Lecture 06: Continuous Distributions

Mathematical Statistics I, MATH 60061/70061

Thursday September 16, 2021

Reference: Casella & Berger, 3.3

Normal distribution

The **Normal distribution** is a famous continuous distribution with a bell-shaped PDF.

It is extremely widely used in statistics because of the **central limit theorem**: "Under very weak assumptions, the sum of a large number of **independent and identically distributed** (i.i.d.) random variables has an approximately Normal distribution, regardless of the distribution of the individual random variables."

Standard Normal distribution

A continuous random variable Z is said to have the **standard Normal distribution** if its PDF φ is given by

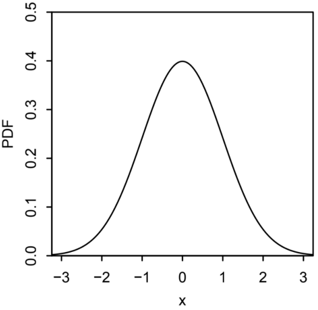
$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty.$$

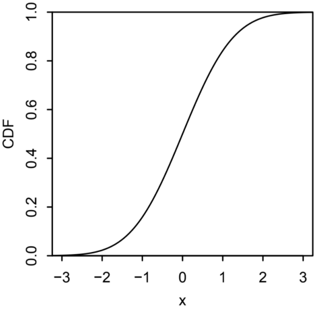
We write this as $Z \sim \mathcal{N}(0, 1)$. Z has mean 0 and variance 1.

The standard Normal CDF Φ is the accumulated area under the PDF:

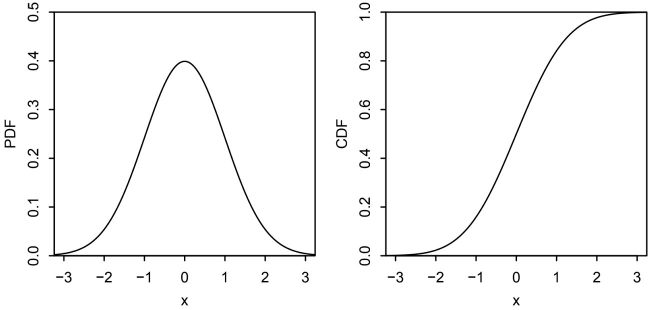
$$\Phi(z) = \int_{-\infty}^{z} \varphi(t) dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Standard Normal PDF and CDF





Standard Normal PDF and CDF



- Symmetry of PDF: $\varphi(z) = \varphi(-z)$.
- Symmetry of tail areas: $\Phi(z) = 1 - \Phi(-z)$.
- Symmetry of Z and $-Z$: if $Z \sim \mathcal{N}(0, 1)$, then $-Z \sim \mathcal{N}(0, 1)$

$$P(-Z \leq z) = P(Z \geq -z) = 1 - \Phi(-z) = \Phi(z).$$

Validity of the standard Normal PDF

Show $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 1$.

Validity of the standard Normal PDF

Show $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 1$.

$$\begin{aligned}
\left(\int_{-\infty}^{\infty} e^{-z^2/2} dz\right)\left(\int_{-\infty}^{\infty} e^{-z^2/2} dz\right) &= \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2} dy\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy \\
&= \int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2/2} rdrd\theta \\
&= \int_{0}^{2\pi}\left(\int_{0}^{\infty} e^{-r^2/2} rdr\right) d\theta \\
&= \int_{0}^{2\pi} 1 d\theta = 2\pi.
\end{aligned}$$

Therefore,

$$\int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi}.$$

Expectation and variance of $Z \sim \mathcal{N}(0, 1)$

Expectation: $E(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz = 0$ [odd function]

By LOTUS,

$$
\begin{aligned}
\mathrm{Var}(Z) &= E(Z^2) - (EZ)^2 = E(Z^2) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz \qquad \text{[even function]} \\
&= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} z^2 e^{-z^2/2} dz
\end{aligned}
$$

