



STAT 131 - Intro to Probability Theory

Lecture 7: Covariance, correlation, and transformations

Suzana Şerboi, Ph.D.
Mathematics and Statistics, UCSC

Covariance

Just as the mean and variance provided single-number summaries of the distribution of a single r.v., covariance is a single-number summary of the joint distribution of two r.v.s.

The **covariance** between r.v.s X and Y is:

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)].$$

If the differences $(X - E(X))$ and $(Y - E(Y))$ have different signs, then the covariance is negative, and if they have the same signs, then the covariance is positive.

We can think of the covariance as measure of the tendency of two r.v.s to go up or down together, relative to their means:

- positive covariance between X and Y indicates that when X goes up, Y also tends to go up, and
- negative covariance indicates that when X goes up, Y tends to go down.

Multiplying $(X - EX)(Y - EY)$ out and using linearity of expectation,

$$\begin{aligned}\text{Cov}(X, Y) &= E[(X - EX)(Y - EY)] \\ &= E[X \cdot Y - X \cdot EY - EX \cdot Y - EX \cdot EY] \\ &= E(X \cdot Y) - E(X \cdot EY) - E(EX \cdot Y) + E(EX \cdot EY) \\ &= E(XY) - EY \cdot EX - EX \cdot EY + EX \cdot EY \\ &= E(XY) - EX \cdot EY.\end{aligned}$$

Thus we have an equivalent expression:

The covariance between r.v.s X and Y is:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

If X and Y are independent, then they are **uncorrelated** (i.e. have zero covariance).

Proof. We'll show this in the case where X and Y are continuous, with PDFs f_X and f_Y . The proof in the discrete case is the same, with PMFs instead of PDFs and sums instead of integrals.

Since X and Y are independent, their joint PDF is the product of the marginal PDFs. That is, $f_{X,Y}(x, y) = f_X(x) f_Y(y)$.

By 2D LOTUS,

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} y f_Y(y) \left(\int_{-\infty}^{\infty} x f_X(x) dx \right) dy \\ &= \int_{-\infty}^{\infty} x f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= E(X)E(Y) \end{aligned}$$

Thus $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0$.

Zero covariance doesn't imply independence

The converse of this theorem is false: **just because X and Y are uncorrelated does not mean they are independent.**

For example, let $X \sim \mathcal{N}(0, 1)$, and let $Y = X^2$.

Then $E(XY) = E(X^3) = 0$ because odd moments of a symmetric distribution are 0.

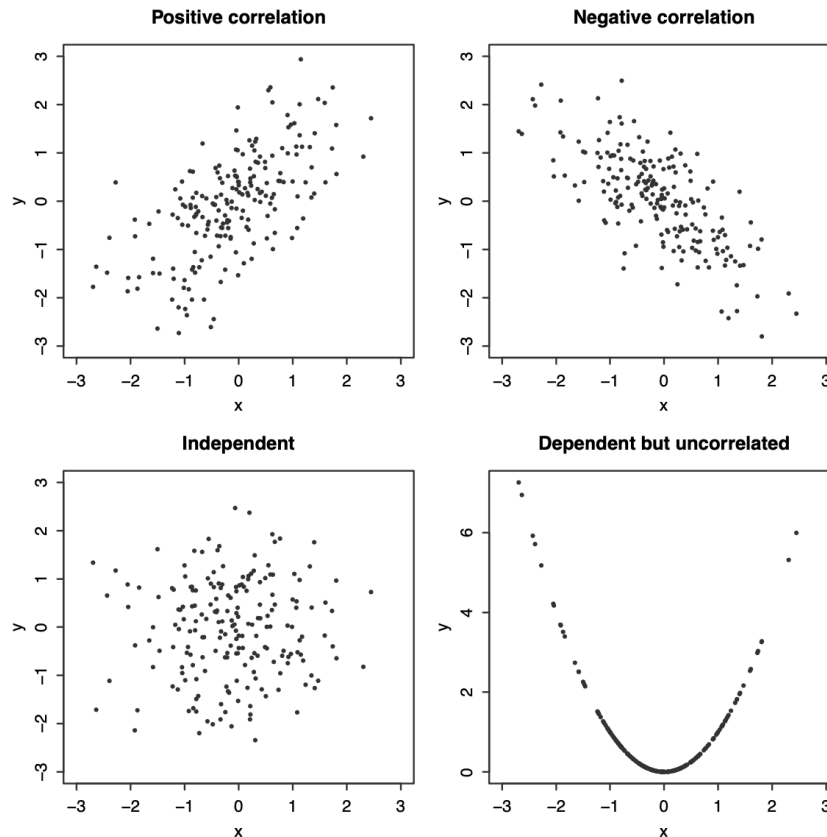
Thus X and Y are uncorrelated,

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0 - 0 = 0.$$

But they are certainly not independent: Y is a function of X , so knowing X gives us perfect information about Y (one is a function of the other.)

