Continuous Random Variables

A *continuous random variable* is one that assumes all values in an interval (or intervals) on the real line.

Continuous Random Variables

A *continuous random variable* is one that assumes all values in an interval (or intervals) on the real line.

Measurements of continuous quantities are usually represented as continuous random variables:

- temperature
- salinity
- pH
- elapsed time

Continuous Random Variables

The distinguishing feature of a continuous random variable is that its sample space Ω consists of one or more intervals on the real line.

Continuous Random Variables

The distinguishing feature of a continuous random variable is that its sample space Ω consists of one or more intervals on the real line.

When a probability experiment with a continuous sample space Ω is performed, the outcome is exactly one element $x \in \Omega$.

Continuous Random Variables

The distinguishing feature of a continuous random variable is that its sample space Ω consists of one or more intervals on the real line.

When a probability experiment with a continuous sample space Ω is performed, the outcome is exactly one element $x \in \Omega$.

The random variable X associated with the experiment is usually defined to be the value of x.

Continuous Random Variables

Example: The uniform distribution. The functions

- spreadsheet: `RAND()`
- R: `runif(1)`

perform the following experiment:

Choose a number from the interval $[0, 1]$ with each number in the interval equally likely to be chosen.

Continuous Random Variables

Example: The uniform distribution. The functions

- spreadsheet: `RAND()`
- R: `runif(1)`

perform the following experiment:

Choose a number from the interval $[0, 1]$ with each number in the interval equally likely to be chosen.

The sample space for this experiment is $\Omega = [0, 1]$

Continuous Random Variables

Example: The uniform distribution. The functions

- spreadsheet: `RAND()`
- R: `runif(1)`

perform the following experiment:

Choose a number from the interval $[0, 1]$ with each number in the interval equally likely to be chosen.

The sample space for this experiment is $\Omega = [0, 1]$

The value of random variable X associated with the outcome is just the value chosen.

Continuous Random Variables

Example: The uniform distribution. The functions

- spreadsheet: `RAND()`
- R: `runif(1)`

perform the following experiment:

Choose a number from the interval $[0, 1]$ with each number in the interval equally likely to be chosen.

The sample space for this experiment is $\Omega = [0, 1]$

The value of random variable X associated with the outcome is just the value chosen.

The continuous probability distribution associated with this experiment is called the **uniform** distribution.

Continuous Random Variables

Example: The normal distribution. The functions

- spreadsheet: `NORMSINV(RAND())`
- R: `rnorm(1)`

perform the following experiment:

Choose a number from the interval $(-\infty, \infty)$ according to the standard normal or bell curve distribution.

Continuous Random Variables

Example: The normal distribution. The functions

- spreadsheet: `NORMSINV(RAND())`
- R: `rnorm(1)`

perform the following experiment:

Choose a number from the interval $(-\infty, \infty)$ according to the standard normal or bell curve distribution.

The sample space for this experiment is $\Omega = (-\infty, \infty)$

Continuous Random Variables

Example: The normal distribution. The functions

- spreadsheet: `NORMSINV(RAND())`
- R: `rnorm(1)`

perform the following experiment:

Choose a number from the interval $(-\infty, \infty)$ according to the standard normal or bell curve distribution.

The sample space for this experiment is $\Omega = (-\infty, \infty)$

The value of random variable X associated with the outcome is just the value chosen.

Continuous Random Variables

Example: The normal distribution. The functions

- spreadsheet: `NORMSINV(RAND())`
- R: `rnorm(1)`

perform the following experiment:

Choose a number from the interval $(-\infty, \infty)$ according to the standard normal or bell curve distribution.

The sample space for this experiment is $\Omega = (-\infty, \infty)$

The value of random variable X associated with the outcome is just the value chosen.

The continuous probability distribution associated with this experiment is called the **normal** distribution.

Continuous Random Variables

Example: The chi-square distribution. The functions

- spreadsheet: `CHIINV(RAND(),1)`
- R: `rchisq(1,1)`

perform the following experiment:

Choose a number from the interval $(0, \infty)$ according to the chi-square distribution with one degree of freedom.

Continuous Random Variables

Example: The chi-square distribution. The functions

- spreadsheet: `CHIINV(RAND(),1)`
- R: `rchisq(1,1)`

perform the following experiment:

Choose a number from the interval $(0, \infty)$ according to the chi-square distribution with one degree of freedom.