Using integration by parts with $u = z$ and $dv = z e^{-z^2/2} dz$, so $du = dz$ and $v = -e^{-z^2/2}$:

$$
\begin{aligned}
\mathrm{Var}(Z) &= \frac{2}{\sqrt{2\pi}} \left(-z e^{-z^2/2} \Big|_0^\infty + \int_0^\infty e^{-z^2/2} dz \right) \\
&= \frac{2}{\sqrt{2\pi}} \left(0 + \frac{\sqrt{2\pi}}{2} \right) \\
&= 1.
\end{aligned}
$$

Normal distribution

If $Z \sim \mathcal{N}(0,1)$, then $X = \mu + \sigma Z$ is said to have the **Normal distribution** with mean μ and variance σ^2, for any real μ and σ^2 with $\sigma > 0$. We denote this by $X \sim \mathcal{N}(\mu, \sigma^2)$.

Expectation and variance of X:

$$E(\mu + \sigma Z) = E(\mu) + \sigma E(Z) = \mu,$$
$$\mathrm{Var}(\mu + \sigma Z) = \mathrm{Var}(\sigma Z) = \sigma^2 \mathrm{Var}(Z) = \sigma^2.$$

The standardized version of X is

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1).$$

Normal CDF and PDF

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then the CDF of X is

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

and the PDF of X is

$$f(x) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}.$$

Normal CDF and PDF

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then the CDF of X is

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

and the PDF of X is

$$f(x) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}.$$

CDF:

$$F(x) = P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

PDF:

$$\begin{aligned} f(x) &= \frac{d}{dx}\Phi\left(\frac{x-\mu}{\sigma}\right) = \varphi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma} \qquad \text{[chain rule]} \\ &= \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \end{aligned}$$

68-95-99.7% rule

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\begin{aligned} P(|X-\mu| < \sigma) &\approx 0.68, \\ P(|X-\mu| < 2\sigma) &\approx 0.95, \\ P(|X-\mu| < 3\sigma) &\approx 0.997. \end{aligned}$$

After standardization,

$$\begin{aligned} P(|Z| < 1) &\approx 0.68, \\ P(|Z| < 2) &\approx 0.95, \\ P(|Z| < 3) &\approx 0.997. \end{aligned}$$

Log-Normal distribution

Let $X \sim \mathcal{N}(0, 1)$. The distribution of $Y = e^X$ is the **Log-Normal**.

Log-Normal distribution

Let $X \sim \mathcal{N}(0,1)$. The distribution of $Y = e^X$ is the **Log-Normal**.

Since $g(x) = e^x$ is strictly increasing, we can use the change of variables formula to find the PDF of Y. Let $y = e^x$, so $x = \log y$ and $dy/dx = e^x$. Then

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = \varphi(x) \frac{1}{e^x} = \varphi(\log y) \frac{1}{y}, \quad y > 0.$$

Log-Normal distribution

Let $X \sim \mathcal{N}(0,1)$. The distribution of $Y = e^X$ is the **Log-Normal**.

Since $g(x) = e^x$ is strictly increasing, we can use the change of variables formula to find the PDF of Y. Let $y = e^x$, so $x = \log y$ and $dy/dx = e^x$. Then

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \varphi(x)\frac{1}{e^x} = \varphi(\log y)\frac{1}{y}, \quad y > 0.$$

Alternatively, the CDF of Y is

$$F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P(X \leq \log y) = \Phi(\log y),$$

so the PDF is

$$f_Y(y) = \frac{d}{dy}\Phi(\log y) = \varphi(\log y)\frac{1}{y}, \quad y > 0.$$

Chi-Square distribution

Let $X \sim \mathcal{N}(0, 1)$, $Y = X^2$. The distribution of Y is an example of a **Chi-Square distribution**.

Chi-Square distribution

Let $X \sim \mathcal{N}(0,1)$, $Y = X^2$. The distribution of Y is an example of a **Chi-Square distribution**.

