[XY] - E[X]E[Y]$
  - $Cov[aX + b, cY + d] = acCoV[XY]$
  - $Cov[X, X] = Var[X]$
- Covariance and Independence
  - If $X \perp Y$, then $Cov[X, Y] = 0$. thus independence $\Rightarrow Cov[X, Y] = 0$.
  - If $Cov[X, Y] = 0$, then $X \perp Y$? **NO!** $Cov[X, Y] = 0 \nRightarrow$ independence.

Covariance and Independence

- Properties of covariances:
  - $Cov[X, Y] = E[XY] - E[X]E[Y]$
  - $Cov[aX + b, cY + d] = acCoV[XY]$
  - $Cov[X, X] = Var[X]$
- Covariance and Independence
  - If $X \perp Y$, then $Cov[X, Y] = 0$. thus independence $\Rightarrow Cov[X, Y] = 0$.
  - If $Cov[X, Y] = 0$, then $X \perp Y$? **NO!** $Cov[X, Y] = 0 \not\Rightarrow$ independence.

Covariance and Independence

- Properties of covariances:
  - $Cov[X, Y] = E[XY] - E[X]E[Y]$
  - $Cov[aX + b, cY + d] = acCoV[XY]$
  - $Cov[X, X] = Var[X]$
- Covariance and Independence
  - If $X \perp Y$, then $Cov[X, Y] = 0$. thus independence $\Rightarrow Cov[X, Y] = 0$.
  - If $Cov[X, Y] = 0$, then $X \perp Y$? **NO!** $Cov[X, Y] = 0 \nRightarrow$ independence.

Covariance and Independence

- Properties of covariances:
  - $Cov[X, Y] = E[XY] - E[X]E[Y]$
  - $Cov[aX + b, cY + d] = acCoV[XY]$
  - $Cov[X, X] = Var[X]$
- Covariance and Independence
  - If $X \perp Y$, then $Cov[X, Y] = 0$. thus independence $\Rightarrow Cov[X, Y] = 0$.
  - If $Cov[X, Y] = 0$, then $X \perp Y$? **NO!** $Cov[X, Y] = 0 \not\Rightarrow$ independence.

Covariance and Correlation

- Covariance is not scale-free. **Correlation** is a special form of covariance after dividing out the scales of the respective variables.

Correlation

The correlation between X and Y is defined as

$$\rho_{XY} = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}$$

- Correlation properties:
  - $-1 \leq \rho \leq 1$
  - If $|\rho_{XY}| = 1$, then X and Y are perfectly correlated with a linear relationship: $Y = a + bX$

Covariance and Correlation

- Covariance is not scale-free. Correlation is a special form of covariance after dividing out the scales of the respective variables.

Correlation

The correlation between X and Y is defined as

$$\rho_{XY} = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}$$

- Correlation properties:
  - $-1 \leq \rho \leq 1$
  - If $|\rho_{XY}| = 1$, then X and Y are perfectly correlated with a linear relationship: $Y = a + bX$

Covariance and Correlation

- Covariance is not scale-free. Correlation is a special form of covariance after dividing out the scales of the respective variables.

Correlation

The correlation between X and Y is defined as

$$\rho_{XY} = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}$$

- Correlation properties:
  - $-1 \leq \rho \leq 1$
  - If $\mid \rho_{XY} \mid = 1$, then X and Y are perfectly correlated with a linear relationship: $Y = a + bX$

Covariance and Correlation

- Covariance is not scale-free. Correlation is a special form of covariance after dividing out the scales of the respective variables.

Correlation

The correlation between X and Y is defined as

$$\rho_{XY} = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}$$

- Correlation properties:
  - $-1 \leq \rho \leq 1$
  - If $|\rho_{XY}| = 1$, then X and Y are perfectly correlated with a linear relationship: $Y = a + bX$

Covariance and Correlation

- Covariance is not scale-free. Correlation is a special form of covariance after dividing out the scales of the respective variables.

Correlation

The correlation between X and Y is defined as

$$\rho_{XY} = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}$$

- Correlation properties:
  - $-1 \leq \rho \leq 1$
  - If $|\rho_{XY}| = 1$, then X and Y are perfectly correlated with a linear relationship: $Y = a + bX$

Conditional Distributions

Conditional Probability function

- The conditional probability mass functional(conditional p.m.f) of Y conditional of X is

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

TABLE 2.3 Joint and Conditional Distributions of Computer Crashes (*M*) and Computer Age (*A*)

A. Joint Distribution

	$M=0$	$M=1$	$M=2$	$M=3$	$M=4$	Total
Old computer ($A = 0$)	0.35	0.065	0.05	0.025	0.01	0.50
New computer ($A = 1$)	0.45	0.035	0.01	0.005	0.00	0.50
Total	0.80	0.10	0.06	0.03	0.01	1.00

B. Conditional Distributions of *M* given *A*

	$M=0$	$M=1$	$M=2$	$M=3$	$M=4$	Total
$\Pr(M \mid A = 0)$	0.70	0.13	0.10	0.05	0.02	1.00
$\Pr(M \mid A = 1)$	0.90	0.07	0.02	0.01	0.00	1.00

