Math refresher B: Fundamentals of probability

Econometrics for minor Finance, Lecture 1–2

Tunga Kantarcı, Fall 2025

Fundamentals of probability: Random variables and their probability distributions: Random variable

A random variable is a numerical quantity whose value is uncertain before an experiment, but becomes fixed once the outcome of the experiment is revealed.

Fundamentals of probability: Random variables and their probability distributions: Random variable: Example

Consider flipping a fair coin 10 times as an experiment. The number of times the coin lands on heads is a random variable. This number is random because it is the probabilistic outcome of each flip. Before flipping the coin, the experiment, we know that the possible values range from 0 to 10, but the actual number of heads is uncertain.

Once we perform the 10 flips and count the number of heads, we obtain a specific realization of the random variable for that trial. From another trial, we could get a different realization.

Fundamentals of probability: Random variables and their probability distributions: Random variable: Notation

We indicate collections of random variables using an uppercase letter and subscripts to distinguish individual elements within the group:

$$\{X_i : i = 1, \ldots, n\}$$

Each X_i is a random variable that can take on values from a set of possible outcomes, typically denoted in lowercase:

$$\{x_1, x_2, \ldots, x_k\}$$

After the experiment or sampling process is complete, we observe specific numerical values for each X_i. These are called realizations, and we denote them using lowercase letters with matching subscripts:

$$\{x_i : i = 1, \ldots, n\}$$

Fundamentals of probability: Random variables and their probability distributions: Random variable: Notation

The lowercase x notation is commonly used for both possible outcomes and observed realizations. In this context, the set $\{x_1, x_2, \ldots, x_k\}$ refers to the possible outcomes, while x_i refers to the actual observed value of X_i.

Fundamentals of probability: Random variables and their probability distributions: Random variable: Notation: Example

Suppose we record last year's income for 20 randomly selected households. We represent these income values as random variables:

$$X_1, X_2, \ldots, X_{20}$$

Each X_i reflects the income of the i-th household and is subject to randomness due to the sampling process. Once the data is collected, we observe specific numerical values, the realizations of these random variables:

$$x_1, x_2, \ldots, x_{20}$$

Realizations form the actual dataset used for analysis, while the random variables represent the underlying probabilistic structure before observation.

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable

A discrete random variable is one that takes on only a finite number of values.

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable

To fully describe the behavior of a random variable, we only need the probability that it takes on values.

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable: Example

A Bernoulli random variable is the simplest example of a discrete random variable. It can take on only the values 0 and 1. It is conventional to refer to the event $X = 0$ as a "failure" and the event $X = 1$ as a "success".

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable: Example

To fully describe the behavior of a Bernoulli variable, we only need the probability that it takes on the value 1. In the coin-flipping example, if the coin is "fair," then

$$P(X = 1) = \frac{1}{2}$$

which reads as "the probability that X equals one is one-half". Since probabilities must sum to one, it follows that

$$P(X = 0) = \frac{1}{2}$$

as well.

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable

More generally, any discrete random variable is completely described by listing

- its possible values, and
- the associated probability that it takes on each value.

If X takes on the k possible values $\{x_1, x_2, \ldots, x_k\}$, then the probabilities $p_1, p_2, \ldots, p_k$ are defined by

$$p_j = P(X = x_j), \quad j = 1, 2, \ldots, k$$

where each p_j is between 0 and 1, and

$$p_1 + p_2 + \cdots + p_k = 1.$$

This is read as: "The probability that X takes on the value x_j is equal to p_j". We summarize this information as follows.

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable: PMF

The probability mass function (PMF) of X summarizes the information concerning the possible outcomes of X and the corresponding probabilities:

$$f_X(x_j) = p_j$$

for $j = 1, 2, \ldots, k$.

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable: PMF: Notation

Note how we subscript the PMF of X in question:

$$f_X(x)$$

This reads as: the value of the PMF of the random variable X, evaluated at the point x. In other words, $f_X(x)$ is the probability that the random variable X takes on the particular realized value x.