The event $X^2 \leq y$ is equivalent to the event $-\sqrt{y} \leq X \leq \sqrt{y}$, so the CDF of Y is

$$\begin{aligned} F_Y(y) &= P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \\ &= 2\Phi(\sqrt{y}) - 1, \end{aligned}$$

so

$$f_Y(y) = 2\varphi(\sqrt{y}) \cdot \frac{1}{2} y^{-1/2} = \varphi(\sqrt{y}) y^{-1/2}, \quad y > 0.$$

Cauchy distribution

Let X and Y be i.i.d. $\mathcal{N}(0,1)$, and let $T = X/Y$. The distribution of T is called the **Cauchy distribution**.

Cauchy distribution

Let X and Y be i.i.d. $\mathcal{N}(0,1)$, and let $T = X/Y$. The distribution of T is called the **Cauchy distribution**.

To find the CDF of T:

$$\begin{aligned}
F_T(t) &= P(T \leq t) \\
&= P(\frac{X}{Y} \leq t) \\
&= P(X \leq tY \mid Y > 0) + P(X \geq tY \mid Y < 0) \\
&= P(X \leq tY \mid Y > 0) + P(X \leq t(-Y) \mid Y < 0) \\
&= P(X \leq t|Y|) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{t|y|} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dx dy.
\end{aligned}$$

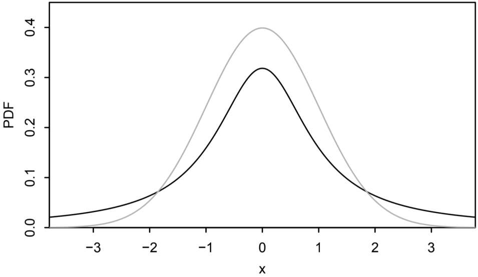
Cauchy distribution

$$
\begin{aligned}
F_T(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \left(\int_{-\infty}^{t|y|} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right) dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \Phi(t|y|) dy \\
&= \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-y^2/2} \Phi(ty) dy
\end{aligned}
$$

Differentiating the CDF with respect to t gives the PDF:

$$
\begin{aligned}
f_T(t) = F_T'(t) &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} \frac{\partial}{\partial t} \left(e^{-y^2/2} \Phi(ty) \right) dy \\
&= \sqrt{\frac{2}{\pi}} \int_0^{\infty} y e^{-y^2/2} \varphi(ty) dy \\
&= \frac{1}{\pi} \int_0^{\infty} y e^{-\frac{(1+t^2)y^2}{2}} dy \\
&= \frac{1}{\pi(1+t^2)}. \quad [u = (1+t^2)y^2/2, du = (1+t^2) y dy]
\end{aligned}
$$

Cauchy PDF



Cauchy PDF (dark) and $\mathcal{N}(0,1)$ PDF (light).

- The Cauchy distribution has much *heavier tails* than the Normal distribution.
- The expected value of a Cauchy random variable does *not* exist.
  - For large t, $\frac{t}{1+t^2} \approx \frac{1}{t}$, and $\int_1^\infty \frac{1}{t} dt = \infty$.

Exponential distribution

The **Exponential distribution** is the continuous counterpart to the Geometric distribution.

- Geometric random variable counts the *number of failures* before the first success in a sequence of Bernoulli trials.
- Exponential random variable represents the *waiting time* until the first arrival of a success.
  - Successes arrive at a rate of λ successes per unit of time.
  - The average # of successes in a time interval of length t is λt.

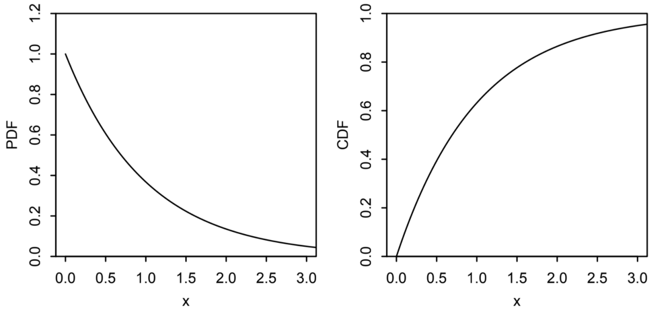
Exponential PDF and CDF

A continuous random variable X is said to have the **Exponential distribution** with parameter λ, $X \sim \mathrm{Expo}(\lambda)$, where $\lambda > 0$, if its PDF is

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

The corresponding CDF is

$$F(x) = 1 - e^{-\lambda x}, \quad x > 0.$$



Exponential mean and variance

Let $X \sim \mathrm{Expo}(1)$, then $f(x) = e^{-x}$, for $x > 0$.

