Continuous Random Variables

James Heald[1]

[1]Gatsby Computational Neuroscience Unit
University College London

Gatsby Bridging Programme 2023

GATSBY

Objectives

- Introduce the concept and formal definition of a continuous random variable X and a probability density function.
- Learn how to find the probability that a continuous random variable falls in some interval [a, b].
- Learn that if X is continuous, the probability that X takes on any specific value is 0.
- Introduce the concept and formal definition of a cumulative distribution function of a continuous random variable.
- Learn how to find the cumulative distribution function of a continuous random variable X from the probability density function of X.

Discrete vs. continuous random variables

Unlike discrete random variables, which can take on a finite or countable number of possible values (e.g. faces of a die or cards of a deck), continuous random variables can take on an uncountable number of possible values (e.g. all the real numbers in an interval).

Discrete vs. continuous random variables

Unlike discrete random variables, which can take on a finite or countable number of possible values (e.g. faces of a die or cards of a deck), continuous random variables can take on an uncountable number of possible values (e.g. all the real numbers in an interval).

Examples

- the voltage membrane potential of a cell
- the interspike interval of a neuron
- the force generated by a muscle
- the velocity of an eye movement

Discrete vs. continuous random variables

Unlike discrete random variables, which can take on a finite or countable number of possible values (e.g. faces of a die or cards of a deck), continuous random variables can take on an uncountable number of possible values (e.g. all the real numbers in an interval).

Examples

- the voltage membrane potential of a cell
- the interspike interval of a neuron
- the force generated by a muscle
- the velocity of an eye movement

Many concepts introduced for discrete random variables (e.g. probability mass functions, cumulative distributions functions) have analogs in the continuous setting (sums replaced by integrals).

Continuous random variables

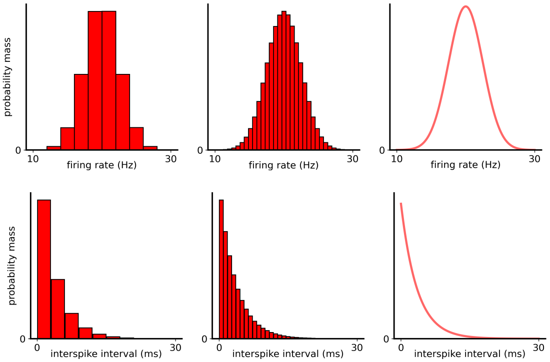
Definition

A random variable X is continuous if:

1. possible values comprise either a single interval on the number line (i.e. for some $a < b$, any number x between a and b is a possible value) or a union of disjoint intervals, and
2. $P(X = c) = 0$ for any number c that is a possible value of X.

Discrete probability distributions in the limit

Continuous random variables can be discretised into bins to form a discrete distribution that can be viewed as a probability histogram. As the bins become narrower, the histogram approaches a smooth curve.



The probability density function

Definition

The **probability density function** (PDF) of a continuous random variable X is a function $f(x)$ defined on the interval $(-\infty, \infty)$ such that for any two numbers a and b with $a \leq b$,

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

That is, the probability that X takes on a value in the interval [a, b] is the area under the graph of the density function above this interval.

The probability density function

Definition

The **probability density function** (PDF) of a continuous random variable X is a function $f(x)$ defined on the interval $(-\infty, \infty)$ such that for any two numbers a and b with $a \leq b$,

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

That is, the probability that X takes on a value in the interval [a, b] is the area under the graph of the density function above this interval.

A valid probability density function $f(x)$ must have the following properties to respect the axioms of probability:

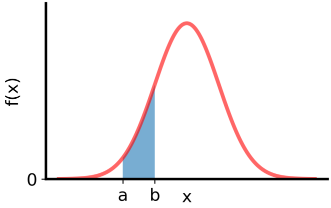
$$f(x) \geq 0 \text{ for all } x \tag{1}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1. \tag{2}$$

Probabilities as integrals

The probability that a continuous random variable X takes on a value in the interval [a, b] is given by the area under the probability density function $f(x)$:

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



Density as probability per unit of x

If $f(x)$ is not the probability of x, what is it?

Density as probability per unit of x

If $f(x)$ is not the probability of x, what is it?

The probability that X will lie in an infinitesimal interval dx about x is $f(x)dx$:

$$\begin{aligned} P(x \leq X \leq x + dx) &= \int_{x}^{x+dx} f(t)dt \\ &= f(x)dx \end{aligned}$$

Density as probability per unit of x

If $f(x)$ is not the probability of x, what is it?