Sometimes the subscript X is suppressed, and we simply write $f(x)$ when the context makes the random variable clear.

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable: PMF

Given the PMF of any discrete random variable, it is simple to compute the probability of any event involving that random variable.

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable: PMF: Example

In basketball, a free throw is when a player gets to shoot the ball from a special line without anyone trying to block them. If the ball goes in the hoop, they score one point. Suppose a player gets two chances to shoot. Let X be the number of successful free throws. Then X can be:

- $X = 0$: The player missed both shots
- $X = 1$: The player made one successful shot
- $X = 2$: Both shots are successful

So, X can take on three values: $\{0, 1, 2\}$ depending on how many free throws the player makes.

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable: PMF: Example

Assume that the PMF of X is given by:

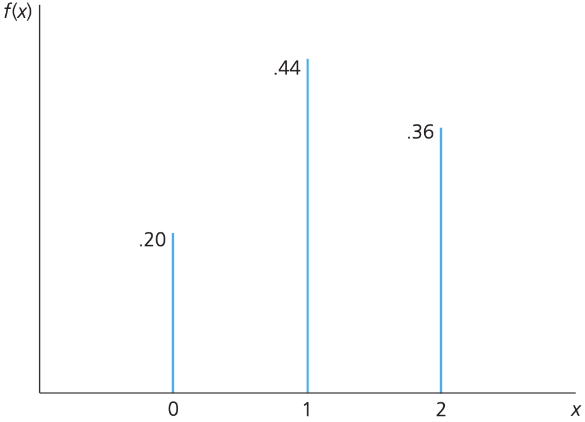
- f(0) = 0.20
- f(1) = 0.44
- f(2) = 0.36

The three probabilities sum to one, as they must. Using this PMF, we can calculate the probability that the player makes at least one free throw:

$$P(X \geq 1) = P(X = 1) + P(X = 2) = 0.44 + 0.36 = 0.80.$$

Fundamentals of probability: Random variables and their probability distributions: Discrete random variable: PMF: Example

The PMF of the number of free throws made out of two attempts.



Fundamentals of probability: Random variables and their probability distributions: Continuous random variable

A continuous random variable can take any real value. Although real-world measurements are recorded in discrete units, variables with a wide range of values are often modeled as continuous for simplicity.

Fundamentals of probability: Random variables and their probability distributions: Continuous random variable: Example

The price of a product might be measured in cents, which is technically discrete. However, since prices can vary by fractions of a cent and span a large range, it is more practical to treat price as a continuous variable.

Fundamentals of probability: Random variables and their probability distributions: Continuous random variable: PDF

A probability density function (PDF) describes the likelihood of outcomes for a continuous random variable. We denote a PDF with

$$f_X(x)$$

Fundamentals of probability: Random variables and their probability distributions: Continuous random variable: PDF

A continuous random variable can take on infinitely many possible values. Therefore the probability of it taking on any exact value is 0. This is not a limitation, it is a property of continuous distributions.

Fundamentals of probability: Random variables and their probability distributions: Continuous random variable: PDF

The value of the PDF at a point x, denoted $f(x)$, does not represent the probability that $X = x$, since that probability is 0. Instead, $f(x)$ reflects the relative likelihood of values near x compared to other regions of the distribution. If, say, $f(x_1) > f(x_2)$, then values near x_1 are more likely to occur than values near x_2.

Fundamentals of probability: Random variables and their probability distributions: Continuous random variable: PDF

To find actual probabilities, we use the PDF over intervals. That is, for constants a and b, the probability that X lies between them is given by the area under the PDF between those points:

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

The total area under the PDF is always equal to 1.

Fundamentals of probability: Random variables and their probability distributions: Continuous random variable: PDF: Example

The PDF of the t-distribution is given by

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where

- t is the variable
- ν is the degrees of freedom, and determines the shape of the PDF, and
- $\Gamma(\cdot)$ denotes the gamma function.