$E(X)$ and $\mathrm{Var}(X)$ can be obtained using standard integration by parts:

$$
\begin{aligned}
E(X) &= \int_0^\infty x e^{-x} dx = 1, \quad [u = x, dv = e^{-x} dx] \\
E(X^2) &= \int_0^\infty x^2 e^{-x} dx = 2, \quad [u = x^2, dv = e^{-x} dx] \\
\mathrm{Var}(X) &= E(X^2) - (EX)^2 = 1.
\end{aligned}
$$

Exponential mean and variance

Let $X \sim \mathrm{Expo}(1)$, then $f(x) = e^{-x}$, for $x > 0$.

$E(X)$ and $\mathrm{Var}(X)$ can be obtained using standard integration by parts:

$$\begin{aligned}
E(X) &= \int_0^\infty x e^{-x} dx = 1, \quad [u = x, dv = e^{-x} dx] \\
E(X^2) &= \int_0^\infty x^2 e^{-x} dx = 2, \quad [u = x^2, dv = e^{-x} dx] \\
\mathrm{Var}(X) &= E(X^2) - (EX)^2 = 1.
\end{aligned}$$

The expected value and variance of $Y = X/\lambda \sim \mathrm{Expo}(\lambda)$:

$$E(Y) = \frac{1}{\lambda} E(X) = \frac{1}{\lambda}, \quad \mathrm{Var}(Y) = \frac{1}{\lambda^2} \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Memoryless property

A continuous distribution is said to have the **memoryless property** if a random variable X from that distribution satisfies

$$P(X \geq s + t \mid X \geq s) = P(X \geq t)$$

for all $s, t \geq 0$.

Exponential distribution has the memoryless property. Let $X \sim \mathrm{Expo}(\lambda)$, then

$$P(X \geq s+t \mid X \geq s) = \frac{P(X \geq s + t)}{P(X \geq s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X \geq t).$$

In fact, no other continuous distribution on $(0, \infty)$ is memoryless.

The importance of the Exponential

- The Exponential distribution is an important model in its own right, since some physical phenomena, such as radioactive decay, do exhibit the memoryless property.
- The Exponential distribution is well-connected to other named distributions, such as Geometric and Poisson.
- The Exponential distribution serves as a building block for more flexible distributions, such as Weibull.

Minimum of independent Exponentials

Let $X_1, \ldots, X_n$ be independent with $X_j \sim \mathrm{Expo}(\lambda_j)$. Let $L = \min(X_1, \ldots, X_n)$. What is the distribution of L?

Gamma distribution

The Gamma distribution is a continuous distribution on the positive real line, which generalizes the Exponential distribution.

A random variable Y is said to have the **Gamma distribution** with parameters a and λ, $Y \sim \mathrm{Gamma}(a, \lambda)$, where $a > 0$ and $\lambda > 0$, if its PDF is

$$f(y) = \frac{1}{\Gamma(a)}(\lambda y)^a e^{-\lambda y}\frac{1}{y}, \quad y > 0.$$

where Γ is the **gamma function**, defined by

$$\Gamma(a) = \int_0^\infty x^a e^{-x}\frac{dx}{x},$$

for real numbers $a > 0$.

Gamma function

Two important properties of the gamma function.

- $\Gamma(a+1) = a\Gamma(a)$ for all $a > 0$. This follows from integration by parts:

$$\Gamma(a+1) = \int_0^\infty x^a e^{-x} dx = -x^a e^{-x}\Big|_0^\infty + a\int_0^\infty x^{a-1} e^{-x} dx = a\Gamma(a)$$

- $\Gamma(n) = (n-1)!$ if n is a positive integer. This can be proved by induction, starting with $n = 1$ and using the recursive relation $\Gamma(a+1) = a\Gamma(a)$.