The probability that X will lie in an infinitesimal interval dx about x is $f(x)dx$:

$$\begin{aligned} P(x \leq X \leq x + dx) &= \int_x^{x+dx} f(t)dt \\ &= f(x)dx \end{aligned}$$

Thus, density is probability per unit of x (rate of probability accumulation):

$$\frac{P(x \leq X \leq x + dx)}{dx} = f(x)$$

Each possible value has zero probability

The probability that X takes on a particular value a is 0, as

$$\begin{aligned} P(X = a) &= \int_a^a f(x)dx \\ &= \lim_{\epsilon \to 0} \int_{a-\epsilon}^{a+\epsilon} f(x)dx \\ &= 0. \end{aligned}$$

Each possible value has zero probability

The probability that X takes on a particular value a is 0, as

$$\begin{aligned} P(X = a) &= \int_a^a f(x)dx \\ &= \lim_{\epsilon \to 0} \int_{a-\epsilon}^{a+\epsilon} f(x)dx \\ &= 0. \end{aligned}$$

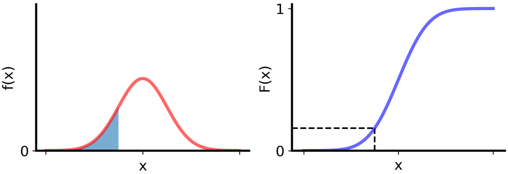
This implies that probabilities don't depend on interval end points:

$$P(a \leq X \leq b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b),$$

as $P(X = a) = P(X = b) = 0$.

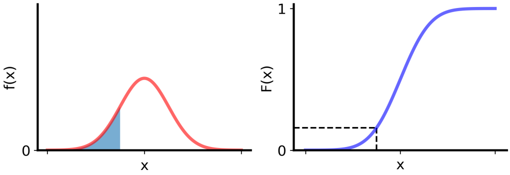
The cumulative distribution function

The cumulative distribution function (CDF) $F(x)$ is the area under the probability density function $f(x)$ to the left of x.



The cumulative distribution function

The cumulative distribution function (CDF) $F(x)$ is the area under the probability density function $f(x)$ to the left of x.



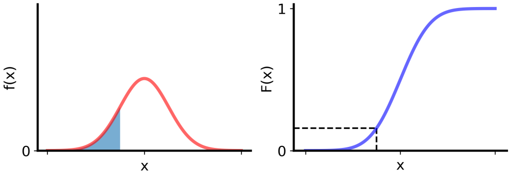
The CDF $F(x)$ has the following properties:

a. $F(x)$ is a non-decreasing (monotonic) function of x

b. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

The cumulative distribution function

The cumulative distribution function (CDF) $F(x)$ is the area under the probability density function $f(x)$ to the left of x.



The CDF $F(x)$ has the following properties:

a. $F(x)$ is a non-decreasing (monotonic) function of x

b. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Collectively, a. and b. imply that $F : \mathbb{R} \mapsto [0, 1]$.

The cumulative distribution function

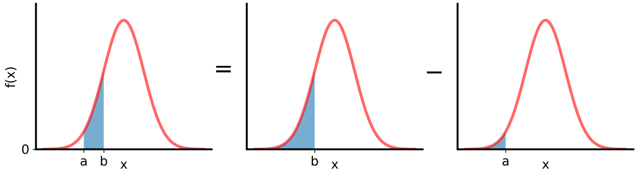
Definition

Let X be a continuous random variable with probability density function $f(x)$, then the **cumulative distribution function** $F(x)$ is defined as

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} f(t)dt. \end{aligned}$$

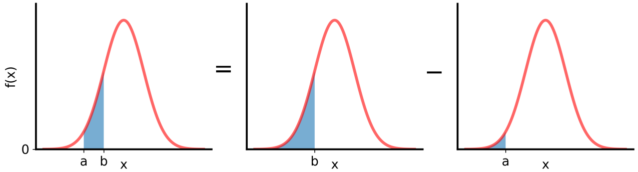
Computing probabilities using the CDF

$$P(a \leq X \leq b) = F(b) - F(a).$$

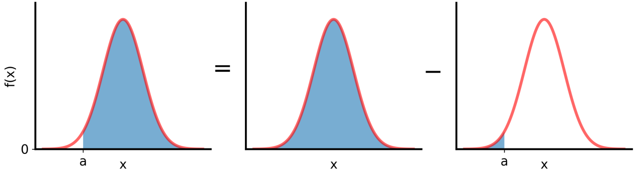


Computing probabilities using the CDF

$$P(a \leq X \leq b) = F(b) - F(a).$$



$$\begin{aligned} P(X > a) &= F(\infty) - F(a) \\ &= 1 - F(a). \end{aligned}$$



Obtaining the PDF from the CDF

How would you obtain the PDF from the CDF?