Covariance is a measure of linear association

Random variables can be dependent in nonlinear ways and still have zero covariance, as this example demonstrates. Bottom left: X and Y are independent, hence uncorrelated. Bottom right: Y is a deterministic function of X , but X and Y are uncorrelated.



The covariance has the following properties:

$$(1) \text{Cov}(X, X) = \text{Var}(X)$$

$$(2) \text{Cov}(X, Y) = \text{Cov}(Y, X)$$

$$(3) \text{Cov}(X, c) = 0 \text{ for any constant } c.$$

$$(4) \text{Cov}(aX, Y) = a \text{Cov}(X, Y) \text{ for any constant } a.$$

$$(5) \text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$$

$$(6) \text{Cov}(X + c, Y) = \text{Cov}(X, Y)$$

(7) More generally,

$$\text{Cov}\left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j\right) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j \text{Cov}(X_i, Y_j).$$

(8)

$$\text{Cov}(X + Y, Z + W) = \text{Cov}(X, Z) + \text{Cov}(X, W) + \text{Cov}(Y, Z) + \text{Cov}(Y, W)$$

$$(9) \text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2 \text{Cov}(X, Y)$$

In the case when X and Y are independent r.v.s we have $\text{Cov}(X, Y) = 0$ and thus $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$.

Correlation

Since covariance is influenced by the units of measurement for X and Y — if X is measured in centimeters instead of meters, the covariance increases by a factor of 100 - a more interpretable, unitless measure called correlation is often preferred.

The **correlation** between r.v.s X and Y is:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\text{SD}(X) \cdot \text{SD}(Y)}.$$

Recall $\text{SD}(X)$ denotes the standard deviation of X .

Note that shifting and scaling r.v. has no effect on their correlation:

$$\text{Corr}(cX, Y) = \frac{\text{Cov}(cX, Y)}{\sqrt{c^2 \text{Var}(X) \text{Var}(Y)}} = \text{Corr}(X, Y).$$

Correlation bounds

Correlation is easy to interpret because it is independent of the units of measurement and always falls within the range of -1 to 1 .

For any r.v.s X and Y ,

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

Proof. Without loss of generality we can assume X and Y have variance 1, since scaling does not change the correlation.

Let $\rho = \text{Corr}(X, Y) = \text{Cov}(X, Y)$.

Using the fact that variance is nonnegative, along with Property 9 of covariance, we have

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y) = 2 + 2\rho \geq 0 \Rightarrow \rho \geq -1$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2 \text{Cov}(X, Y) = 2 - 2\rho \geq 0 \Rightarrow \rho \leq 1$$

Thus, $-1 \leq \rho \leq 1$.

Example. Rolling two dice, let X be the number shown on the first die, Y the one shown on the second die, $Z = X + Y$ the sum of the two numbers.

Clearly, X and Y are independent, $\text{Cov}(X, Y) = 0$, $\text{Corr}(X, Y) = 0$.

For X and Z ,

$$\begin{aligned}\text{Cov}(X, Z) &= \text{Cov}(X, X + Y) = \text{Cov}(X, X) + \text{Cov}(X, Y) \\ &= \text{Var } X + 0 = \text{Var } X\end{aligned}$$

$$\begin{aligned}\text{Var}(Z) &= \text{Var}(X + Y) = \text{Var } X + \text{Var } Y \quad \text{indep.}! \\ &= \text{Var } X + \text{Var } X = 2 \text{Var } X,\end{aligned}$$

since $\text{Var}(X) = \text{Var}(Y)$ (as X and Y have the same distribution).

$$\begin{aligned}\text{Corr}(X, Z) &= \frac{\text{Cov}(X, Z)}{\text{SD}(X) \cdot \text{SD}(Z)} \\ &= \frac{\text{Var } X}{\sqrt{\text{Var } X} \cdot \sqrt{2 \text{Var } X}} = \frac{1}{\sqrt{2}}.\end{aligned}$$

Multivariate Normal distribution

The Multivariate Normal is a continuous distribution that extends the Normal distribution to multiple dimensions. Instead of dealing with the complex joint PDF of the Multivariate Normal, we define it based on its connection to the standard Normal distribution.