The $\mathrm{Gamma}(a, 1)$ distribution

Dividing both sides of the $\Gamma(a)$ definition by $\Gamma(a)$ gives

$$1 = \int_0^\infty \frac{1}{\Gamma(a)} x^a e^{-x} \frac{dx}{x},$$

where the function under the integral is a valid PDF supported on $(0, \infty)$. This is the PDF of the Gamma distribution with parameters a and 1, i.e., $X \sim \mathrm{Gamma}(a, 1)$ if its PDF is

$$f_X(x) = \frac{1}{\Gamma(a)} x^a e^{-x} \frac{1}{x}, \quad x > 0.$$

The $\mathrm{Gamma}(a, \lambda)$ distribution

From the $\mathrm{Gamma}(a, 1)$ distribution, we can obtain the general Gamma distribution by a scale transformation: if $X \sim \mathrm{Gamma}(a, 1)$ and $\lambda > 0$, then $Y = X/\lambda \sim \mathrm{Gamma}(a, \lambda)$.

By the change of variables formula with $x = \lambda y$ and $dx/dy = \lambda$, the PDF of Y is

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = \frac{1}{\Gamma(a)} (\lambda y)^a e^{-\lambda y} \frac{1}{\lambda y} \lambda = \frac{1}{\Gamma(a)} (\lambda y)^a e^{-\lambda y} \frac{1}{y}$$

for $y > 0$.

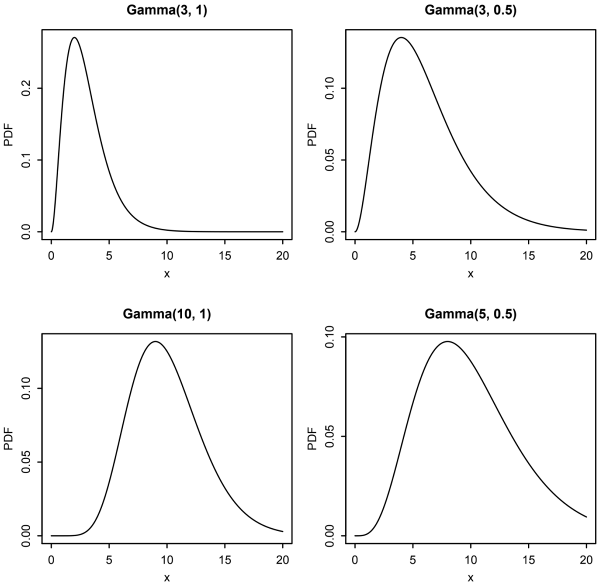
Parameters of the Gamma distribution

The PDF of the $\mathrm{Gamma}(a, \lambda)$ is

$$f(y) = \frac{1}{\Gamma(a)}(\lambda y)^a e^{-\lambda y}\frac{1}{y}, \quad y > 0.$$

- The $\mathrm{Gamma}(1, \lambda)$ PDF is $f(y) = \lambda e^{-\lambda y}$, so the $\mathrm{Gamma}(1, \lambda)$ and $\mathrm{Expo}(\lambda)$ distributions are the same.
- For small values of a, the PDF is skewed, but as a increases, the PDF starts to look more symmetrical and bell-shaped.
- Increasing λ compresses the PDF toward smaller values.

Gamma PDFs


Gamma(3, 1)
Gamma(3, 0.5)
Gamma(10, 1)
Gamma(5, 0.5)
PDF
x
0.0
0.1
0.2
0.00
0.05
0.10
0
5
10
15
20

Mean and variance of $\mathrm{Gamma}(a, \lambda)$

Mean of $X \sim \mathrm{Gamma}(a, 1)$:

$$E(X) = \int_0^\infty x \cdot \frac{1}{\Gamma(a)} x^a e^{-x} \frac{dx}{x} = \frac{1}{\Gamma(a)} \int_0^\infty x^{a+1} e^{-x} \frac{dx}{x}$$
$$= \frac{\Gamma(a+1)}{\Gamma(a)} = a.$$