Obtaining the PDF from the CDF

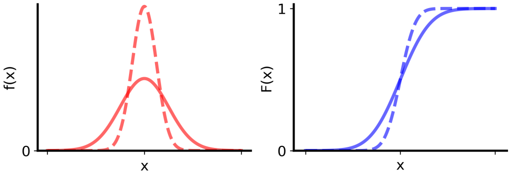
How would you obtain the PDF from the CDF?

The CDF is the integral of the PDF, and so the PDF is the derivative of the CDF.

Obtaining the PDF from the CDF

How would you obtain the PDF from the CDF?

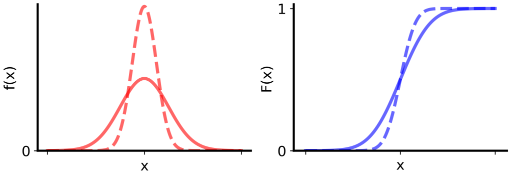
The CDF is the integral of the PDF, and so the PDF is the derivative of the CDF.



Obtaining the PDF from the CDF

How would you obtain the PDF from the CDF?

The CDF is the integral of the PDF, and so the PDF is the derivative of the CDF.



At every x at which the derivative $\frac{\delta F(x)}{\delta x}$ exists, $\frac{\delta F(x)}{\delta x} = f(x)$.

Example: the uniform distribution

Definition

A random variable X is said to have a **uniform distribution** on the interval $[a, b]$ if the PDF of X is

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

The CDF of X is given by

$$F(x; a, b) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b. \end{cases}$$

When X has a uniform distribution on $[a, b]$, we write this as $X \sim \mathrm{U}(a, b)$.

Example: the uniform distribution

Definition

A random variable X is said to have a **uniform distribution** on the interval $[a, b]$ if the PDF of X is

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$

The CDF of X is given by

$$F(x; a, b) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b \\ 1 & \text{if } x > b. \end{cases}$$

When X has a uniform distribution on $[a, b]$, we write this as $X \sim \mathrm{U}(a, b)$.

Example

When X has a uniform distribution on the interval [a, b], for $a \leq x \leq b$:

$$\frac{\delta F(x)}{\delta x} = \frac{\delta}{\delta x}\left(\frac{x-a}{b-a}\right) = \frac{1}{b-a} = f(x)$$

Percentiles of a continuous distribution

Definition

Let p be a number between 0 and 1. The **(100p)th percentile** of the distribution of a continuous random variable X, denoted by $\eta(p)$, is defined by

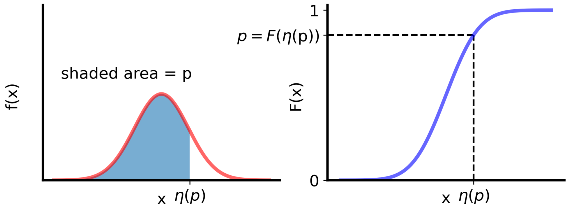
$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(x)dx$$

Percentiles of a continuous distribution

Definition

Let p be a number between 0 and 1. The **(100p)th percentile** of the distribution of a continuous random variable X, denoted by $\eta(p)$, is defined by

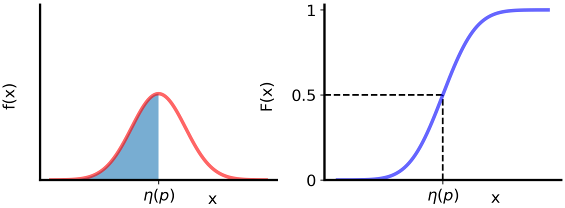
$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(x)dx$$



Median

Definition

The **median** of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile, so $\tilde{\mu}$ satisfies $F(\tilde{\mu}) = 0.5$. That is, half the are area under the probability density function is to the left of $\tilde{\mu}$ and half is to the right of $\tilde{\mu}$.



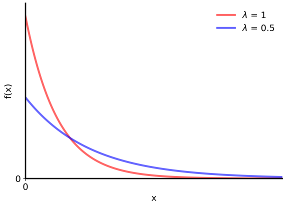
Example: the exponential distribution

Definition

A random variable X is said to have an **exponential distribution** on the interval $[0, \infty)$ if the PDF of X is:

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda > 0$ is a rate parameter that governs the rate of decay of $f(x)$. When X has an exponential distribution with parameter λ, we write $X \sim \mathrm{Exp}(\lambda)$.



Example: the exponential distribution

Example

Write down the formula for the $(100p)$th percentile of the distribution of $X \sim \mathrm{Exp}\,(\lambda)$, and use it to find the median of X.