The sample space for this experiment is $\Omega = (0, \infty)$

Continuous Random Variables

Example: The chi-square distribution. The functions

- spreadsheet: `CHIINV(RAND(),1)`
- R: `rchisq(1,1)`

perform the following experiment:

Choose a number from the interval $(0, \infty)$ according to the chi-square distribution with one degree of freedom.

The sample space for this experiment is $\Omega = (0, \infty)$

The value of random variable X associated with the outcome is just the value chosen.

Continuous Random Variables

Example: The chi-square distribution. The functions

- spreadsheet: `CHIINV(RAND(),1)`
- R: `rchisq(1,1)`

perform the following experiment:

Choose a number from the interval $(0, \infty)$ according to the chi-square distribution with one degree of freedom.

The sample space for this experiment is $\Omega = (0, \infty)$

The value of random variable X associated with the outcome is just the value chosen.

The continuous probability distribution associated with this experiment is called the **chi-square** distribution (with one degree of freedom).

Events

The approach we took with discrete random variables of considering every possible subset of Ω to be an event does not work when Ω is uncountably infinite (which every interval is).

Events

The approach we took with discrete random variables of considering every possible subset of Ω to be an event does not work when Ω is uncountably infinite (which every interval is).

For a continuous random variable, we have to restrict our definition of an event to include only certain subsets of Ω.

Fortunately, for a continuous random variable X, events of interest nearly always fall into one of three types:

- The value of X is less than or equal to some value a
- The value of X lies between two values a and b
- The value of X is greater than or equal to some value b

Probability Density Functions

If X is a continuous random variable, the *probability density function* (pdf) of X is a function $f(x)$ with the property that, for any $a \leq b$,

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

Probability Density Functions

If X is a continuous random variable, the *probability density function* (pdf) of X is a function $f(x)$ with the property that, for any $a \leq b$,

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

Any function with the following properties is a valid pdf:

$$f(x) \geq 0 \quad \text{for all } x \in \Omega$$

$$\int_\Omega f(x)\, dx = 1$$

Probability Density Functions

A curious fact about continuous random variables is that the probability of any exact value is zero:

$$P(X = a) \; = \; P(a \leq X \leq a) = \int_a^a f(x)\, dx \; = \; 0$$

Probability Density Functions

A curious fact about continuous random variables is that the probability of any exact value is zero:

$$P(X = a) \;=\; P(a \le X \le a) = \int_a^a f(x)\,dx \;=\; 0$$

In a probability experiment with a discrete sample space, an event with probability zero cannot occur.

Probability Density Functions

A curious fact about continuous random variables is that the probability of any exact value is zero:

$$P(X = a) \; = \; P(a \leq X \leq a) = \int_a^a f(x)\,dx \; = \; 0$$

In a probability experiment with a discrete sample space, an event with probability zero cannot occur.

However, in a probability experiment with a continuous sample space, an event with probability zero actually occurs every time we perform the experiment.

Probability Density Functions

A curious fact about continuous random variables is that the probability of any exact value is zero:

$$P(X = a) \; = \; P(a \leq X \leq a) = \int_a^a f(x)\, dx \; = \; 0$$

In a probability experiment with a discrete sample space, an event with probability zero cannot occur.

However, in a probability experiment with a continuous sample space, an event with probability zero actually occurs every time we perform the experiment.

For most people this idea takes some getting used to.

Cumulative Distribution Functions

If X is a continuous random variable with probability density function f, the *cumulative distribution function* (cdf) F is defined as:

$$F(x) \; = \; P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$$

Cumulative Distribution Functions

If X is a continuous random variable with probability density function f, the *cumulative distribution function* (cdf) F is defined as:

$$F(x) \; = \; P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$$

The cdf of a random variable always assumes a value in the interval $[0, 1]$, with

$$F(-\infty) = 0$$

and

$$F(\infty) = 1$$

Expected Value and Variance

We calculate the expected value of a continuous random variable in much the same way as a discrete one. Suppose:

- X is a continuous random variable with probability density function f
- Y is a discrete random variable with probability mass function p

Expected Value and Variance

We calculate the expected value of a continuous random variable in much the same way as a discrete one. Suppose:

