Review A: Probability

Professor Ji-Woong Chung
Korea University

Outline

Outline

Sample Space & Events

Probability theory starts with the idea of an experiment.

The sample space, denoted Ω, is the set of possible outcomes of an experiment.

Realizations (or outcomes) of the experiments are points in the sample space, $\omega \in \Omega$.

Collections of realizations are called events $E \subseteq \Omega$.

Example: Consider tossing a coin twice. Then $\Omega = \{HH, HT, TH, TT\}$, where for example $\omega = HH$ is the outcome of landing heads twice. The event of the first toss being tail is $E = \{TH, TT\}$.

Indicator Functions

Definition (Indicator Function)

Let Ω be a sample space and $E \subseteq \Omega$ denote an event. The indicator function of E is defined as

$$\mathbb{1}\{\omega \in E\} = \mathbb{1}_E(\omega) = \begin{cases} 1, & \text{if } \omega \in E, \\ 0, & \text{if } \omega \notin E, \end{cases} \quad \forall \omega \in \Omega.$$

Example: Consider tossing a fair coin twice as in the previous example. Let $E_1 = \{TT\}$ and $E_2 = \{TH, TT\}$. We have

$$\mathbb{1}\{TT \in E_1\} = \mathbb{1}\{TT \in E_2\} = \mathbb{1}\{TH \in E_2\} = 1$$

$$\mathbb{1}\{TH \in E_1\} = 0.$$

Indicator Functions (Contd.)

Indicator functions allow us to succinctly express "yes-or-no" questions. As these questions become more convoluted, this approach proves helpful thanks to a few key properties of indicator functions:

Lemma

Let Ω be a sample space and $E_1, E_2 \subseteq \Omega$ denote two events. The following hold $\forall \omega \in \Omega$:

1. $\mathbb{1}\{\omega \in E_1\}^k = \mathbb{1}\{\omega \in E_1\}, \forall k \in \mathbb{R} \setminus \{0\}$;
2. $\mathbb{1}\{\omega \notin E_1\} = 1 - \mathbb{1}\{\omega \in E_1\}$;
3. $\mathbb{1}\{\omega \in E_1 \cap E_2\} = \mathbb{1}\{\omega \in E_1\}\mathbb{1}\{\omega \in E_2\}$;
4. $\mathbb{1}\{\omega \in E_1 \cup E_2\} = \mathbb{1}\{\omega \in E_1\} + \mathbb{1}\{\omega \in E_2\} - \mathbb{1}\{\omega \in E_1 \cap E_2\}$.

Probabilities

Probabilities characterize the likelihood of an event in a sample space.

Definition (Probability Measure)

A probability measure on Ω is a function $P : \Omega \rightarrow [0, 1]$ satisfying:

1. $P(\Omega) = 1$;
2. $P(E) \geq 0, \forall E \subseteq \Omega$;
3. $P(E_1 \cup E_2) = P(E_1) + P(E_2), \forall E_1, E_2 \subseteq \Omega : E_1 \cap E_2 = \emptyset$.

Example: Consider tossing a coin twice. Let $P(\omega) = 1/4, \forall \omega \in \Omega$ defined in the earlier example. Then $P(HH) = P(TT) = 1/4$ and $P(\{HT, TH\}) = 1/2$.

Outline

Probability

Random Variables

- CDFs, pmfs, and pdfs
- Important Univariate Distributions

Random Vectors

- Joint CDFs, marginals and conditionals pmfs and pdfs
- Independence
- Bivariate Normal Distribution

Outline

Probability

Random Variables
- CDFs, pmfs, and pdfs
- Important Univariate Distributions

Random Vectors
- Joint CDFs, marginals and conditionals pmfs and pdfs
- Independence
- Bivariate Normal Distribution

Random Variables

Random variables allow us to form a bridge between the sample space of an experiment and data.

Definition (Random Variable)

A random variable is a function

$$X : \Omega \to \mathbb{R}$$

that assigns a real number $X(\omega)$ to each outcome $\omega \in \Omega$.

Example: Consider flipping a coin twice and let $X(\omega)$ be the number of heads in ω. Then for $\omega = TH$ we have $X(\omega) = 1$.

Note: More mathematical rigor is necessary for a technical definition of a random variable, but that would exceed the scope of this course.

Cumulative Distribution Functions

The cumulative distribution function allows for succinctly characterizing random variables.