Example: the exponential distribution

Example

Write down the formula for the $(100p)$th percentile of the distribution of $X \sim \mathrm{Exp}(\lambda)$, and use it to find the median of X.

$$p = \int_0^{\eta(p)} \lambda e^{-\lambda x} dx$$

Example: the exponential distribution

Example

Write down the formula for the $(100p)$th percentile of the distribution of $X \sim \mathrm{Exp}(\lambda)$, and use it to find the median of X.

$$\begin{aligned} p &= \int_0^{\eta(p)} \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x}\Big|_0^{\eta(p)} \end{aligned}$$

Example

Write down the formula for the $(100p)$th percentile of the distribution of $X \sim \mathrm{Exp}(\lambda)$, and use it to find the median of X.

$$\begin{aligned} p &= \int_0^{\eta(p)} \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x}\Big|_0^{\eta(p)} \\ &= 1 - e^{-\lambda \eta(p)}. \end{aligned}$$

Example: the exponential distribution

Example

Write down the formula for the $(100p)$th percentile of the distribution of $X \sim \mathrm{Exp}(\lambda)$, and use it to find the median of X.

$$\begin{aligned} p &= \int_0^{\eta(p)} \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x}\Big|_0^{\eta(p)} \\ &= 1 - e^{-\lambda \eta(p)}. \end{aligned}$$

Therefore,

$$\eta(p) = -\frac{\log(1-p)}{\lambda}$$

Example: the exponential distribution

Example

Write down the formula for the $(100p)$th percentile of the distribution of $X \sim \mathrm{Exp}(\lambda)$, and use it to find the median of X.

$$\begin{aligned} p &= \int_0^{\eta(p)} \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x}\Big|_0^{\eta(p)} \\ &= 1 - e^{-\lambda \eta(p)}. \end{aligned}$$

Therefore,

$$\eta(p) = -\frac{\log(1-p)}{\lambda}$$

and

$$\eta(0.5) = -\frac{\log(0.5)}{\lambda}.$$

Mean and variance

The expected value (mean) of a continuous random variable X is:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x.$$

Mean and variance

The expected value (mean) of a continuous random variable X is:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x.$$

The expected value of a function $g(x)$ of X is:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \mathrm{d}x.$$

Mean and variance

The expected value (mean) of a continuous random variable X is:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x.$$

The expected value of a function $g(x)$ of X is:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \mathrm{d}x.$$

The variance of X is:

$$\begin{aligned}\mathrm{Var}[X] &= \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 f(x) \mathrm{d}x \\ &= \mathbb{E}[(x - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2.\end{aligned}$$

Example: the uniform distribution

Example

When X has a uniform distribution on the interval [a,b], its expected value is:

$$\mathbb{E}[X] = \int_a^b x \frac{1}{b-a} \mathrm{d}x$$

Example: the uniform distribution

Example

When X has a uniform distribution on the interval [a,b], its expected value is:

$$\begin{aligned}\mathbb{E}[X] &= \int_a^b x \frac{1}{b-a} \mathrm{d}x \\ &= \frac{b^2 - a^2}{2(b-a)}\end{aligned}$$

Example: the uniform distribution

Example

When X has a uniform distribution on the interval [a,b], its expected value is:

$$\begin{aligned}\mathbb{E}[X] &= \int_a^b x\frac{1}{b-a}\mathrm{d}x \\ &= \frac{b^2-a^2}{2(b-a)} \\ &= \frac{a+b}{2},\end{aligned}$$

Example: the uniform distribution

Example

When X has a uniform distribution on the interval [a,b], its expected value is:

$$\begin{aligned}\mathbb{E}[X] &= \int_a^b x \frac{1}{b-a} \mathrm{d}x \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{a+b}{2},\end{aligned}$$

and its variance is:

$$\mathrm{Var}[X] = \int_a^b x^2 \frac{1}{b-a} \mathrm{d}x - \left(\frac{a+b}{2}\right)^2$$

Example: the uniform distribution

Example

When X has a uniform distribution on the interval [a,b], its expected value is:

$$\begin{aligned}\mathbb{E}[X] &= \int_a^b x \frac{1}{b-a} \mathrm{d}x \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{a+b}{2},\end{aligned}$$

and its variance is:

$$\begin{aligned}\mathrm{Var}[X] &= \int_a^b x^2 \frac{1}{b-a} \mathrm{d}x - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2\end{aligned}$$

Example: the uniform distribution

Example

When X has a uniform distribution on the interval [a,b], its expected value is:

$$\begin{aligned}\mathbb{E}[X] &= \int_a^b x \frac{1}{b-a} \mathrm{d}x \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{a+b}{2},\end{aligned}$$

and its variance is:

$$\begin{aligned}\mathrm{Var}[X] &= \int_a^b x^2 \frac{1}{b-a} \mathrm{d}x - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{(b-a)^2}{12}.\end{aligned}$$

The probability integral transform

Theorem

Suppose X is a continuous random variable having PDF $f_X(x)$ and CDF $F_X(x)$, and suppose further that $f_X(x) > 0$ on an interval (a, b) and $f_X(x) = 0$ otherwise. Then, if $U \sim \mathrm{U}(0, 1)$, the random variable $Y = F_X^{-1}(U)$ has the same distribution as X, that is its CDF F_Y satisfies $F_Y(y) = F_X(y)$ for all y.