Conditional Density Function

Conditional probability density function:

c.d.f. of Y conditional on X is

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

- Based on the definition of the conditional p.m.f./p.d.f., we have the following equation

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x)$$

Conditional Density Function

Conditional probability density function:

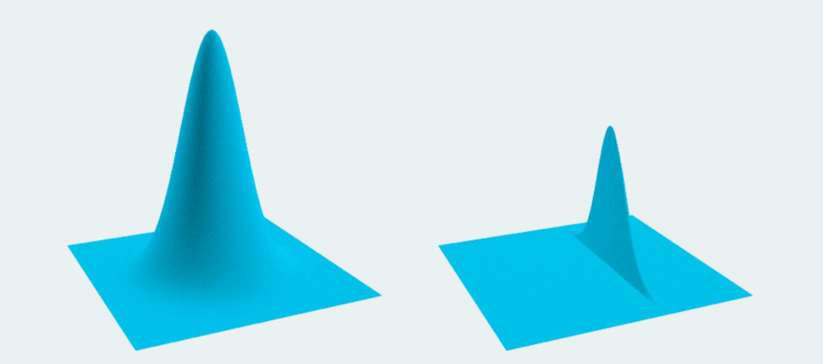
c.d.f. of Y conditional on X is

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

- Based on the definition of the conditional p.m.f./p.d.f., we have the following equation

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x)$$

Conditional Density Function



- c.d.f is proportional to joint p.d.f along x_0 like a slice of total volume.

Conditional Independence

Conditional Independence

X and Y are conditional Independent given Z, denoted as $X \perp Y \mid Z$, if

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)\, f_{Y|Z}(y|z)$$

- X and Y are independent within levels of Z.
- Example:
  - X = swimming accidents, Y = ice cream sold.
  - In general, two variable is highly correlated.
  - If conditional on Z = temperature, then they are independent.

Conditional Independence

Conditional Independence

X and Y are conditional Independent given Z, denoted as $X \perp Y \mid Z$, if

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)\, f_{Y|Z}(y|z)$$

- X and Y are independent within levels of Z.
- Example:
  - X = swimming accidents, Y = ice cream sold.
  - In general, two variable is highly correlated.
  - If conditional on Z = temperature, then they are independent.

Conditional Independence

Conditional Independence

X and Y are conditional Independent given Z, denoted as $X \perp Y \mid Z$, if

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)\, f_{Y|Z}(y|z)$$

- X and Y are independent within levels of Z.
- Example:
  - $X =$ swimming accidents, $Y =$ ice cream sold.
  - In general, two variable is highly correlated.
  - If conditional on $Z =$ temperature, then they are independent.

Conditional Independence

Conditional Independence

X and Y are conditional Independent given Z, denoted as $X \perp Y \mid Z$, if

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)\, f_{Y|Z}(y|z)$$

- X and Y are independent within levels of Z.
- Example:
  - $X =$ swimming accidents, $Y =$ ice cream sold.
  - In general, two variable is highly correlated.
  - If conditional on $Z =$ temperature, then they are independent.

Conditional Independence

Conditional Independence

X and Y are conditional Independent given Z, denoted as $X \perp Y \mid Z$, if

$$f_{X,Y|Z}(x,y|z) = f_{X|Z}(x|z)\, f_{Y|Z}(y|z)$$

- X and Y are independent within levels of Z.
- Example:
  - $X =$ swimming accidents, $Y =$ ice cream sold.
  - In general, two variable is highly correlated.
  - If conditional on $Z =$ temperature, then they are independent.

Conditional Independence

Conditional Independence

X and Y are conditional Independent given Z, denoted as $X \perp Y \mid Z$, if

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)\, f_{Y|Z}(y|z)$$

- X and Y are independent within levels of Z.
- Example:
  - $X =$ swimming accidents, $Y =$ ice cream sold.
  - In general, two variable is highly correlated.
  - If conditional on $Z =$ temperature, then they are independent.

Conditional Expectation Function

Conditional Expectation

Conditional on X, Y's Conditional Expectation is

$$E(Y|X) = \begin{cases} \sum y P_{Y|X}(y|x) & \textit{discrete } Y \\ \int y f_{Y|X}(y|x) dy & \textit{continuous } Y \end{cases}$$

- **Conditional Expectation Function(CEF)** is a function of x, since X is a random variable, so CEF is also a random variable.
- Intuition：期望就是求平均值，而条件期望就是“分组取平均”或“在... 条件下的均值”。

Conditional Expectation Function

Conditional Expectation

Conditional on X, Y's Conditional Expectation is

$$E(Y|X) = \begin{cases} \sum y P_{Y|X}(y|x) & discrete\, Y \\ \int y f_{Y|X}(y|x) dy & continuous\, Y \end{cases}$$

- **Conditional Expectation Function(CEF)** is a function of x, since X is a random variable, so CEF is also a random variable.
- Intuition：期望就是求平均值，而条件期望就是“分组取平均”或“在... 条件下的均值”。