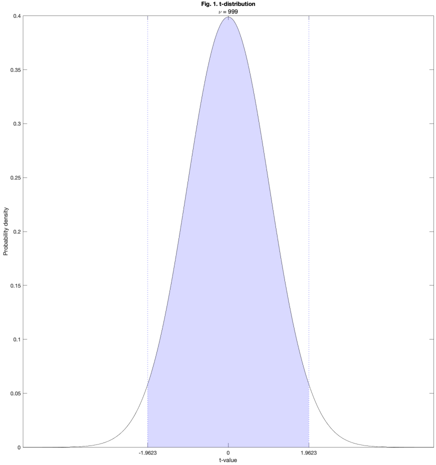
Fundamentals of probability: Random variables and their probability distributions: Continuous random variable: PDF: Example

Consider

$$P(-1.9623 \leq X \leq 1.9623) = \int_{-1.9623}^{1.9623} f_T(t)\, dt \approx 0.9409$$

Let's plot it for $\nu = 999$.

Fundamentals of probability: Random variables and their probability distributions: Continuous random variable: PDF: Example



Fundamentals of probability: Joint distributions, conditional distributions, and independence

The conditional probability of event y given event x is defined as:

$$P(y \mid x) = \frac{P(x, y)}{P(x)}$$

This represents the probability that event y occurs, assuming that event x has occurred. Here the

- numerator represents the probability that both events y and x happen together, and
- denominator restricts our attention to the cases where x happens.

So we are asking: out of all the times x occurs, how often does x also occur?

Fundamentals of probability: Joint distributions, conditional distributions, and independence

What do we use conditional probability for? In econometrics, we are often interested in how one random variable, say X, affects another, Y. The most we can know about how X affects Y is captured by the conditional distribution of Y given X. This relationship is summarized by the conditional PMF, defined as:

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

This is interpreted as the probability that $Y = y$ given that $X = x$.

Fundamentals of probability: Joint distributions, conditional distributions, and independence

Let X and Y be discrete random variables. The pair (X, Y) has a joint distribution, which is fully characterized by its joint PMF:

$$f_{X,Y}(x, y) = P(X = x, Y = y) = P(Y = y \mid X = x) \cdot P(X = x)$$

Fundamentals of probability: Joint distributions, conditional distributions, and independence

Random variables X and Y are said to be independent if and only if

$$\begin{aligned} P(X = x, Y = y) &= P(X = x) \cdot P(Y = y \mid X = x) \\ &= P(X = x) \cdot P(Y = y) \end{aligned}$$

Independence means that the probability of one variable's outcome does not change depending on the value of the other. When this holds, the probability that $X = x$ and $Y = y$ occur together is simply the product of their individual probabilities: These individual probabilities are called marginal PMFs.

Fundamentals of probability: Joint distributions, conditional distributions, and independence

"if and only if", for the second term of the product above, means:

- If conditional PMF = marginal PMF, then independence
- If independent, then conditional PMF = marginal PMF

So it is a perfect two-way relationship.

Fundamentals of probability: Features of probability distributions: Measure of central tendency: Expected value

Let

- X be a discrete random variable taking on a finite set of possible outcomes $\{x_1, x_2, \ldots, x_k\}$, and
- $f(x)$ represent the PMF of X

The expected value of X is given by

$$\mathbb{E}(X) = x_1 f(x_1) + x_2 f(x_2) + \ldots + x_k f(x_k) = \sum_{i=1}^{k} x_i f(x_i)$$

This expression represents a weighted average of all possible values that X can take, where the weights are given by the corresponding probabilities.

Fundamentals of probability: Features of probability distributions: Measure of central tendency: Expected value

The expected value is the theoretical average of all possible outcomes of a random variable. For this reason, it is often referred to as the population mean, where X denotes a variable defined over an entire population.

The sample mean, on the other hand, is an estimate of the expected value calculated from observed data.