- X is a continuous random variable with probability density function f
- Y is a discrete random variable with probability mass function p

The expected values of Y and X are defined by:

$$E(Y) = \sum_{Range(Y)} y \cdot p(y) \qquad E(X) = \int_{\Omega} x \cdot f(x)\, dx$$

Expected Value and Variance

We calculate the expected value of a continuous random variable in much the same way as a discrete one. Suppose:

- X is a continuous random variable with probability density function f
- Y is a discrete random variable with probability mass function p

The expected values of Y and X are defined by:

$$E(Y) = \sum_{Range(Y)} y \cdot p(y) \qquad E(X) = \int_{\Omega} x \cdot f(x)\, dx$$

As with event probabilities, we rarely need to compute these from the definition. Usually there is a well-known formula that we can look up when we need it.

Expected Value and Variance

We calculate the variance of a continuous random variable in much the same way as a discrete one. As before, suppose:

- X is a continuous random variable with probability density function f
- Y is a discrete random variable with probability mass function p

Expected Value and Variance

We calculate the variance of a continuous random variable in much the same way as a discrete one. As before, suppose:

- X is a continuous random variable with probability density function f
- Y is a discrete random variable with probability mass function p

The variances of Y and X are defined by:

$$V(Y) = \sum_{Range(Y)} (y - \mu_y)^2 \cdot p(y) \qquad V(X) = \int_{\Omega} (x - \mu_x)^2 \cdot f(x)\, dx$$

Expected Value and Variance

We calculate the variance of a continuous random variable in much the same way as a discrete one. As before, suppose:

- X is a continuous random variable with probability density function f
- Y is a discrete random variable with probability mass function p

The variances of Y and X are defined by:

$$V(Y) = \sum_{Range(Y)} (y - \mu_y)^2 \cdot p(y) \qquad V(X) = \int_\Omega (x - \mu_x)^2 \cdot f(x)\, dx$$

Expected Value and Variance

We calculate the variance of a continuous random variable in much the same way as a discrete one. As before, suppose:

- X is a continuous random variable with probability density function f
- Y is a discrete random variable with probability mass function p

Expected Value and Variance

We calculate the variance of a continuous random variable in much the same way as a discrete one. As before, suppose:

- X is a continuous random variable with probability density function f
- Y is a discrete random variable with probability mass function p

The variances of Y and X are defined by:

$$V(Y) = \sum_{Range(Y)} (y - \mu_y)^2 \cdot p(y) \qquad V(X) = \int_\Omega (x - \mu_x)^2 \cdot f(x)\, dx$$

Covariance

If $X = [X_1, X_2]'$ represents a bivariate continuous random variable, there will be a joint probability density function

$$f(x_1, x_2)$$

and the $Cov(X_1, X_2)$ will be defined by the double integral:

$$Cov(X_1, X_2) = \int\int_{\Omega} (x_1 - \mu_1)(x_2 - \mu_2) f(x_1, x_2)\, dx_1\, dx_2$$

Covariance

If $X = [X_1, X_2]'$ represents a bivariate continuous random variable, there will be a joint probability density function

$$f(x_1, x_2)$$

and the $Cov(X_1, X_2)$ will be defined by the double integral:

$$Cov(X_1, X_2) = \int\int_{\Omega} (x_1 - \mu_1)(x_2 - \mu_2) f(x_1, x_2)\, dx_1\, dx_2$$

Fortunately, we will rarely need to compute a covariance from the definition.

Covariance

If $X = [X_1, X_2]'$ represents a bivariate continuous random variable, there will be a joint probability density function

$$f(x_1, x_2)$$

and the $Cov(X_1, X_2)$ will be defined by the double integral:

$$Cov(X_1, X_2) = \int\int_\Omega (x_1 - \mu_1)(x_2 - \mu_2) f(x_1, x_2)\, dx_1\, dx_2$$

Fortunately, we will rarely need to compute a covariance from the definition.

Our results for expected values and variances of linear combinations of variables extend with no changes to continuous random variables.

Common Continuous Distributions

The following slides contain a quick tour of some common continuous distributions.

Common Continuous Distributions

The following slides contain a quick tour of some common continuous distributions.

There are many, many continuous distributions, and preference has been given to the ones that occur in statistics.