Conditional Expectation Function

Conditional Expectation

Conditional on X, Y's Conditional Expectation is

$$E(Y|X) = \begin{cases} \sum y P_{Y|X}(y|x) & \textit{discrete } Y \\ \int y f_{Y|X}(y|x) dy & \textit{continuous } Y \end{cases}$$

- **Conditional Expectation Function(CEF)** is a function of x, since X is a random variable, so CEF is also a random variable.
- Intuition：期望就是求平均值，而条件期望就是"分组取平均"或"在... 条件下的均值"。

Properties of Conditional Expectation

- Let X, Y, Z are random variables; $a, b \in \mathbb{R}$; $g(\cdot)$ is a real valued function, then we have

$$E[a \mid Y] = a$$

- $E[(aX + bZ) \mid Y] = aE[X \mid Y] + bE[Z \mid Y]$
- $E[c(X) \mid X] = c(X)$ for any function $c(X)$.Thus if we know X, then we also know $c(X)$.
  - eg. $E[(X^2 + 2X^3) \mid X] = X^2 + 2X^3$
- if X and Y are independent r.v.s, then

$$E[Y \mid X = x] = E[Y]$$

- if X and Y independent conditional on Z, thus $X \perp Y \mid Z$,

$$E[Y \mid X = x, Z = z] = E[Y \mid Z = z]$$

Properties of Conditional Expectation

- Let X, Y, Z are random variables; $a, b \in \mathbb{R}$; $g(\cdot)$ is a real valued function, then we have
$$E[a \mid Y] = a$$
- $E[(aX + bZ) \mid Y] = aE[X \mid Y] + bE[Z \mid Y]$
- $E[c(X) \mid X] = c(X)$ for any function $c(X)$.Thus if we know X, then we also know $c(X)$.
  - eg. $E[(X^2 + 2X^3) \mid X] = X^2 + 2X^3$
- if X and Y are independent r.v.s, then
$$E[Y \mid X = x] = E[Y]$$
- if X and Y independent conditional on Z, thus $X \perp Y \mid Z$,
$$E[Y \mid X = x, Z = z] = E[Y \mid Z = z]$$

Properties of Conditional Expectation

- Let X, Y, Z are random variables; $a, b \in \mathbb{R}$; $g(\cdot)$ is a real valued function, then we have
$$E[a \mid Y] = a$$
- $E[(aX + bZ) \mid Y] = aE[X \mid Y] + bE[Z \mid Y]$
- $E[c(X) \mid X] = c(X)$ for any function $c(X)$.Thus if we know X, then we also know $c(X)$.
  - eg. $E[(X^2 + 2X^3) \mid X] = X^2 + 2X^3$
- if X and Y are independent r.v.s, then
$$E[Y \mid X = x] = E[Y]$$
- if X and Y independent conditional on Z, thus $X \perp Y \mid Z$,
$$E[Y \mid X = x, Z = z] = E[Y \mid Z = z]$$

Properties of Conditional Expectation

- Let X, Y, Z are random variables; $a, b \in \mathbb{R}$; $g(\cdot)$ is a real valued function, then we have
$$E[a \mid Y] = a$$
- $E[(aX + bZ) \mid Y] = aE[X \mid Y] + bE[Z \mid Y]$
- $E[c(X) \mid X] = c(X)$ for any function $c(X)$.Thus if we know X, then we also know $c(X)$.
  - eg. $E[(X^2 + 2X^3) \mid X] = X^2 + 2X^3$
- if X and Y are independent r.v.s, then
$$E[Y \mid X = x] = E[Y]$$
- if X and Y independent conditional on Z, thus $X \perp Y \mid Z$,
$$E[Y \mid X = x, Z = z] = E[Y \mid Z = z]$$

Properties of Conditional Expectation

- Let X, Y, Z are random variables; $a, b \in \mathbb{R}$; $g(\cdot)$ is a real valued function, then we have
$$E[a \mid Y] = a$$
- $E[(aX + bZ) \mid Y] = aE[X \mid Y] + bE[Z \mid Y]$
- $E[c(X) \mid X] = c(X)$ for any function $c(X)$.Thus if we know X, then we also know $c(X)$.
  - eg. $E[(X^2 + 2X^3) \mid X] = X^2 + 2X^3$
- if X and Y are independent r.v.s, then
$$E[Y \mid X = x] = E[Y]$$
- if X and Y independent conditional on Z, thus $X \perp Y \mid Z$,
$$E[Y \mid X = x, Z = z] = E[Y \mid Z = z]$$