Fundamentals of probability: Features of probability distributions: Measure of central tendency: Expected value: Example

Suppose that X takes on the values -1, 0, 2, with probabilities $\frac{1}{8}$, $\frac{1}{2}$, $\frac{3}{8}$, respectively. Then,

$$\mathbb{E}(X) = (-1) \cdot \frac{1}{8} + 0 \cdot \frac{1}{2} + 2 \cdot \frac{3}{8} = \frac{5}{8}$$

Fundamentals of probability: Features of probability distributions: Measure of central tendency: Expected value: Example

Now suppose we draw a random sample of 8 observations from this distribution: $\{-1, 0, 0, 0, 0, 2, 2, 2\}$. Notice that -1 occurs 1 out of 8 times, 0 occurs 4 out of 8 times, and 2 occurs 3 out of 8 times. These frequencies reflect the original probabilities: $\frac{1}{8}$, $\frac{1}{2}$, and $\frac{3}{8}$, respectively. The sample mean is:

$$\bar{x} = \frac{(-1) + 0 + 0 + 0 + 0 + 2 + 2 + 2}{8} = \frac{5}{8}$$

In this case, the sample mean $\bar{x}$ happens to equal the expected value $\mathbb{E}(X)$, because the sample proportions match the theoretical probabilities.

Fundamentals of probability: Features of probability distributions: Measure of central tendency: Expected value: Properties

exp. val. property 1: For any constant c:

$$\mathbb{E}(c) = c$$

This shows that if a random variable always takes the same value c, then its average over all outcomes is simply c itself.

Fundamentals of probability: Features of probability distributions: Measure of central tendency: Expected value: Properties

exp. val. property 2: For any constants a and b:

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

This shows that expectation is a linear operator: scaling a random variable scales its expected value, and shifting it adds the same constant to the expectation.

Fundamentals of probability: Features of probability distributions: Measure of variability: Variance

For a random variable X, let $M = \mathbb{E}(X)$.

There are various ways to measure how far X is from its expected value, but the simplest one to work with is the squared difference

$$(X - M)^2$$

The squaring eliminates the sign from the distance measure; the resulting positive value corresponds to our intuitive notion of distance, and treats values above and below M symmetrically.

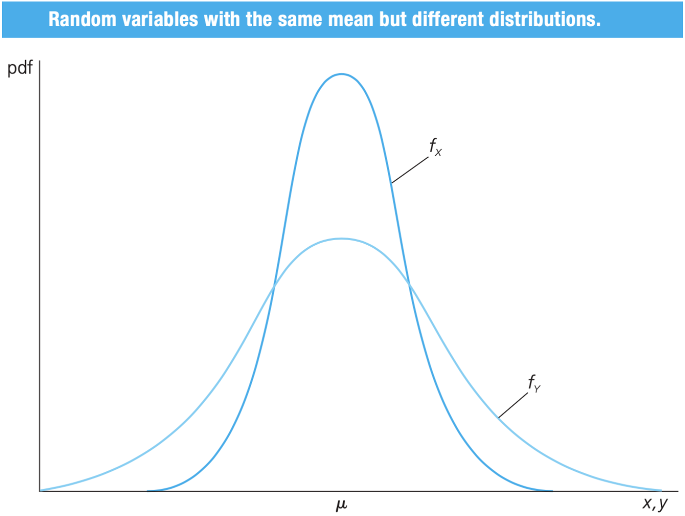
Fundamentals of probability: Features of probability distributions: Measure of variability: Variance

Since X can take on different values depending on the outcome, the squared difference $(X - M)^2$ is itself a random variable. To summarize this variability with a single number, just as we use the population mean to summarize central tendency, take the population average of these squared differences across all possible outcomes. This gives a measure of how much X typically deviates from its mean:

$$\mathrm{Var}(X) = \mathbb{E}[(X - M)^2]$$

Variance is often denoted σ_X^2, or simply σ^2, when the context is clear.