Variance of $X \sim \mathrm{Gamma}(a, 1)$:

$$E(X^2) = \int_0^\infty \frac{1}{\Gamma(a)} x^{a+2} e^{-x} \frac{dx}{x} = \frac{\Gamma(a+2)}{\Gamma(a)} = (a+1)a,$$

$$\mathrm{Var}(X) = (a+1)a - a^2 = a.$$

For $Y = X/\lambda \sim \mathrm{Gamma}(a, \lambda)$,

$$E(Y) = \frac{1}{\lambda} E(X) = \frac{a}{\lambda}, \qquad \mathrm{Var}(Y) = \frac{1}{\lambda^2} \mathrm{Var}(Y) = \frac{a}{\lambda^2}.$$

Sum of Exponential RVs and sum of Gamma RVs

Let $X_1, \ldots, X_n$ be i.i.d. $\mathrm{Expo}(\lambda)$. What is the distribution of $X_1 + \cdots + X_n$?

Let $X_1, \ldots, X_n$ be independent with $X_j \sim \mathrm{Gamma}(a_j, \lambda)$. What is the distribution of $X_1 + \cdots + X_n$?

Beta distribution

A random variable X is said to have the **Beta distribution** with parameters a and b, where $a > 0$ and $b > 0$, if its PDF is

$$f(x) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1,$$

where the constant $\beta(a,b)$ is chosen to make the PDF integrate to 1. We write this as $X \sim \mathrm{Beta}(a,b)$.

The Beta distribution is a continuous distribution on the interval $(0,1)$. It is a generalization of the $\mathrm{Unif}(0,1)$ distribution, allowing the PDF to be non-constant on $(0,1)$

Beta integral

By definition, the constant $\beta(a, b)$ satisfies

$$\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx.$$

An integral of this form is called a **beta integral**.

The beta integral is related to the gamma function through the following identity:

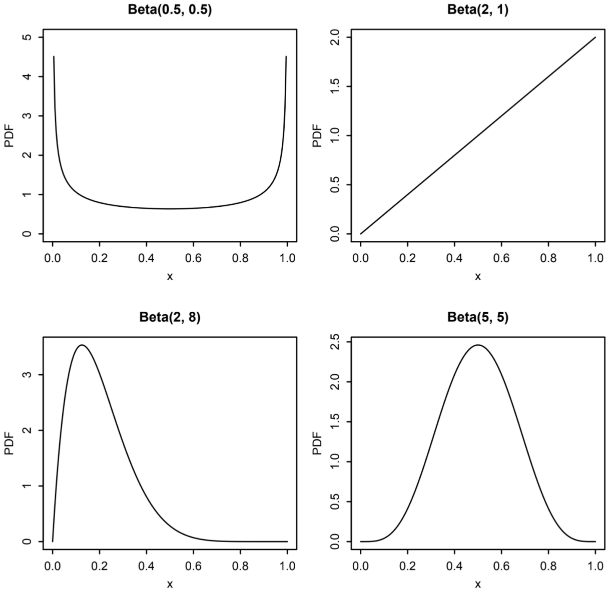
$$\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Beta distribution

The parameters a and b determine the shape of a Beta distribution.

- If $a = b = 1$, the $\mathrm{Beta}(1, 1)$ PDF is constant on $(0, 1)$, so the $\mathrm{Beta}(1, 1)$ and $\mathrm{Unif}(0, 1)$ distributions are the same.
- If $a = b$, the PDF is symmetric about $1/2$.
- If $a > b$, the PDF favors values larger than $1/2$; if $a < b$, the PDF favors values smaller than $1/2$.
- If $a < 1$ and $b < 1$, the PDF is U-shaped and opens upward.
- If $a > 1$ and $b > 1$, the PDF opens down.

Beta PDFs


Beta(0.5, 0.5)
Beta(2, 1)
Beta(2, 8)
Beta(5, 5)
PDF
x
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
0.5
1.5
2.0
2.5