Definition (Cumulative Distribution Function)

The cumulative distribution function (CDF) of a random variable X is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) = P(X \leq x), \forall x \in \mathbb{R}.$$

Notation: Capital letters X typically denote random variables, while lower case letters x typically denote realized values (i.e., a number). We write $X \sim F_X$ to state that X has distribution F_X.

Cumulative Distribution Functions (Contd.)

The next result is crucial: it states that the CDF effectively contains all the information about a random variable.

Theorem

Let X and Y be random variables with CDFs F_X and F_Y, respectively. If $F_X(x) = F_Y(x)$, $\forall x \in \mathbb{R}$, then $P(X \in E) = P(Y \in E)$.

For two random variables X and Y with CDFs F_X and F_Y, respectively, we say that X and Y are identically distributed – denoted by $X \stackrel{d}{=} Y$ – if $F_X(x) = F_Y(x)$, $\forall x \in \mathbb{R}$.

Cumulative Distribution Functions (Contd.)

Example: Consider flipping a fair coin twice as before. Let X be the number of heads. Then

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1/4 & \text{if } x \in [0,1), \\ 3/4 & \text{if } x \in [1,2), \\ 1 & \text{if } x \geq 2. \end{cases}$$

Now consider a second random variable Y equal to the number of tails. We have $X \overset{d}{=} Y$.

The example highlights that $X \overset{d}{=} Y$ does not imply $X = Y$.

Two random variables can be identically distributed, but their realizations do not have to be equal.

Discrete Random Variables

Definition (Discrete Random Variable)

A random variable X is discrete if it takes countably many values $\{x_1, x_2, \dots\}$. The probability mass function (pmf) of X is defined as

$$f_X(x) = P(X = x), \forall x \in \mathbb{R}.$$

The support of X is given by

$$\text{supp } X = \{x \in \mathbb{R} \mid f_X(x) > 0\}.$$

The support of X is the set of values it can take. By the definition of probabilities, it holds that $f_X(x) \geq 0, \forall x \in \mathbb{R}$ and

$$\sum_{x \in \text{supp } X} f_X(x) = 1.$$

The pmf and CDF of X are related via

$$F_X(x) = P(X \leq x) = \sum_{x' \in \text{supp } X} f_X(x') \mathbb{1}\{x' \leq x\}.$$

Discrete Random Variables (Contd.)

Example: Consider flipping a fair coin twice, with $X(\omega)$ is the number of heads. We have

$$\text{supp } X = \{0, 1, 2\}$$

and the corresponding pmf is

$$f_X(x) = \begin{cases} 1/4 & \text{if } x = 0, \\ 1/2 & \text{if } x = 1, \\ 1/4 & \text{if } x = 2, \\ 0 & \text{otherwise.} \end{cases}$$

We may easily calculate $F_X(1)$ via the earlier equation as

$$F_X(1) = f_X(0) + f_X(1) = 3/4.$$

Continuous Random Variables

Definition (Continuous Random Variable)

A random variable X is continuous if there exists a function $f_X : \mathbb{R} \to \mathbb{R}$ satisfying:

1. $f_X(x) \geq 0$, $\forall x \in \mathbb{R}$;
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$;
3. $P(a < X \leq b) = \int_a^b f_X(x)dx$, $\forall a \leq b \in \mathbb{R}$.

The function f_X is called the probability density function (pdf) of X.

The pdf and CDF of X are related via

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt.$$

By the fundamental theorem of calculus, we have

$$\frac{\partial}{\partial x} F_X(x) = f_X(x).$$

Continuous Random Variables (Contd.)

Example: Consider the idea of choosing a random number between 0 and 1. For this purpose, construct the random variable X with pdf

$$f_X(x) = \begin{cases} 1 & \text{for } x \in [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $f_X(x) \geq 0$, $\forall x \in \mathbb{R}$ and $\int_{-\infty}^{\infty} f_X(t)dt = \int_0^1 1dt = 1$. The corresponding CDF is given by

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0, \\ x & \text{for } x \in [0,1], \\ 1 & \text{for } x > 1. \end{cases}$$

This random variable X is called a *standard uniform random variable*, and write $X \sim U(0,1)$.

Caveat

Discrete and continuous random variables can lead to confusion.