Fundamentals of probability: Features of probability distributions: Measure of variability: Variance



Fundamentals of probability: Features of probability distributions: Measure of variability: Variance: Properties

var. property 1: The variance of any constant is 0.

Fundamentals of probability: Features of probability distributions: Measure of variability: Variance: Properties

var. property 2: For constants a and b,

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

This means that adding a constant to a random variable does not change the variance, but multiplying a random variable by a constant scales the variance by the square of that constant. For example, if X denotes temperature in Celsius and $Y = 32 + \frac{9}{5}X$ is temperature in Fahrenheit, then

$$\mathrm{Var}(Y) = \left(\frac{9}{5}\right)^2 \mathrm{Var}(X)$$

Fundamentals of probability: Features of probability distributions: Measure of variability: Variance: Properties

var. property 3: For constants a and b,

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$$

We are yet to learn about covariance.

Fundamentals of probability: Features of probability distributions: Measure of variability: Standard deviation

Standard deviation of a random variable is simply the square root of the variance:

$$\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$$

The standard deviation is often denoted σ_X, or simply σ when this is clear from the context.

Fundamentals of probability: Features of joint and conditional distributions: Measures of association

The probability distribution $f_X(x)$ of a random variable X describes all the information about its behavior. The expected value and variance of X are single numbers that summarize this entire distribution.

The joint PDF of two random variables, $f_{X,Y}(x, y)$, completely describes the relationship between them. We also need a single number to summarize this joint distribution. That is, it is useful to have a summary measure of how, on average, two random variables vary with one another.

Fundamentals of probability: Features of joint and conditional distributions: Measures of association: Covariance

Let $M_X = \mathbb{E}(X)$ and $M_Y = \mathbb{E}(Y)$, and consider the random variable

$$(X - M_X)(Y - M_Y)$$

If

- $X > M_X$ and $Y > M_Y$, or
- $X < M_X$ and $Y < M_Y$,

then the product is positive. On the other hand, if

- $X > M_X$ and $Y < M_Y$, or
- vice versa,

then the product is negative. Can this product tell us anything about the relationship between X and Y?

Fundamentals of probability: Features of joint and conditional distributions: Measures of association: Covariance

Covariance between two random variables X and Y is defined as:

$$\text{Cov}(X, Y) = \mathbb{E}[(X - M_X)(Y - M_Y)]$$

It says, if

- positive, then, on on average, when X is above its mean, Y is also above its mean
- negative, then, on average, when X is above its mean, Y is below its mean

Covariance measures the amount of linear dependence between two random variables. If positive, two random variables move in the same direction. If negative, they move in opposite directions. It is often denoted σ_{XY}.

Fundamentals of probability: Features of joint and conditional distributions: Measures of association: Covariance: Properties

Because covariance is a measure of how two random variables are related, it is natural to ask how covariance is related to the notion of independence:

cov. property 1: If X and Y are independent, then

$$\mathrm{Cov}(X, Y) = 0$$

Fundamentals of probability: Features of joint and conditional distributions: Measures of association: Covariance: Properties

It is important to note that if the covariance between two random variables X and Y is zero, that is,

$$\text{Cov}(X, Y) = 0,$$

this does not necessarily imply that X and Y are independent. 0 covariance only indicates that there is no linear relationship between the variables. They may still be dependent in a nonlinear way.

Fundamentals of probability: Features of joint and conditional distributions: Measures of association: Covariance: Properties

cov. property 2: For any constants a and b,

$$\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X, Y)$$

An important implication of this property is that the covariance between two random variables is sensitive to the units of measurement. That is, multiplying one or both variables by a constant, such as converting dollars to euros, will scale the covariance accordingly. This matters in economics because variables like prices, wages, and inflation rates can be expressed in different units without changing their underlying relationships, yet the numerical value of their covariance will change depending on those units. This is a deficiency and it is overcome by the correlation coefficient.