Properties of Conditional Expectation

- Let X, Y, Z are random variables; $a, b \in \mathbb{R}$; $g(\cdot)$ is a real valued function, then we have
$$E[a \mid Y] = a$$
- $E[(aX + bZ) \mid Y] = aE[X \mid Y] + bE[Z \mid Y]$
- $E[c(X) \mid X] = c(X)$ for any function $c(X)$.Thus if we know X, then we also know $c(X)$.
  - eg. $E[(X^2 + 2X^3) \mid X] = X^2 + 2X^3$
- if X and Y are independent r.v.s, then
$$E[Y \mid X = x] = E[Y]$$
- if X and Y independent conditional on Z, thus $X \perp Y \mid Z$,
$$E[Y \mid X = x, Z = z] = E[Y \mid Z = z]$$

The Law of Iterated Expectations(LIE)

- It states that an unconditional expectation can be written as the unconditional average of conditional expectation function.

$$E(Y) = E[E(Y|X)]$$

- And if $g(x)$ and $h(Y)$ are a real value functions then it can easily extend to

$$E(g(X)h(Y)) = E[E(g(X)h(Y)|X)] = E[g(X)E(h(Y)|X)]$$

The Law of Iterated Expectations(LIE)

- It states that an unconditional expectation can be written as the unconditional average of conditional expectation function.

$$E(Y) = E[E(Y|X)]$$

- And if $g(x)$ and $h(Y)$ are a real value functions then it can easily extend to

$$E(g(X)h(Y)) = E[E(g(X)h(Y)|X)] = E[g(X)E(h(Y)|X)]$$

The Law of Iterated Expectations(LIE)

Proof.

$$
\begin{aligned}
E[E(Y|X)] &= \int E[Y \mid X = u] g_X(u) du \\
&= \int \left[\int t f_Y(t \mid X = u) dt \right] g_X(u) du \\
&= \iint t f_Y(t \mid X = u) g_X(u) dt\, du \\
&= \int t \left[\int f_Y(t \mid X = u) g_X(u) du \right] dt \\
&= \int t \left[\int f_{XY}(t, u) du \right] dt \\
&= \int t f_Y(t) dt = E(Y)
\end{aligned}
$$

Conditional Variance

- Conditional on X, Y's Conditional Variance is defined as

$$Var(Y|X) = E\left[(Y - E[Y|X])^2 \mid X\right]$$

- Usual variance formula applied to conditional distribution.
  - Discrete

$$V[Y \mid X] = \sum_y (y - E[Y \mid X])^2 f_{Y|X}(y|x)$$

  - Continuous

$$V[Y \mid X] = \int_y (y - E[Y \mid X])^2 f_{Y|X}(y|x)$$

Conditional Variance

- Conditional on X, Y's Conditional Variance is defined as

$$Var(Y|X) = E\left[(Y - E[Y|X])^2 \mid X\right]$$

- Usual variance formula applied to conditional distribution.
  - Discrete

$$V[Y \mid X] = \sum_y (y - E[Y \mid X])^2 f_{Y|X}(y|x)$$

  - Continuous

$$V[Y \mid X] = \int_y (y - E[Y \mid X])^2 f_{Y|X}(y|x)$$

Conditional Variance

- Conditional on X, Y's Conditional Variance is defined as

$$Var(Y|X) = E\left[(Y - E[Y|X])^2 \mid X\right]$$

- Usual variance formula applied to conditional distribution.
  - Discrete

$$V[Y \mid X] = \sum_y (y - E[Y \mid X])^2 f_{Y|X}(y|x)$$

  - Continuous

$$V[Y \mid X] = \int_y (y - E[Y \mid X])^2 f_{Y|X}(y|x)$$

Conditional Variance

- Conditional on X, Y's Conditional Variance is defined as

$$Var(Y|X) = E\left[(Y - E[Y|X])^2 \mid X\right]$$

- Usual variance formula applied to conditional distribution.
  - Discrete

$$V[Y \mid X] = \sum_{y}(y - E[Y \mid X])^2 f_{Y|X}(y|x)$$

  - Continuous

$$V[Y \mid X] = \int_{y}(y - E[Y \mid X])^2 f_{Y|X}(y|x)$$

Several Famous Distributions

Families of distributions

- There are several important families of distributions:
  - The p.m.f./p.d.f. within the family has the same form, with parameters that might vary across the family.
  - The parameters determine the shape of the distribution
- Statistical modeling in a nutshell: to study probability distribution function.
  - Assume the data, $X_1, X_2, \ldots, X_n$, are independent draws from a common distribution $f_\theta(x)$ within a family of distributions (normal, poisson, etc)
  - Use a function of the observed data to estimate the value of the $\theta : \hat{\theta}(X_1, X_2, \ldots, X_n)$

Families of distributions

- There are several important families of distributions:
    - The p.m.f./p.d.f. within the family has the same form, with parameters that might vary across the family.
    - The parameters determine the shape of the distribution
- Statistical modeling in a nutshell: to study probability distribution function.
    - Assume the data, $X_1, X_2, \ldots, X_n$, are independent draws from a common distribution $f_\theta(x)$ within a family of distributions (normal, poisson, etc)
    - Use a function of the observed data to estimate the value of the $\theta : \hat{\theta}(X_1, X_2, \ldots, X_n)$