When X is a continuous random variable, do not interpret $f_X(x)$ as $P(X = x)$. By Definition, $P(X = x) = \int_x^x f_X(t)dt = 0$ which is not equal to $f_X(x)$ (in general). $f_X(x) = P(X = x)$ only works for discrete random variables.

Note also that pdfs may take values larger than 1 or even be unbounded, but pmfs must map to [0, 1].

Example: Uniform distribution on the interval $[0, 1/2]$ has probability density $f_X(x) = 2$ for $0 \leq x \leq 1/2$ and $f_X(x) = 0$ elsewhere.

Example: pdf of $f_X(x) = \frac{1}{2\sqrt{x}}$ is unbounded.

Cumulative Distribution Functions (Contd.)

Lemma below allows us to readily express different kinds of probabilities using the CDF of the corresponding random variable.

Lemma

Let X be a random variable and F be the corresponding CDF. Then:

1. $P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a)$;
2. $P(X > x) = 1 - P(X \le x) = 1 - F(x)$;
3. *If X is continuous, then*

$$\begin{aligned} F(b) - F(a) &= P(a < X < b) = P(a \le X \le b) \\ &= P(a < X \le b) = P(a \le X \le b). \end{aligned}$$

Quantile Functions

Another characterization of a random variable is its quantile function.

Definition

Let X be a random variable and F be the corresponding CDF. The quantile function (or inverse CDF) is the function $F^{-1} : [0, 1] \to \text{supp } X$ defined by

$$F^{-1}(q) = \inf\{x \mid F(x) \geq q\}, \forall q \in [0, 1].$$

When F is strictly increasing and continuous, then $F^{-1}(q)$ is the unique real number that satisfies

$$P(X \leq F^{-1}(q)) = q.$$

Note: If you are unfamiliar with the infimum operator $\inf$, just think of it as the minimum (that will suffice for this class).

Outline

Important Univariate Distributions

Discussed two well-known probability distributions:

- Binomial random variables
- Uniform random variable

Will introduce other examples of frequently-occurring discrete and continuous random variables.

- Use the examples to gain intuition about how random variables can be leveraged for modeling a real-world experiment.
- The most important example provided is the normal distribution: Study this carefully!

Important Discrete Distributions

Definition (Discrete Uniform Distribution)

Let $k > 1$ be a given integer. Suppose that X has pmf given by

$$f_X(x) = \begin{cases} 1/k, & \text{for } x = 1, \ldots, k, \\ 0, & \text{otherwise.} \end{cases}$$

We say that X has a uniform distribution on $\{1, \ldots, k\}$, and write $X \sim U\{1, \ldots, k\}$.

The discrete uniform distribution is for randomly choosing a single value from a finite set of values with equal probability.

Important Discrete Distributions (Contd.)

Definition (Bernoulli Distribution)

Let $p \in (0, 1)$ be a given scalar. Suppose that X has pmf given by

$$f_X(x) = \begin{cases} p, & \text{if } x = 1, \\ 1 - p, & \text{if } x = 0, \\ 0, & \text{otherwise.} \end{cases}$$

We say that X has a Bernoulli distribution and write $X \sim \text{Bernoulli}(p)$.

The Bernoulli distribution represents a single coin flip where the probability of a success is denoted by p.

Note that for $x \in \{0, 1\}$, we may write the pmf as

$$f_X(x) = p^x (1 - p)^{1-x}.$$

Important Discrete Distributions (Contd.)

Definition (Binomial Distribution)

Let $p \in (0,1)$ and $n \in \mathbb{N}$ be given. Suppose that X has pmf given by

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{for } x = 0, \dots, n, \\ 0, & \text{otherwise.} \end{cases}$$

We say that X has a Binomial distribution and write $X \sim \text{Binomial}(p, n)$.

The Binomial distribution represents the number of successes in a sequence of n coin flips, where the probability of a success is p.

Notation: $\binom{n}{x}$ denotes the number of possible combinations of x out of n elements – that is,

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Important Continuous Distributions

Definition (Uniform Distribution)

Let $a < b \in \mathbb{R}$ be given scalars. Suppose that X has pdf given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{for } x \in [a,b], \\ 0, & \text{otherwise.} \end{cases}$$

We say that X has a uniform distribution on $[a,b]$, and write $X \sim U(a,b)$.