Fundamentals of probability: Features of joint and conditional distributions: Measures of association: Correlation coefficient

The correlation coefficient between X and Y is given by

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\text{sd}(X) \cdot \text{sd}(Y)} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Because σ_X and σ_Y are positive, $\text{Cov}(X, Y)$ and $\text{Corr}(X, Y)$ always have the same sign, and

$$\text{Corr}(X, Y) = 0 \quad \text{if and only if} \quad \text{Cov}(X, Y) = 0.$$

The corr. coef. is sometimes denoted ρ_{XY}.

Fundamentals of probability: Features of joint and conditional distributions: Measures of association: Correlation coefficient: Properties

corr. coef. property 1: $-1 \leq \text{Corr}(X, Y) \leq 1$

If

- $\rho_{XY} = 0$, there is no linear relationship between X and Y, and X and Y are said to be uncorrelated random variables
- ρ_{XY} is close to 1 or –1, the linear relationship is stronger

This shows that correlation between X and Y is invariant to the units of measurement of either X or Y. This is stated more generally in the second property.

Fundamentals of probability: Features of joint and conditional distributions: Measures of association: Correlation coefficient: Properties

corr. coef. property 2: For any constants a and b,

$$\text{Corr}(aX, bY) \stackrel{!}{=} \text{Corr}(X, Y)$$

If $ab < 0$, then

$$\text{Corr}(aX, bY) = -\text{Corr}(X, Y)$$

This says that multiplying one or both variables by a constant, such as converting dollars to euros , does not scale the correlation coefficient accordingly.

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation

Covariance and correlation measure the linear relationship between two random variables, without any reference to which one depends on the other. In social sciences, more often we would like to explain one variable, called Y, in terms of another variable, say, X. Further, if Y is related to X in a nonlinear fashion, we would like to know this. Call Y the "explained variable" and X the "explanatory variable".

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation

We already introduced the notion of conditional PDF of Y given $X = x$: $f(y \mid x)$. Thus, we might want to see how the distribution of y, say wages, changes with x, say education level. However, we usually want to have a simple way of summarizing this distribution. A single number will no longer suffice, because the distribution of Y given $X = x$ generally depends on the value of x.

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation

We can summarize the relationship between Y and X by looking at the conditional expectation of Y given X, sometimes called the conditional mean. The idea is this: suppose we know that X has taken on a particular value, say, x. Then, we can compute the expected value of Y, given that we know this outcome of X. We denote this expected value by

$$\mathbb{E}(Y \mid X = x)$$

Generally, as x changes, so does $\mathbb{E}(Y \mid x)$. We always use $\mathbb{E}(Y \mid x)$ as shorthand.

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation

If Y is a discrete random variable taking on possible outcomes $\{y_1, y_2, \ldots, y_n\}$, then

$$\mathbb{E}(Y \mid X = x) = \sum_{i=1}^{n} y_i \cdot f(y_i \mid x)$$

This expression tells us to fix X at a specific value x, consider the PMF of Y at that x, and the mean of this conditional PMF will give the expected value of Y at that x : a weighted average of the possible outcomes y_i, where the weights are the conditional probabilities $f(y_i \mid x)$.

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Example

Let (X, Y) represent the population of all working individuals, where X is years of education and Y is hourly wage. Then,

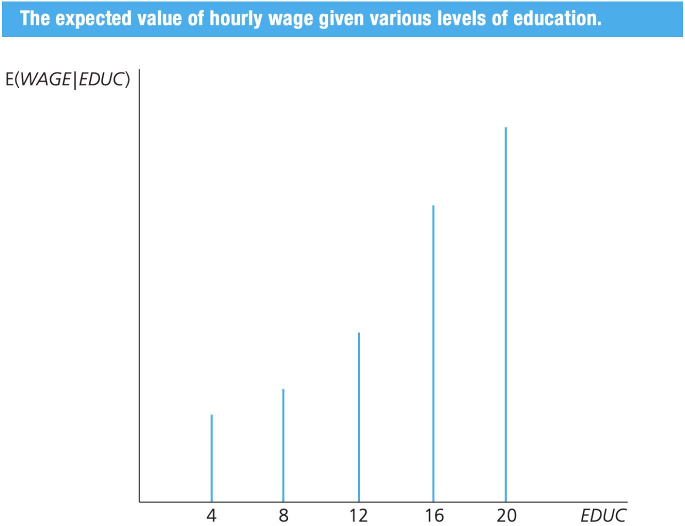
$$\mathbb{E}(Y \mid X = 12)$$

is the average hourly wage for all people in the population with 12 years of education, roughly a high school education. Similarly,