A k -dimensional random vector $\mathbf{X} = (X_1, \dots, X_n)$ is said to have a **Multivariate Normal (MVN) distribution** if every linear combination of the X_j has a Normal distribution. That is, we require:

$$t_1 X_1 + \dots + t_k X_k$$

to have a Normal distribution for any constants t_1, \dots, t_k . An important special case is $k = 2$, this distribution is called the **Bivariate Normal (BVN)**.

- If (X_1, \dots, X_n) is MVN, then the marginal distribution of X_1 is Normal, since we can take t_1 to be 1 and all other 0.
- It is possible to have Normally distributed r.v.s X_1, \dots, X_n , such that (X_1, \dots, X_n) is not Multivariate Normal.

Parameters of an MVN random vector

A Multivariate Normal distribution is fully specified by knowing the mean of each component, the variance of each component, and the covariance or correlation between any two components. Another way to say this is that the parameters of an MVN random vector (X_1, \dots, X_k) are as follows:

- the mean vector (μ_1, \dots, μ_k) , where $E(X_j) = \mu_j$
- the covariance matrix, which is the $k \times k$ matrix of covariances between components, arranged so that the row i , column j entry is $\text{Cov}(X_i, X_j)$.

For example, in order to fully specify a Bivariate Normal distribution for (X, Y) , we need to know five parameters:

- the means $E(X) = \mu_X, E(Y) = \mu_Y$
- the variances $\text{Var}(X) = \sigma_X^2, \text{Var}(Y) = \sigma_Y^2$
- the correlation $\text{Corr}(X, Y) = \rho$

To find the Bivariate Normal joint PDF, we first need to talk about transformations of random variables.

Transformations

After applying a function to a random variable X or random vector \mathbf{X} , we would like to find the distribution of the transformed random variable or joint distribution of the transformed random vector.

Transformations of random variables appear all over the place in statistics. Kinds of transformations we'll be looking at: unit conversion, sums, averages, extreme values.

It is especially important to us to understand transformations because of the approach we've taken to learning about the named distributions. Starting from a few basic distributions, we have defined other distributions as transformations of these elementary building blocks, in order to understand how the named distributions are related to one another.

If we are only looking for the expectation of $g(X)$, LOTUS tells us that the PMF or PDF of X is enough for calculating $E(g(X))$. LOTUS also applies to functions of several r.v.s, as we learned in the previous chapter. If we need the full distribution of $g(X)$, not just its expectation, our approach depends on whether X is discrete or continuous.

Discrete Case

In the discrete case, we get the PMF of $g(X)$ by translating the event $g(X) = y$ into an equivalent event involving X .

To do so, we look for all values x such that $g(x) = y$; as long as X equals any of these x 's, the event $g(X) = y$ will occur. This gives the formula

$$P(g(X) = y) = \sum_{x:g(x)=y} P(X = x)$$

For a **one-to-one** g , the situation is particularly simple, because there is only one value of x such that $g(x) = y$, namely $g^{-1}(y)$. Then we can use

$$P(g(X) = y) = P(X = g^{-1}(y))$$

to convert between the PMFs of X and $g(X)$.

For example, it is extremely easy to convert between the Geometric and First Success distributions.

Continuous case

In the continuous case, a universal approach is to start from the CDF of $g(X)$, and translate the event $g(X) \leq y$ into an equivalent event involving X .

For general g , we may have to think carefully about how to express $g(X) \leq y$ in terms of X , and there is no easy formula we can plug into.

But when **g is continuous and strictly increasing**, the translation is easy: $g(X) \leq y$ is the same as $X \leq g^{-1}(y)$, so

$$F_{g(X)}(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

We can then differentiate with respect to y to get the PDF of $g(X)$.

This gives a one-dimensional version of the change of variables formula, which generalizes to invertible transformations in multiple dimensions.