The uniform distribution represents choosing a number from the interval $[a,b]$ at random. The CDF of the uniform distribution:

$P(X \leq x) = \frac{x-a}{b-a} = F_X(x)$

Normal Distribution

Definition (Normal Distribution)

Let $\mu \in \mathbb{R}$ and $\sigma > 0$ be given scalars. Suppose that X has pdf given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \forall x \in \mathbb{R}.$$

We say that X has a normal distribution and write $X \sim N(\mu, \sigma^2)$.

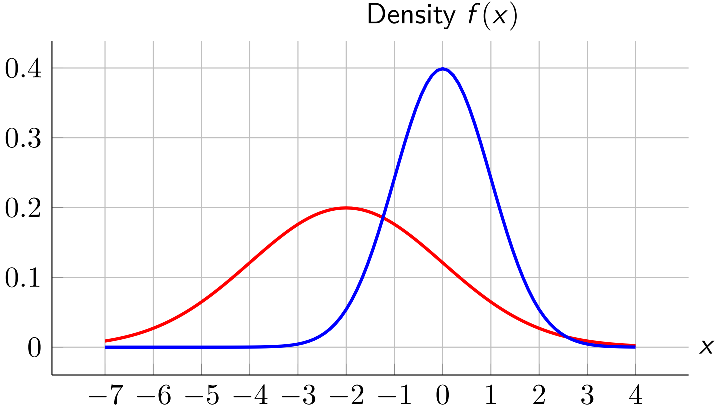
The normal distribution is immensely important in statistics and econometrics.

- Normal distributions often end up being suitable approximations, as formalized by the Central Limit Theorem. .

The normal distribution is symmetric around μ:

$$f_X(\mu + \delta) = f_X(\mu - \delta), \forall \delta \in \mathbb{R}.$$

Normal Distribution (Visualization)



Notes. Normal densities with $\mu = -2, \sigma = 2$ in red, and $\mu = 0, \sigma = 1$ in blue.

Normal Distribution (Contd.)

If $\mu = 0$ and $\sigma = 1$, we say that X has a standard normal distribution:

- Denote its pdf by $\phi(x)$;
- Denote its CDF by $\Phi(x)$;
- Denote its quantile function by $\Phi^{-1}(x)$.

There exists no closed-form expression for $\Phi(x)$. Conventions in statistics and econometrics make it worthwhile to memorize some key values:

- $\Phi(-1.96) \approx 0.025 \quad \text{and} \quad \Phi^{-1}(0.025) \approx -1.96.$
- $\Phi(-1.64) \approx 0.050 \quad \text{and} \quad \Phi^{-1}(0.050) \approx -1.64.$
- $\Phi(1.96) \approx 0.975 \quad \text{and} \quad \Phi^{-1}(0.975) \approx 1.96.$
- $\Phi(1.64) \approx 0.950 \quad \text{and} \quad \Phi^{-1}(0.950) \approx 1.64.$

Normal Distribution (Contd.)

We state the following useful properties without proof:

Lemma

Let $X \approx N(\mu, \sigma^2)$ *and* $Z \approx N(0, 1)$. *Then,*

a $\frac{X-\mu}{\sigma} \stackrel{d}{=} Z$

b $\mu + \sigma Z \stackrel{d}{=} X$

This lemma implies $P(a < X \leq b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$

Outline

Outline

Random Vectors

So far, we've only discussed univariate distributions.

We often need tools to characterize relationships between random variables.

A random vector is a function from the sample space to $\mathbb{R}^d$, for some $d \in \mathbb{N}$, i.e., $X : \Omega \to \mathbb{R}^d$.

- It's a simple generalization of random variables.
- Each component of a random vector is itself a random variable.

For ease of exposition, let's focus on bivariate random vectors $(d = 2)$.

Concepts generalize naturally to higher dimensions $d > 2$.

Joint Cumulative Distribution Functions

The joint CDF succinctly characterizes random vectors.

Definition (Joint Cumulative Distribution Function)

The joint cumulative distribution function (joint CDF) of a random vector (X, Y) is the function $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ defined by

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y), \forall (x, y) \in \mathbb{R}^2.$$

Note that this definition applies regardless of whether X and Y are continuous or discrete random variables.