The Uniform Distribution

The uniform distribution arises from the following probability experiment: Select a number randomly from the interval $[0, 1]$, with each number having an equal chance of being chosen.

The Uniform Distribution

The uniform distribution arises from the following probability experiment: Select a number randomly from the interval $[0, 1]$, with each number having an equal chance of being chosen.

The characteristics of the uniform distribution are:

- sample space: $\Omega = [0, 1]$
- parameters: none
- probability density function: $f(x) = 1, \quad 0 \leq x \leq 1$
- expected value $E(X)$: $1/2$
- variance $V(X)$: $1/12$
- R cumulative distribution function $F(x)$: `punif(x)`

The Normal Distribution

The normal or bell curve distribution can arise from adding independent random variables.

The Normal Distribution

The normal or bell curve distribution can arise from adding independent random variables.

The characteristics of the uniform distribution are:

- sample space: $\Omega = (-\infty, \infty)$
- parameters: mean μ standard deviation σ
- expected value $E(X)$: μ
- variance $V(X)$: σ^2
- R cumulative distribution function $F(x)$: `pnorm(x,`μ`,`σ`)`

The Normal Distribution

The normal or bell curve distribution can arise from adding independent random variables.

The characteristics of the uniform distribution are:

- sample space: $\Omega = (-\infty, \infty)$
- parameters: mean μ standard deviation σ
- expected value $E(X)$: μ
- variance $V(X)$: σ^2
- R cumulative distribution function $F(x)$: `pnorm(x,`μ`,`σ`)`

The case $\mu = 0$ and $\sigma = 1$ is called the **standard normal** distribution.

The Chi-square Distribution

The square of a standard normal random variable has a chi-square distribution with one degree of freedom.

The Chi-square Distribution

The square of a standard normal random variable has a chi-square distribution with one degree of freedom.

The characteristics of the uniform distribution are:

- sample space: $\Omega = (0, \infty)$
- parameters: degrees of freedom n
- expected value $E(X)$: n
- variance $V(X)$: $2n$
- R cumulative distribution function $F(x)$: `pchisq(x,n)`

The Chi-square Distribution

The square of a standard normal random variable has a chi-square distribution with one degree of freedom.

The characteristics of the uniform distribution are:

- sample space: $\Omega = (0, \infty)$
- parameters: degrees of freedom n
- expected value $E(X)$: n
- variance $V(X)$: $2n$
- R cumulative distribution function $F(x)$: `pchisq(x,n)`

The sum of the squares of k independent standard normal variates is a chi-square with k degrees of freedom.

The t Distribution

The ratio of a standard normal random variable to the square root of an independent chi-square variate divided by its degrees of freedom has a t distribution.

The t Distribution

The ratio of a standard normal random variable to the square root of an independent chi-square variate divided by its degrees of freedom has a t distribution.

The characteristics of the uniform distribution are:

- sample space: $\Omega = (-\infty, \infty)$
- parameters: degrees of freedom n, $n = 3, 4, 5, \ldots$
- expected value $E(X)$: 0
- variance $V(X)$: $n/(n-2)$ for $n > 2$
- R cumulative distribution function $F(x)$: `pt(x,n)`

The t Distribution

The ratio of a standard normal random variable to the square root of an independent chi-square variate divided by its degrees of freedom has a t distribution.

The characteristics of the uniform distribution are:

- sample space: $\Omega = (-\infty, \infty)$
- parameters: degrees of freedom n, $n = 3, 4, 5, \ldots$
- expected value $E(X)$: 0
- variance $V(X)$: $n/(n-2)$ for $n > 2$
- R cumulative distribution function $F(x)$: `pt(x,n)`

Quickly converges to a standard normal as n becomes large.

The F Distribution

The ratio of two independent chi-square random variables divided by their respective degrees of freedom has an F distribution.

The F Distribution

The ratio of two independent chi-square random variables divided by their respective degrees of freedom has an F distribution.

The characteristics of the uniform distribution are:

- sample space: $\Omega = (0, \infty)$
- parameters: degrees of freedom in the numerator n_1 and denominator n_2
- expected value $E(X)$: $n_2/(n_2 - 2)$ for $n_2 > 2$
- variance $V(X)$: (complicated formula)
- R cumulative distribution function $F(x)$: `pf(x,`n_1`,`n_2`)`