Families of distributions

- There are several important families of distributions:
    - The p.m.f./p.d.f. within the family has the same form, with parameters that might vary across the family.
    - The parameters determine the shape of the distribution
- Statistical modeling in a nutshell: to study probability distribution function.
    - Assume the data, $X_1, X_2, \ldots, X_n$, are independent draws from a common distribution $f_\theta(x)$ within a family of distributions (normal, poisson, etc)
    - Use a function of the observed data to estimate the value of the $\theta : \hat{\theta}(X_1, X_2, \ldots, X_n)$

Families of distributions

- There are several important families of distributions:
  - The p.m.f./p.d.f. within the family has the same form, with parameters that might vary across the family.
  - The parameters determine the shape of the distribution
- Statistical modeling in a nutshell: to study probability distribution function.
  - Assume the data, $X_1, X_2, ..., X_n$, are independent draws from a common distribution $f_\theta(x)$ within a family of distributions (normal, poisson, etc)
  - Use a function of the observed data to estimate the value of the $\theta : \hat{\theta}(X_1, X_2, ..., X_n)$

Families of distributions

- There are several important families of distributions:
  - The p.m.f./p.d.f. within the family has the same form, with parameters that might vary across the family.
  - The parameters determine the shape of the distribution
- Statistical modeling in a nutshell: to study probability distribution function.
  - Assume the data, $X_1, X_2, ..., X_n$, are independent draws from a common distribution $f_\theta(x)$ within a family of distributions (normal, poisson, etc)
  - Use a function of the observed data to estimate the value of the $\theta : \hat{\theta}(X_1, X_2, ..., X_n)$

Families of distributions

- There are several important families of distributions:
  - The p.m.f./p.d.f. within the family has the same form, with parameters that might vary across the family.
  - The parameters determine the shape of the distribution
- Statistical modeling in a nutshell: to study probability distribution function.
  - Assume the data, $X_1, X_2, ..., X_n$, are independent draws from a common distribution $f_\theta(x)$ within a family of distributions (normal, poisson, etc)
  - Use a function of the observed data to estimate the value of the $\theta : \hat{\theta}(X_1, X_2, ..., X_n)$

The Bernoulli Distribution

Definition

X has a Bernoulli distribution if it have a binary values $X \in \{0, 1\}$ and its probability mass function is

$$f_X(x) = P(X = x) = \begin{cases} p & if x = 1 \\ 1 - p & if x = 0 \end{cases}$$

Question:

What is the *Expectation and Variance* of X?

$$E(X) = \sum_{j=1}^{k} x_j p_j = 0 \times (1 - p) + 1 \times p = p$$

$$Var(X) = E[X - E(X)]^2 = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p)$$

The Bernoulli Distribution

Definition

X has a Bernoulli distribution if it have a binary values $X \in \{0, 1\}$ and its probability mass function is

$$f_X(x) = P(X = x) = \begin{cases} p & if x = 1 \\ 1 - p & if x = 0 \end{cases}$$

Question:

What is the *Expectation and Variance* of X?

$$E(X) = \sum_{j=1}^{k} x_j p_j = 0 \times (1 - p) + 1 \times p = p$$

$$Var(X) = E[X - E(X)]^2 = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p)$$

The Normal Distribution

- The p.d.f of a normal random variable X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], \ -\infty < X < +\infty$$

- if X is normally distributed with expected value μ and variance σ^2, denoted as $X \sim N(\mu, \sigma^2)$
  - if we know these two parameters, we know everything about the distribution.
- Examples: Human heights, weights, test scores,
- If X represents wage, income or consumption etc, it will has a log-normal distribution, thus

$$log(X) \sim N(\mu, \sigma^2)$$

The Normal Distribution

- The p.d.f of a normal random variable X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], \ -\infty < X < +\infty$$

- if X is normally distributed with expected value μ and variance σ^2, denoted as $X \sim N(\mu, \sigma^2)$
  - if we know these two parameters, we know everything about the distribution.
- Examples: Human heights, weights, test scores,
- If X represents wage, income or consumption etc, it will has a log-normal distribution, thus

$$log(X) \sim N(\mu, \sigma^2)$$

The Normal Distribution

- The p.d.f of a normal random variable X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], \ -\infty < X < +\infty$$

- if X is normally distributed with expected value μ and variance σ^2, denoted as $X \sim N(\mu, \sigma^2)$
  - if we know these two parameters, we know everything about the distribution.
- Examples: Human heights, weights, test scores,
- If X represents wage, income or consumption etc, it will has a log-normal distribution, thus

$$log(X) \sim N(\mu, \sigma^2)$$

The Normal Distribution

- The p.d.f of a normal random variable X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], \ -\infty < X < +\infty$$

- if X is normally distributed with expected value μ and variance σ^2, denoted as $X \sim N(\mu, \sigma^2)$
  - if we know these two parameters, we know everything about the distribution.
- Examples: Human heights, weights, test scores,
- If X represents wage, income or consumption etc, it will has a log-normal distribution, thus

$$log(X) \sim N(\mu, \sigma^2)$$

The Normal Distribution

- The p.d.f of a normal random variable X is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right],\ -\infty < X < +\infty$$