The probability integral transform

Theorem

Suppose X is a continuous random variable having PDF $f_X(x)$ and CDF $F_X(x)$, and suppose further that $f_X(x) > 0$ on an interval (a, b) and $f_X(x) = 0$ otherwise. Then, if $U \sim \mathrm{U}(0, 1)$, the random variable $Y = F_X^{-1}(U)$ has the same distribution as X, that is its CDF F_Y satisfies $F_Y(y) = F_X(y)$ for all y.

Proof.

$$F_Y(y) = P(Y \leq y)$$

The probability integral transform

Theorem

Suppose X is a continuous random variable having PDF $f_X(x)$ and CDF $F_X(x)$, and suppose further that $f_X(x) > 0$ on an interval (a, b) and $f_X(x) = 0$ otherwise. Then, if $U \sim \mathrm{U}(0, 1)$, the random variable $Y = F_X^{-1}(U)$ has the same distribution as X, that is its CDF F_Y satisfies $F_Y(y) = F_X(y)$ for all y.

Proof.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(F_X^{-1}(U) \leq y) \end{aligned}$$

The probability integral transform

Theorem

Suppose X is a continuous random variable having PDF $f_X(x)$ and CDF $F_X(x)$, and suppose further that $f_X(x) > 0$ on an interval (a, b) and $f_X(x) = 0$ otherwise. Then, if $U \sim \mathrm{U}(0, 1)$, the random variable $Y = F_X^{-1}(U)$ has the same distribution as X, that is its CDF F_Y satisfies $F_Y(y) = F_X(y)$ for all y.

Proof.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(F_X^{-1}(U) \leq y) \\ &= P(F_X(F_X^{-1}(U)) \leq F_X(y)) \end{aligned}$$

The probability integral transform

Theorem

Suppose X is a continuous random variable having PDF $f_X(x)$ and CDF $F_X(x)$, and suppose further that $f_X(x) > 0$ on an interval (a, b) and $f_X(x) = 0$ otherwise. Then, if $U \sim \mathrm{U}(0, 1)$, the random variable $Y = F_X^{-1}(U)$ has the same distribution as X, that is its CDF F_Y satisfies $F_Y(y) = F_X(y)$ for all y.

Proof.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(F_X^{-1}(U) \leq y) \\ &= P(F_X(F_X^{-1}(U)) \leq F_X(y)) \\ &= P(U \leq F_X(y)) \end{aligned}$$

The probability integral transform

Theorem

Suppose X is a continuous random variable having PDF $f_X(x)$ and CDF $F_X(x)$, and suppose further that $f_X(x) > 0$ on an interval (a, b) and $f_X(x) = 0$ otherwise. Then, if $U \sim \mathrm{U}(0, 1)$, the random variable $Y = F_X^{-1}(U)$ has the same distribution as X, that is its CDF F_Y satisfies $F_Y(y) = F_X(y)$ for all y.

Proof.

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(F_X^{-1}(U) \leq y) \\
&= P(F_X(F_X^{-1}(U)) \leq F_X(y)) \\
&= P(U \leq F_X(y)) \\
&= F_X(y) \quad \square
\end{aligned}
$$

The probability integral transform

Theorem

Suppose X is a continuous random variable having PDF $f_X(x)$ and CDF $F_X(x)$, and suppose further that $f_X(x) > 0$ on an interval (a, b) and $f_X(x) = 0$ otherwise. Then, if $U \sim \mathrm{U}(0, 1)$, the random variable $Y = F_X^{-1}(U)$ has the same distribution as X, that is its CDF F_Y satisfies $F_Y(y) = F_X(y)$ for all y.

Proof.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(F_X^{-1}(U) \leq y) \\ &= P(F_X(F_X^{-1}(U)) \leq F_X(y)) \\ &= P(U \leq F_X(y)) \\ &= F_X(y) \quad \square \end{aligned}$$

because

- F_X is strictly increasing on (a, b) and
- $P(U \leq F_X(y)) = F_X(y)$ when $U \sim \mathrm{U}(0, 1)$.

The inverse transform method

The probability integral transform forms the basis of the inverse transform method, a procedure for sampling a continuous random variable given the inverse of its cumulative distribution function.

The inverse transform method

The probability integral transform forms the basis of the inverse transform method, a procedure for sampling a continuous random variable given the inverse of its cumulative distribution function.