$$\mathbb{E}(Y \mid X = 16)$$

is the average hourly wage for all people with 16 years of education. Tracing out the expected value for various levels of education provides important information on how wages and education are related. The following plot for an illustration.

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Example



Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Example

In principle, the expected value of hourly wage can be found at each level of education, and these expectations can be summarized in a table. Because education can vary widely, and can even be measured in fractions of a year, this is a cumbersome way to show the relationship between average wage and amount of education.

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Example

In econometrics, we typically specify simple functions that capture this relationship. For example, suppose that the expected value of $WAGE$ given $EDUC$ is the linear function:

$$\mathbb{E}(WAGE \mid EDUC) = 1.05 + 0.45 \cdot EDUC$$

If this relationship holds in the population of working people, the average wage for people with eight years of education is:

$$1.05 + 0.45 \cdot 8 = \$4.65$$

The coefficient on education implies that each year of education increases the expected hourly wage by \$0.45, or 45 cents.

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Example

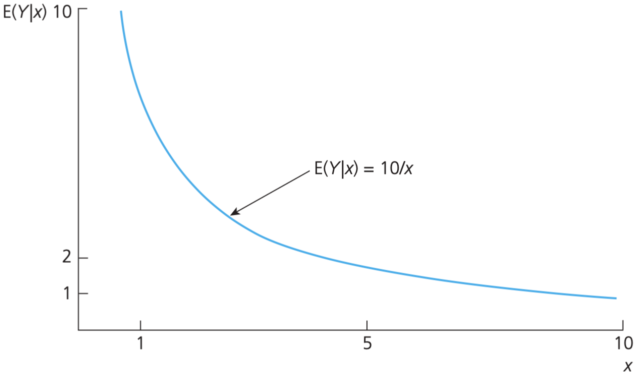
Conditional expectations can also be nonlinear functions. For example, suppose that:

$$\mathbb{E}(Y \mid x) = \frac{10}{x}$$

where X is a positive random variable. This could represent a demand function, where Y is quantity demanded and X is price. If Y and X are related in this way, an analysis of linear association, such as correlation analysis, would be inadequate. Let's plot this function.

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Example

Graph of $E(Y|x) = 10/x$.



Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Properties

con. exp. property 1: For any function $c(X)$,

$$\mathbb{E}[c(X) \mid X] = c(X)$$

This means that functions of X behave as constants when we compute expectations conditional on X. For example, $\mathbb{E}[X^2 \mid X] = X^2$. Intuitively, this simply means that if we know X, then we also know X^2.

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Properties

con. exp. property 2: For functions $a(X)$ and $b(X)$,

$$\mathbb{E}[a(X)Y + b(X) \mid X] = a(X) \cdot \mathbb{E}(Y \mid X) + b(X)$$

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Properties: Example

We can compute the conditional expectation of $XY + 28$ as

$$\mathbb{E}[XY + 28 \mid X] = X \cdot \mathbb{E}(Y \mid X) + 28$$

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Properties

con. exp. property 2: If X and Y are independent, then

$$\mathbb{E}(Y \mid X) = \mathbb{E}(Y)$$

This property means that, if X and Y are independent, then the expected value of Y given X does not depend on X. In this case, $\mathbb{E}(Y \mid X)$ always equals the (unconditional) expected value of Y.