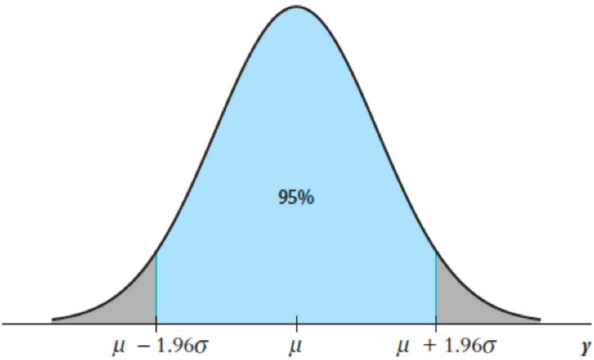
- if X is normally distributed with expected value μ and variance σ^2, denoted as $X \sim N(\mu, \sigma^2)$
  - if we know these two parameters, we know everything about the distribution.
- Examples: Human heights, weights, test scores,
- If X represents wage, income or consumption etc, it will has a log-normal distribution, thus

$$log(X) \sim N(\mu, \sigma^2)$$

The Normal Distribution

FIGURE 2.5 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 $N(\mu, \sigma^2)$.



The Standard Normal Distribution

- A special case of the normal distribution where the mean is zero ($\mu = 0$) and the variance is one ($\sigma^2 = \sigma = 1$), then its p.d.f is

$$f_X(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \ -\infty < X < +\infty$$

- if X is standard normally distributed, then denoted as $X \sim N(0, 1)$
- The standard normal cumulative distribution function is denoted

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

where z is a standardize r.v. thus $z = \frac{x - \mu_X}{\sigma_X}$

The Standard Normal Distribution

- A special case of the normal distribution where the mean is zero $(\mu = 0)$ and the variance is one $(\sigma^2 = \sigma = 1)$, then its p.d.f is

$$f_X(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \; -\infty < X < +\infty$$

- if X is standard normally distributed, then denoted as $X \sim N(0,1)$
- The standard normal cumulative distribution function is denoted

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

where z is a standardize r.v. thus $z = \frac{x - \mu_X}{\sigma_X}$

The Standard Normal Distribution

- A special case of the normal distribution where the mean is zero $(\mu = 0)$ and the variance is one $(\sigma^2 = \sigma = 1)$, then its p.d.f is

$$f_X(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \ -\infty < X < +\infty$$

- if X is standard normally distributed, then denoted as $X \sim N(0, 1)$
- The standard normal cumulative distribution function is denoted

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

where z is a standardize r.v. thus $z = \frac{x - \mu_X}{\sigma_X}$

The Standard Normal Distribution

FIGURE B.8 The standard normal cumulative distribution function.

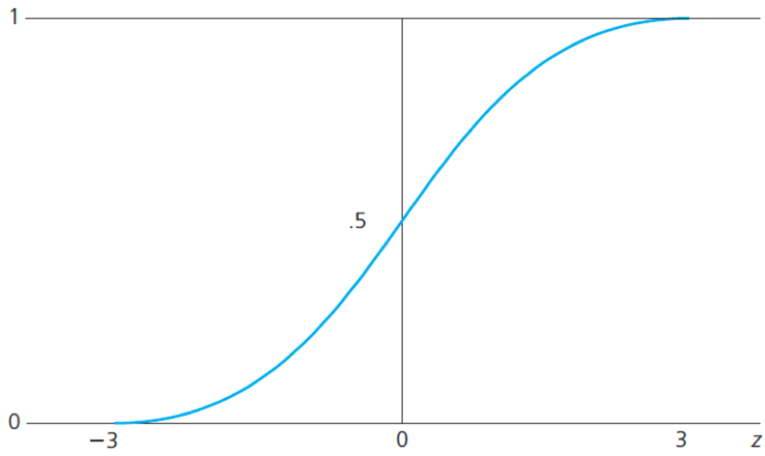
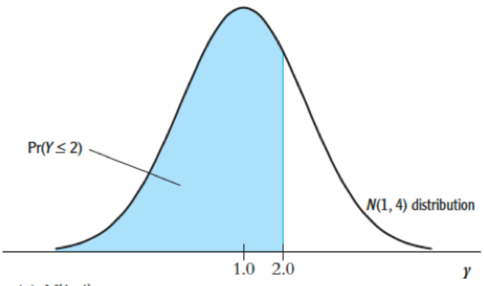
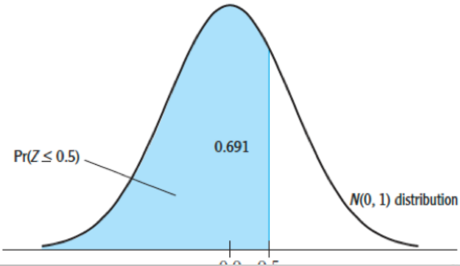


FIGURE 2.6 Calculating the Probability That $Y \le 2$ When Y Is Distributed $N(1, 4)$

To calculate $\Pr(Y \le 2)$, standardize Y, then use the standard normal distribution table. Y is standardized by subtracting its mean ($\mu = 1$) and dividing by its standard deviation ($\sigma = 2$). The probability that $Y \le 2$ is shown in Figure 2.6a, and the corresponding probability after standardizing Y is shown in Figure 2.6b. Because the standardized random variable, $(Y - 1)/2$, is a standard normal (Z) random variable, $\Pr(Y \le 2) = \Pr\left(\frac{Y-1}{2} \le \frac{2-1}{2}\right) = \Pr(Z \le 0.5)$. From Appendix Table 1, $\Pr(Z \le 0.5) = \Phi(0.5) = 0.691$.