Algorithm The inverse transform method

Sample $u \sim \mathrm{U}(0, 1)$
Let $x = F^{-1}(u)$

The inverse transform method

The probability integral transform forms the basis of the inverse transform method, a procedure for sampling a continuous random variable given the inverse of its cumulative distribution function.

Algorithm The inverse transform method

Sample $u \sim \mathrm{U}(0, 1)$
Let $x = F^{-1}(u)$

The inverse transform method draws samples that are distributed as the cumulative distribution function F.

Example: the exponential distribution

Example

For $x \geq 0$, the PDF of the exponential distribution is:

$$f(x) = \lambda e^{-\lambda x},$$

Example: the exponential distribution

Example

For $x \geq 0$, the PDF of the exponential distribution is:

$$f(x) = \lambda e^{-\lambda x},$$

which implies the CDF is:

$$F(x) = 1 - e^{-\lambda x} = u,$$

Example: the exponential distribution

Example

For $x \geq 0$, the PDF of the exponential distribution is:

$$f(x) = \lambda e^{-\lambda x},$$

which implies the CDF is:

$$F(x) = 1 - e^{-\lambda x} = u,$$

and the inverse of the CDF is:

$$F^{-1}(u) = -\frac{\log(1-u)}{\lambda} = x.$$

Example: the exponential distribution

Example

For $x \geq 0$, the PDF of the exponential distribution is:

$$f(x) = \lambda e^{-\lambda x},$$

which implies the CDF is:

$$F(x) = 1 - e^{-\lambda x} = u,$$

and the inverse of the CDF is:

$$F^{-1}(u) = -\frac{\log(1-u)}{\lambda} = x.$$

Algorithm The inverse transform method for the exponential distribution

Sample $u \sim \mathrm{U}(0,1)$
Let $x = -\frac{\log(1-u)}{\lambda}$

The normal distribution ('bell-shaped curve')

Definition

A continuous random variable X has a **normal distribution** with parameters μ and σ (or σ^2), where $-\infty < \mu < \infty$ and $0 < \sigma$, if the probability density function of X is:

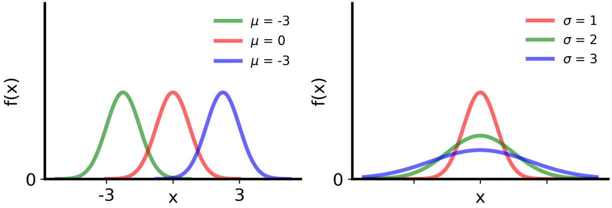
$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

The normal distribution ('bell-shaped curve')

Definition

A continuous random variable X has a **normal distribution** with parameters μ and σ (or σ^2), where $-\infty < \mu < \infty$ and $0 < \sigma$, if the probability density function of X is:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

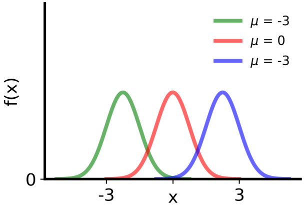


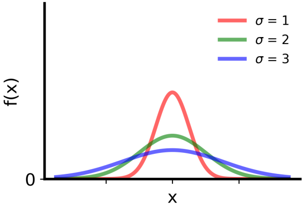
The normal distribution ('bell-shaped curve')

Definition

A continuous random variable X has a **normal distribution** with parameters μ and σ (or σ^2), where $-\infty < \mu < \infty$ and $0 < \sigma$, if the probability density function of X is:

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$





The normal distribution is the most important distribution in all of probability theory. It is ubiquitous in statistical analysis (central limit theorem).

Linear transformation of a normal random variable

Claim

If X is normally distributed with parameters μ and σ, then $Y = aX + b$ is normally distributed with parameters $a\mu + b$ and $a\sigma$.

Linear transformation of a normal random variable

Claim

If X is normally distributed with parameters μ and σ, then $Y = aX + b$ is normally distributed with parameters $a\mu + b$ and $a\sigma$.

Proof.

Let F_Y denote the CDF of $Y = aX + b$, then

$$F_Y(x) = P(aX + b \leq x)$$

Linear transformation of a normal random variable

Claim

If X is normally distributed with parameters μ and σ, then $Y = aX + b$ is normally distributed with parameters $a\mu + b$ and $a\sigma$.

Proof.

Let F_Y denote the CDF of $Y = aX + b$, then

$$\begin{aligned} F_Y(x) &= P(aX + b \leq x) \\ &= P(X \leq (x - b)/a) \end{aligned}$$

Linear transformation of a normal random variable

Claim

If X is normally distributed with parameters μ and σ, then $Y = aX + b$ is normally distributed with parameters $a\mu + b$ and $a\sigma$.

Proof.