Discrete Random Vectors

Definition (Bivariate Discrete Random Vector)

A pair of discrete random variables (X, Y) is a bivariate discrete random vector. The joint probability mass function (joint pmf) is defined as

$$f_{X,Y}(x, y) = P(X = x, Y = y), \forall (x, y) \in \mathbb{R}^2.$$

Example: Consider the random vector (X, Y) with joint pmf given by

	$Y = 0$	$Y = 1$
$X = 0$	$\frac{1}{5}$	$\frac{1}{10}$
$X = 1$	$\frac{3}{10}$	$\frac{2}{5}$

Thus $f_{X,Y}(0, 1) = P(X = 0, Y = 1) = \frac{1}{10}$.

Discrete Random Vectors (Contd.)

Definition (Marginal Probability Mass Function)

If (X, Y) is a discrete random vector with joint pmf $f_{X,Y}$, then the marginal pmf of X is defined by

$$f_X(x) = P(X = x) = \sum_{y \in \text{supp } Y} P(X = x, Y = y) = \sum_{y \in \text{supp } Y} f_{X,Y}(x, y).$$

The marginal pdf of Y is defined analogously.

Example: Consider again the joint pmf of the previous example. We have $P(X = 0) = \frac{3}{10}$ as

	$Y = 0$	$Y = 1$	Total
$X = 0$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{3}{10}$
$X = 1$	$\frac{3}{10}$	$\frac{2}{5}$	$\frac{7}{10}$
Total	$\frac{1}{2}$	$\frac{1}{2}$	1

Continuous Random Vectors

Definition (Bivariate Continuous Random Vector)

A pair of continuous random variables (X, Y) is a bivariate continuous random vector. The joint probability density function (joint pdf) is a function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ that satisfies the following properties:

1. $f_{X,Y}(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2$;
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dxdy = 1$;
3. $P((X, Y) \in \mathcal{A}) = \int \int_A f_{X,Y}(x, y) dxdy, \forall \mathcal{A} \subseteq \mathbb{R}^2$.

Example: Consider randomly choosing a point on the unit square with coordinates (X, Y). Then

$$f_{X,Y}(x, y) = \begin{cases} 1, & \text{if } (x, y) \in [0, 1]^2, \\ 0, & \text{otherwise.} \end{cases}$$

We have $P(X \leq \frac{1}{2}, Y \leq \frac{1}{2}) = \frac{1}{4}$.

Continuous Random Vectors (Contd.)

Definition (Marginal Probability Density Function)

If (X, Y) is a continuous random vector with joint pdf $f_{X,Y}$, then the marginal pdf of X is defined by

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

The marginal pdf of Y is defined analogously.

Example: Recall the uniform distribution on the unit square:

$$f_{X,Y}(x, y) = \begin{cases} 1, & \text{if } (x, y) \in [0, 1]^2, \\ 0, & \text{otherwise.} \end{cases}$$

We have $f_X(x) = \mathbb{1}\{x \in [0, 1]\}$ and $f_Y(y) = \mathbb{1}\{y \in [0, 1]\}$.

Conditional Distributions

Joint distributions characterize the relationship between random variables.

Marginal probability density (or mass) functions are another name for the probability density (or mass) functions we discussed in the setting of random variables.

"*Marginal*" *highlights the context of multiple random variables*.

We now introduce the concept of conditional distributions.

Conditional distributions characterize a random variable when there is information on another random variable.

Conditional Distributions (Contd.)

Definition (Conditional Probability Mass Function)

If (X, Y) is a discrete random vector with joint pmf $f_{X,Y}$, then the conditional pmf of X given Y is defined by

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

$\forall (x, y) \in \mathbb{R}^2 : f_Y(y) > 0$ (and is undefined otherwise). The conditional pmf of Y given X is defined analogously.

Example: Consider again the joint pmf of the previous example. We have

$$P(X = 0 \mid Y = 0) = \frac{P(X = 0, Y = 0)}{P(Y = 0)} = \frac{\frac{1}{5}}{\frac{5}{10}} = \frac{2}{5},$$

$$P(Y = 0 \mid X = 0) = \frac{P(Y = 0, X = 0)}{P(X = 0)} = \frac{\frac{1}{5}}{\frac{3}{10}} = \frac{2}{3}.$$

Conditional Distributions (Contd.)

Definition (Conditional Probability Density Function)

If (X, Y) is a continuous random vector with joint pdf $f_{X,Y}$, then the conditional pdf of X given $Y = y$ is defined by

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

$\forall (x, y) \in \mathbb{R}^2$ where $f_Y(y) > 0$ (and is undefined otherwise). Then,

$$P(X \in \mathcal{A} | Y = y) = \int_{\mathcal{A}} f_{X|Y}(x|y)\, dx.$$

The conditional pdf of Y given $X = x$ is defined analogously.