(a) $N(1, 4)$



The Chi-Square Distribution

- Let $Z_i (i = 1, 2, ..., m)$ be independent random variables, each distributed as **standard normal**. Then a new random variable can be defined as the sum of the squares of Z_i :

$$X = \sum_{i=1}^{m} Z_i^2$$

Then X has a **chi-squared distribution** with m **degrees of freedom**

- The form of the distribution varies with the number of degrees of freedom, i.e. the number of standard normal random variables Z_i included in X.
- The distribution has a long tail, or is skewed, to the right. As the degrees of freedom m gets larger, however, the distribution becomes more symmetric and ‘‘bell-shaped.’’ In fact, as m gets larger, the chi-square distribution converges to, and essentially becomes, a **normal distribution**.

The Chi-Square Distribution

- Let $Z_i (i = 1, 2, ..., m)$ be independent random variables, each distributed as **standard normal**. Then a new random variable can be defined as the sum of the squares of Z_i :

$$X = \sum_{i=1}^{m} Z_i^2$$

Then X has a **chi-squared distribution** with m **degrees of freedom**

- The form of the distribution varies with the number of degrees of freedom, i.e. the number of standard normal random variables Z_i included in X.
- The distribution has a long tail, or is skewed, to the right. As the degrees of freedom m gets larger, however, the distribution becomes more symmetric and ‘‘bell-shaped.’’ In fact, as m gets larger, the chi-square distribution converges to, and essentially becomes, a **normal distribution**.

The Chi-Square Distribution

- Let $Z_i (i = 1, 2, ..., m)$ be independent random variables, each distributed as **standard normal**. Then a new random variable can be defined as the sum of the squares of Z_i :

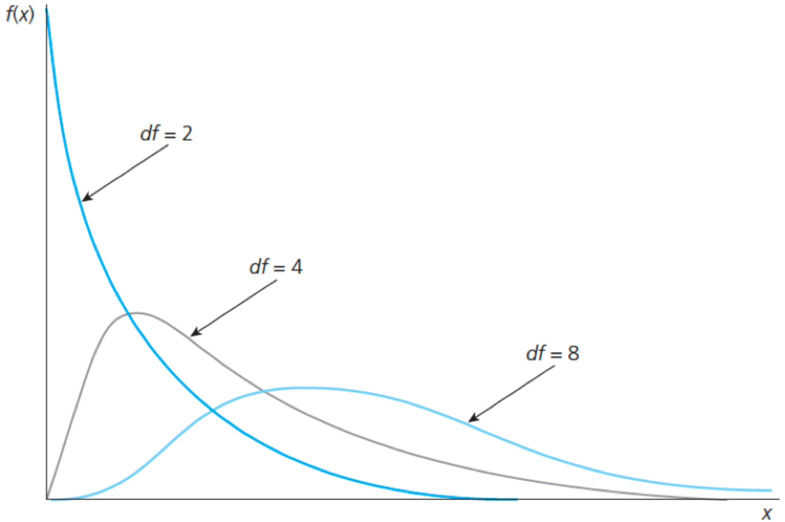
$$X = \sum_{i=1}^{m} Z_i^2$$

Then X has a **chi-squared distribution** with m **degrees of freedom**

- The form of the distribution varies with the number of degrees of freedom, i.e. the number of standard normal random variables Z_i included in X.
- The distribution has a long tail, or is skewed, to the right. As the degrees of freedom m gets larger, however, the distribution becomes more symmetric and ‘‘bell-shaped.’’ In fact, as m gets larger, the chi-square distribution converges to, and essentially becomes, a **normal distribution**.

The Chi-Square Distribution

FIGURE B.9 The chi-square distribution with various degrees of freedom.



The Student t Distribution

- The Student t distribution can be obtained from a standard normal and a chi-square random variable.
- Let Z have a standard normal distribution, let X have a chi-square distribution with m degrees of freedom and assume that Z and X are independent. Then the random variable

$$T = \frac{Z}{\sqrt{X/n}}$$

has has a t-distribution with m degrees of freedom, denoted as $T \sim t_n$.

- The shape of the t-distribution is similar to that of a normal distribution, except that the t-distribution has more probability mass in the tails.
- As the degrees of freedom get large, the t-distribution approaches **the standard normal distribution**.