Let F_Y denote the CDF of $Y = aX + b$, then

$$\begin{aligned} F_Y(x) &= P(aX + b \leq x) \\ &= P(X \leq (x - b)/a) \\ &= F_X((x - b)/a), \end{aligned}$$

where F_X is the CDF of X.

Claim

If X is normally distributed with parameters μ and σ, then $Y = aX + b$ is normally distributed with parameters $a\mu + b$ and $a\sigma$.

Proof.

Let F_Y denote the CDF of $Y = aX + b$, then

$$\begin{aligned} F_Y(x) &= P(aX + b \leq x) \\ &= P(X \leq (x - b)/a) \\ &= F_X((x - b)/a), \end{aligned}$$

where F_X is the CDF of X. By differentiation, the PDF of Y is

$$f_Y(x) = \frac{1}{a} f_X((x - b)/a)$$

$\square$

Linear transformation of a normal random variable

Claim

If X is normally distributed with parameters μ and σ, then $Y = aX + b$ is normally distributed with parameters $a\mu + b$ and $a\sigma$.

Proof.

Let F_Y denote the CDF of $Y = aX + b$, then

$$\begin{aligned} F_Y(x) &= P(aX + b \leq x) \\ &= P(X \leq (x - b)/a) \\ &= F_X((x - b)/a), \end{aligned}$$

where F_X is the CDF of X. By differentiation, the PDF of Y is

$$\begin{aligned} f_Y(x) &= \frac{1}{a} f_X((x - b)/a) \\ &= \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\frac{x-b}{a} - \mu}{\sigma}\right)^2} \end{aligned}$$

□

Linear transformation of a normal random variable

Claim

If X is normally distributed with parameters μ and σ, then $Y = aX + b$ is normally distributed with parameters $a\mu + b$ and $a\sigma$.

Proof.

Let F_Y denote the CDF of $Y = aX + b$, then

$$\begin{aligned} F_Y(x) &= P(aX + b \leq x) \\ &= P(X \leq (x-b)/a) \\ &= F_X((x-b)/a), \end{aligned}$$

where F_X is the CDF of X. By differentiation, the PDF of Y is

$$\begin{aligned} f_Y(x) &= \frac{1}{a} f_X((x-b)/a) \\ &= \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\frac{x-b}{a}-\mu}{\sigma}\right)^2} \\ &= \frac{1}{a\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-(a\mu+b)}{a\sigma}\right)^2}. \end{aligned}$$

$\square$

The standard normal random variable

When X is a normal random variable with parameters μ and σ, the computation of $P(a \leq X \leq b)$ requires evaluating

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

The standard normal random variable

When X is a normal random variable with parameters μ and σ, the computation of $P(a \leq X \leq b)$ requires evaluating

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx.$$

This cannot be calculated in closed form. However, for $\mu = 0$ and $\sigma = 1$, this integral has been approximated and tabulated for certain values of a and b. This table can also be used to compute probabilities for any other values of μ and σ.

The standard normal random variable

When X is a normal random variable with parameters μ and σ, the computation of $P(a \leq X \leq b)$ requires evaluating

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx.$$

This cannot be calculated in closed form. However, for $\mu = 0$ and $\sigma = 1$, this integral has been approximated and tabulated for certain values of a and b. This table can also be used to compute probabilities for any other values of μ and σ.

Definition

The normal distribution with parameter values $\mu = 0$ and $\sigma = 1$ is called the **standard normal distribution**. A random variable Z having a standard normal distribution is called a **standard normal random variable** and has probability density function given by

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}.$$

Tabulated CDF of the standard normal distribution

TABLE 5.1: AREA $\Phi(x)$ UNDER THE STANDARD NORMAL CURVE TO THE LEFT OF X

X	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Standardising a normal random variable

Probabilities involving a nonstandard normal random variable are computed by standardising.

Standardising a normal random variable

Probabilities involving a nonstandard normal random variable are computed by standardising.

Proposition

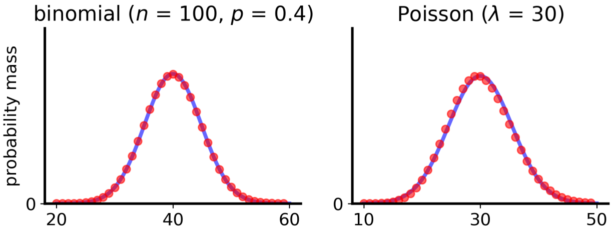
If X has a normal distribution with mean μ and standard deviation σ, then the **standardised variable** $Z = (X - \mu)/\sigma$ *has a standard normal distribution. Thus*

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right), \end{aligned}$$

where Φ is the cumulative distribution function of a standard normal random variable.