Fundamentals of probability: Features of joint and conditional distributions: Conditional expectation: Properties: Example

In the wage and education example, if wages were independent of education, then the average wages of high school and college graduates would be the same:

$$\mathbb{E}(WAGE \mid EDUC) = \mathbb{E}(WAGE)$$

Because this is almost certainly false, we cannot assume that wage and education are independent.

Fundamentals of probability: Features of joint and conditional distributions: Conditional variance

Given random variables X and Y, the variance of Y, conditional on $X = x$, is simply the variance associated with the conditional distribution of Y given $X = x$:

$$\text{Var}(Y \mid X = x) = \mathbb{E}\left[(Y - \mathbb{E}(Y \mid x))^2 \mid x\right]$$

This is simply the variance of Y, except that the deviation is taken from the conditional mean $\mathbb{E}(Y \mid x)$ rather than the unconditional mean $\mathbb{E}(Y)$.

Fundamentals of probability: Features of joint and conditional distributions: Conditional variance: Example

Let $Y = \mathit{SAVINGS}$ and $X = \mathit{INCOME}$, both measured annually for the population of all families. Suppose that

$$\text{Var}(\mathit{SAVINGS} \mid \mathit{INCOME}) = 400 + 0.25 \cdot \mathit{INCOME}$$

This indicates that, as income increases, the variance in savings levels also increases. It is important to recognize that the relationship between the variance of savings and income is entirely separate from the relationship between the expected value of savings and income.

Fundamentals of probability: Features of joint and conditional distributions: Conditional variance: Properties

cond. cov. property 1: If X and Y are independent, then

$$\mathrm{Var}(Y \mid X) = \mathrm{Var}(Y)$$

This property is straightforward: when X and Y are independent, the distribution of Y given X does not depend on X.

Fundamentals of probability: Probability distributions: t distribution

The t-distribution is a workhorse in statistical inference, especially in regression analysis. Let

- Z have a standard normal distribution, and
- X have a chi-square distribution with ν degrees of freedom.

Further, assume that Z and X are independent. Then, the random variable

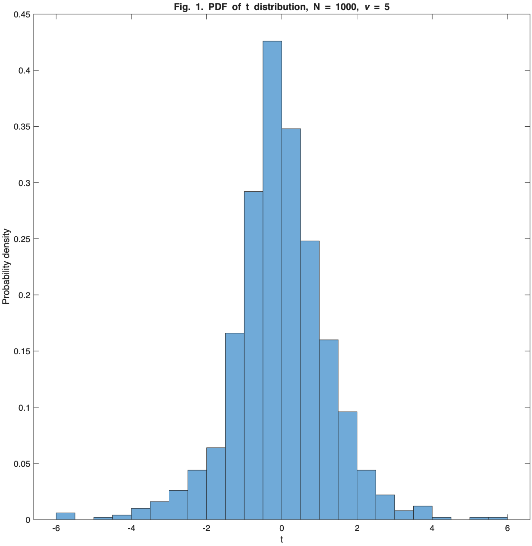
$$T = \frac{Z}{\sqrt{X/\nu}}$$

has a t-distribution with ν degrees of freedom. We denote this by $T \sim t_\nu$.

Fundamentals of probability: Probability distributions: t distribution

Let's generate random samples from the t-distribution and visualize their distribution using a plot.

Fundamentals of probability: Probability distributions: t distribution



Fundamentals of probability: Probability distributions: t distribution

By sampling many values from the t distribution and plotting them, we approximate its shape.

Fundamentals of Probability: Probability Distributions: t distribution

The PDF of the t-distribution is given by:

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where t is the variable, ν is the degrees of freedom, and $\Gamma(\cdot)$ denotes the gamma function. The gamma function $\Gamma(x)$ is a continuous extension of the factorial, defined for $x > 0$ by:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$$

It satisfies $\Gamma(n) = (n-1)!$ for any positive integer n.

Fundamentals of probability: Probability distributions: t distribution

Evaluating the PDF at many values of t yields a smooth curve that represents the t distribution.

Fundamentals of probability: Probability distributions: t distribution