The Student t Distribution

- The Student t distribution can be obtained from a standard normal and a chi-square random variable.
- Let Z have a standard normal distribution, let X have a chi-square distribution with m degrees of freedom and assume that Z and X are independent. Then the random variable

$$T = \frac{Z}{\sqrt{X/n}}$$

  has has a t-distribution with m degrees of freedom, denoted as $T \sim t_n$.
- The shape of the t-distribution is similar to that of a normal distribution, except that the t-distribution has more probability mass in the tails.
- As the degrees of freedom get large, the t-distribution approaches **the standard normal distribution**.

The Student t Distribution

- The Student t distribution can be obtained from a standard normal and a chi-square random variable.
- Let Z have a standard normal distribution, let X have a chi-square distribution with m degrees of freedom and assume that Z and X are independent. Then the random variable

$$T = \frac{Z}{\sqrt{X/n}}$$

has has a t-distribution with m degrees of freedom, denoted as $T \sim t_n$.

- The shape of the t-distribution is similar to that of a normal distribution, except that the t-distribution has more probability mass in the tails.
- As the degrees of freedom get large, the t-distribution approaches **the standard normal distribution**.

The Student t Distribution

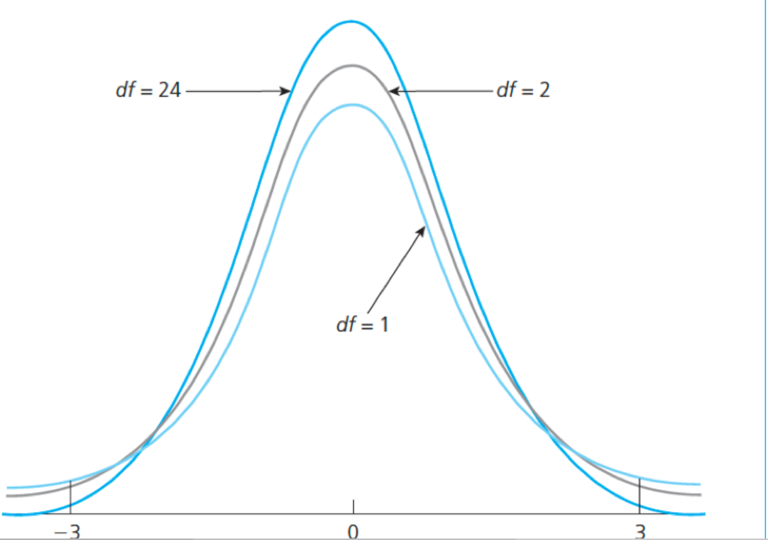
- The Student t distribution can be obtained from a standard normal and a chi-square random variable.
- Let Z have a standard normal distribution, let X have a chi-square distribution with m degrees of freedom and assume that Z and X are independent. Then the random variable

$$T = \frac{Z}{\sqrt{X/n}}$$

  has has a t-distribution with m degrees of freedom, denoted as $T \sim t_n$.
- The shape of the t-distribution is similar to that of a normal distribution, except that the t-distribution has more probability mass in the tails.
- As the degrees of freedom get large, the t-distribution approaches **the standard normal distribution**.

The Student t Distribution

FIGURE B.10 The *t* distribution with various degrees of freedom.



The F Distribution

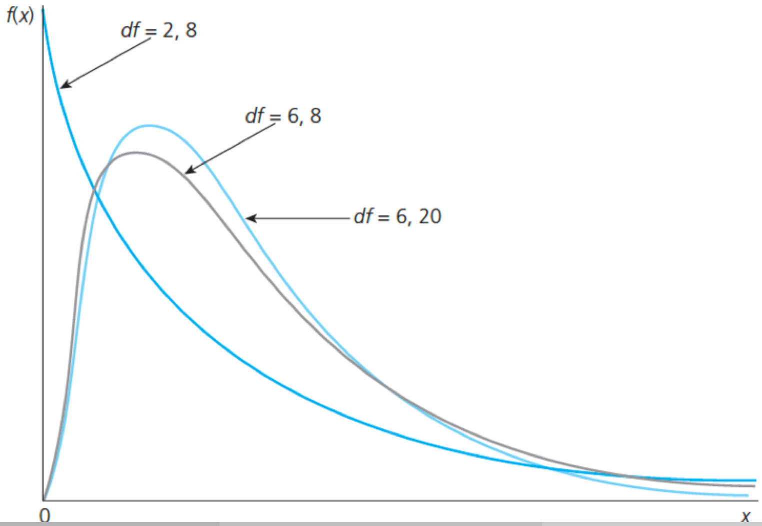
- Let $X_1 \sim \chi^2_m$and $X_2 \sim \chi^2_n$, and assume that X_1 and X_2 are independent,

$$Z = \frac{\frac{X_1}{m}}{\frac{X_2}{n}} \sim F_{m,n}$$

thus Z has an F-distribution with (m, n) degrees of freedom.

The F Distribution

FIGURE B.11 The F_{k_1,k_2} distribution for various degrees of freedom, k_1 and k_2.



In Summary