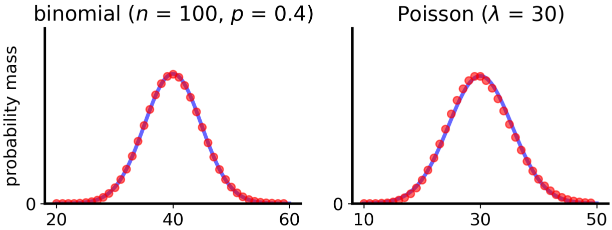
Approximating the binomial and Poisson distributions

The binomial and Poisson distributions are approximately normal for large n or large λ.



Approximating the binomial and Poisson distributions

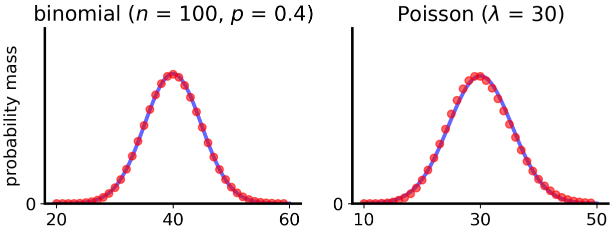
The binomial and Poisson distributions are approximately normal for large n or large λ.



For the binomial with parameters n and p, $\mu = np$ and $\sigma = \sqrt{np(1-p)}$, and for the Poisson with parameter λ, $\mu = \lambda$ and $\sigma = \sqrt{\lambda}$.

Approximating the binomial and Poisson distributions

The binomial and Poisson distributions are approximately normal for large n or large λ.



For the binomial with parameters n and p, $\mu = np$ and $\sigma = \sqrt{np(1-p)}$, and for the Poisson with parameter λ, $\mu = \lambda$ and $\sigma = \sqrt{\lambda}$.

These approximations are a great convenience, especially in conjunction with the '$2/3 - 95\%$ rule'.

The exponential distribution: memorylessness

Theorem

A random variable X satisfies $X \sim \mathrm{Exp}(\lambda)$ for some $\lambda > 0$ if for all positive t and h the following equation is satisfied:

$$P(X > t + h | X > t) = P(X > h),$$

that is, if X is memoryless.

The exponential distribution: memorylessness

Theorem

A random variable X satisfies $X \sim \mathrm{Exp}(\lambda)$ for some $\lambda > 0$ if for all positive t and h the following equation is satisfied:

$$P(X > t + h | X > t) = P(X > h),$$

that is, if X is memoryless.

Proof.

$$P(X > t + h | X > t) = \frac{P(X > t + h, X > t)}{P(X > t)}$$

□

The exponential distribution: memorylessness

Theorem

A random variable X satisfies $X \sim \mathrm{Exp}(\lambda)$ for some $\lambda > 0$ if for all positive t and h the following equation is satisfied:

$$P(X > t + h | X > t) = P(X > h),$$

that is, if X is memoryless.

Proof.

$$\begin{aligned} P(X > t + h | X > t) &= \frac{P(X > t + h, X > t)}{P(X > t)} \\ &= \frac{P(X > t + h)}{P(X > t)} \end{aligned}$$

$\square$

The exponential distribution: memorylessness

Theorem

A random variable X satisfies $X \sim \mathrm{Exp}\,(\lambda)$ for some $\lambda > 0$ if for all positive t and h the following equation is satisfied:

$$P(X > t + h | X > t) = P(X > h),$$

that is, if X is memoryless.

Proof.

$$\begin{aligned} P(X > t + h | X > t) &= \frac{P(X > t + h, X > t)}{P(X > t)} \\ &= \frac{P(X > t + h)}{P(X > t)} \\ &= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} \end{aligned}$$

$\square$

The exponential distribution: memorylessness

Theorem

A random variable X satisfies $X \sim \mathrm{Exp}(\lambda)$ for some $\lambda > 0$ if for all positive t and h the following equation is satisfied:

$$P(X > t + h | X > t) = P(X > h),$$

that is, if X is memoryless.

Proof.

$$\begin{aligned} P(X > t + h | X > t) &= \frac{P(X > t + h, X > t)}{P(X > t)} \\ &= \frac{P(X > t + h)}{P(X > t)} \\ &= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} \\ &= e^{-\lambda h} \end{aligned}$$

□

The exponential distribution: memorylessness

Theorem

A random variable X satisfies $X \sim \mathrm{Exp}(\lambda)$ for some $\lambda > 0$ if for all positive t and h the following equation is satisfied:

$$P(X > t + h | X > t) = P(X > h),$$

that is, if X is memoryless.

Proof.

$$\begin{aligned} P(X > t + h | X > t) &= \frac{P(X > t + h, X > t)}{P(X > t)} \\ &= \frac{P(X > t + h)}{P(X > t)} \\ &= \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} \\ &= e^{-\lambda h} \\ &= P(X > h). \end{aligned}$$

□