From the definitions of the conditional pmf and pdf, we see that

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

Outline

Independence

We now turn to a restriction on the relationship between random variables that is of the highest importance in all the identifying assumptions we will consider in this course.

Definition (Independence)

Two random variables X and Y are independent if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \forall A, B \subseteq \mathbb{R}.$$

Independence is denoted by $X \perp\!\!\!\perp Y$.

Checking the above equation by brute force is challenging. Fortunately, we have the following key result:

Theorem

Let (X, Y) *have joint pdf* (*or pmf*) $f_{X,Y}$. *Then*

$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad f_{X,Y}(x, y) = f_X(x) f_Y(y), \forall (x, y) \in \mathbb{R}^2.$$

Independence (Contd.)

An immediate consequence of the previous theorem is the following result:

Corollary

Let (X, Y) *have joint pdf (or pmf)* $f_{X,Y}$. *Then*

$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad f_{X|Y}(x|y) = f_X(x),\ \forall (x,y) \in \mathbb{R}^2.$$

Proof:

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x) \end{aligned}$$

Independence (Contd.)

Example: Consider again the joint pmf of the previous example. The example showed that $P(X = 0 \mid Y = 0) = \frac{2}{5}$ but we have $P(X = 0) = \frac{3}{10}$. Hence, by the corollary, we can conclude that X and Y are not independent.

Suppose now that the joint pmf of (X, Y) is instead given by

	$Y = 0$	$Y = 1$	Total
$X = 0$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
$X = 1$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
Total	$\frac{1}{2}$	$\frac{1}{2}$	1

Now we have $f_X(x)f_Y(y) = f_{X,Y}(x, y)$, $\forall(x, y) \in \mathbb{R}^2$. Hence, by the theorem, we can conclude that $X \perp\!\!\!\perp Y$.

Independence (Contd.)

The next result is important for working with random variables within economic models.

Corollary

Let (X, Y) be two random variables. Then, for any function h,

$$X \perp\!\!\!\perp Y \Rightarrow X \perp\!\!\!\perp h(Y).$$

Proof: Define $h^{-1}(\mathcal{B}) \equiv \{y \in \mathbb{R} | h(y) \in \mathcal{B}\}$. $\forall \mathcal{A}, \mathcal{B} \subset \mathbb{R}$

$$\begin{aligned} P(x \in \mathcal{A}, h(y) \in \mathcal{B}) &= P(x \in \mathcal{A}, y \in h^{-1}(\mathcal{B})) \\ &= P(x \in \mathcal{A}) P(y \in h^{-1}(\mathcal{B})) \quad \because X \perp\!\!\!\perp Y \\ &= P(x \in \mathcal{A}) P(h(y) \in \mathcal{B}) \end{aligned}$$

Outline

Bivariate Normal Distribution

We now turn to one particularly important bivariate distribution.

Definition (Bivariate Normal Distribution)

Let $(\mu_X, \mu_Y) \in \mathbb{R}^2$ and

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}, \, (\sigma_X, \sigma_Y, \sigma_{XY}) \in \mathbb{R}^3 \text{ such that } \sigma_X^2 \sigma_Y^2 > \sigma_{XY}^2$$

be given. Suppose that the random vector (X, Y) has joint pdf given by

$$f_{X,Y}(x, y) = \frac{1}{2\pi |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} \begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix}^\top \Sigma^{-1} \begin{pmatrix} x - \mu_X \\ y - \mu_Y \end{pmatrix} \right).$$

We say that (X, Y) has a bivariate normal distribution and write $(X, Y) \sim N(\mu, \Sigma)$, where $\mu = (\mu_X, \mu_Y)$.

Bivariate Normal Distribution (Contd.)

Bivariate normals are convenient because their marginal (and conditional) densities can be succinctly expressed. If $(X, Y) \sim N(\mu, \Sigma)$, then

- $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$;
- $Y \mid X = x \sim N\left(\mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(x - \mu_X), \sigma_Y^2 - \frac{\sigma_{XY}^2}{\sigma_X^2}\right)$.

Another useful property of normal random vectors is that independence reduces to a simple condition.

Theorem

Let $(X, Y) \sim N(\mu, \Sigma)$. *Then*

$$X \perp\!\!\!\perp Y \quad \Leftrightarrow \quad \sigma_{XY} = 0.$$

Bivariate Normal Distribution (Contd.)

Lemma

Let $(X, Y) \sim N(\mu, \Sigma)$, *and* $Z \sim N(0, I_2)$. *Then*:

1. $\Sigma^{-\frac{1}{2}} \begin{pmatrix} X \\ Y \end{pmatrix} - \mu \sim Z$;
2. $\mu + \Sigma^{\frac{1}{2}} Z \sim \begin{pmatrix} X \\ Y \end{pmatrix}$;
3. *For given* $a, b \in \mathbb{R}$, *we have*
$$aX + bY \sim N\left(a\mu_X + b\mu_Y, \begin{pmatrix} a & b \end{pmatrix} \Sigma \begin{pmatrix} a \\ b \end{pmatrix}\right);$$
4. *If in addition* $X \perp\!\!\!\perp Y$, *then*
$$aX + bY \sim N\left(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2\right).$$

Notation: For $d \in \mathbb{N}$, I_d denotes the identity matrix in d dimensions.

χ^2-Distribution

By the previous lemma, we're equipped to construct independent normal random variables from any bivariate normal random vector (X, Y) given μ and Σ.

It allows for the construction of another well-known probability distribution: The χ^2-distribution

Theorem

Let $Z \sim N(0, I_2)$. Then

$$Z^\top Z = Z_1^2 + Z_2^2 \sim \chi^2(2),$$

where $\chi^2(df)$ denotes the χ^2-distribution with df-degrees of freedom.
More generally, if $Z \sim N(0, I_m)$ for some $m \in \mathbb{N}$, then

$$Z^\top Z = \sum_{i=1}^{m} Z_i^2 \sim \chi^2(m).$$

χ^2-Distribution

We formulate the following corollary for ease of application:

Corollary

Let $X \sim N(\mu, \Sigma)$, where supp $X = \mathbb{R}^m$. Then

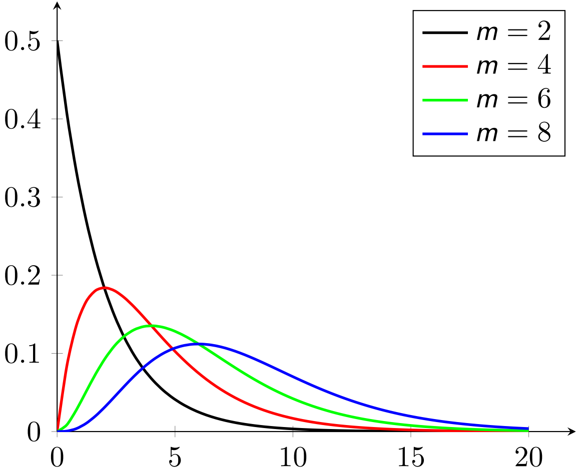
$$(X - \mu)^T \Sigma^{-1} (X - \mu) \sim \chi^2(m)$$

Proof:[1]

$$\begin{aligned}(X - \mu)^T \Sigma^{-1} (X - \mu) &= (X - \mu)^T (\Sigma^{-1/2})^T \Sigma^{-1/2} (X - \mu) \\ &= [\Sigma^{-1/2}(X - \mu)]^T [\Sigma^{-1/2}(X - \mu)] \\ &= Z^T Z = \sum_{i=1}^{m} Z_i^2 \sim \chi^2(m).\end{aligned}$$

[1] Theorem: Let A be a positive semidefinite symmetric matrix. Then there is exactly one positive semidefinite and symmetric matrix B such that $A = BB$.

χ^2-Distribution (Contd)



Notes: χ^2 desities with different dfs.

Summary

Thus far, we focused on distributions of random variables:

- CDFs and pdfs (or pmfs) fully characterize a random variable.
- Joint CDFs and joint pdfs (or pmfs) fully characterize relationships between random variables.

But, knowning everything about a random variable or its relation to other random variables is not always necessary.

- Often, we are content with knowing about key features of a random variable that partly characterize it or its relation to other random variables.
- The causal question did not consider the distribution of hourly wages for college graduates had they not pursued higher education. Instead we were content with knowing the **expected** returns to education.

In Part B of the probability theory review, we will cover concepts that summarize key features of a random variable's (or random vector's) distribution