Change of variables in one dimension

Let X be a continuous r.v. with PDF f_X , and let $Y = g(X)$, where g is a differentiable and strictly increasing (or strictly decreasing) function. Then the PDF of Y is given by:

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|,$$

where $x = g^{-1}(y)$. The support of Y is all $g(x)$ with x in the support of X .

Proof. Let g be strictly increasing. The CDF of Y is $F_Y(y) = P(Y \leq y)$

$$= P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) = F_X(x)$$

so by the chain rule, the PDF of Y is $f_Y(y) = f_X(x) \frac{dx}{dy}$.

The proof for g strictly decreasing is analogous. In that case the PDF ends up as $-f_X(x) \frac{dx}{dy}$, which is nonnegative since $\frac{dx}{dy} < 0$ if g is strictly decreasing.

Using the absolute value $\left| \frac{dx}{dy} \right|$ covers both cases.

Example: Let X be a continuous random variable with PDF $f_X(x) = 4x^3$ for $0 < x \leq 1$ and 0 otherwise. Now let $Y = 1/X$. Find $f_Y(y)$.

First note that the support of Y is $[1, \infty)$.

Also, note that $g(x) = \frac{1}{x}$ is a strictly decreasing and differentiable function on $(0, 1]$.

We have $g'(x) = -1/x^2$.

For any $y \in [1, \infty)$, $x = g^{-1}(y) = 1/y$ and $\frac{dx}{dy} = -\frac{1}{y^2}$. Then

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

$$f_Y(y) = 4(1/y)^3 \left| -1/y^2 \right| = 4/y^5$$

NOTE. When finding the distribution of Y , be sure to:

- Specify the support of Y .
- Check the assumptions of the change of variables theorem carefully if you wish to apply it.
- Express your final answer for the PDF of Y as a function of y .

Example (Log-Normal PDF). Let $X \sim \mathcal{N}(0, 1)$, $Y = e^X$. In Chapter 6 we named the distribution of Y the Log-Normal. We can use the change of variables formula to find the PDF of Y , since $g(x) = e^x$ is strictly increasing.

To determine the support of Y , we just observe that as x ranges from $-\infty$ to ∞ , e^x ranges from 0 to ∞ .

Let $y = e^x$, so $x = \log y$ and $dy/dx = e^x$. Then

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

We can get the same result by working from the definition of the CDF, translating the event $Y \leq y$ into an equivalent event involving X . For $y > 0$,

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

so the PDF is again

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

Example (Chi-Square PDF). Let $X \sim \mathcal{N}(0, 1)$, $Y = X^2$.

The distribution of Y is an example of a Chi-Square distribution.

To find the PDF of Y , we can no longer apply the change of variables formula because $g(x) = x^2$ is not strictly increasing (or decreasing). Instead we start from the CDF.

Note the event $X^2 \leq y$ is equivalent to the event $-\sqrt{y} \leq X \leq \sqrt{y}$.

Then

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

Thus

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

n -dimensional Change of variables

The change of variables formula generalizes to n dimensions, where it tells us how to use the joint PDF of a random vector \mathbf{X} to get the joint PDF of the transformed random vector $\mathbf{Y} = g(\mathbf{X})$.

The formula is analogous to the one-dimensional version, but it involves a multivariate generalization of the derivative called a Jacobian matrix.

See sections A.6 and A.7 of the math appendix for more about Jacobians.

NOTE. This is only for continuous r.v.s. For discrete r.v.s we can transform using the PMF directly. For example, let X be a positive r.v. and $Y = X^3$, then: $P(Y = y) = P(X = y^{1/3})$.

Change of variables. Let $\mathbf{X} = (X_1, \dots, X_n)$ be a continuous random vector with joint PDF $f_{\mathbf{X}}$. Let $g : A_0 \rightarrow B_0$ be an invertible function, where A_0 and B_0 are open subsets of \mathbb{R}^n , A_0 contains the support of \mathbf{X} , and B_0 is the range of g .

Let $\mathbf{Y} = g(\mathbf{X})$, and mirror this by letting $\mathbf{y} = g(\mathbf{x})$. Since g is invertible, we also have $\mathbf{X} = g^{-1}(\mathbf{Y})$ and $\mathbf{x} = g^{-1}(\mathbf{y})$. Suppose that all the partial derivatives $\frac{\partial x_i}{\partial y_j}$ exist and are continuous, so we can form the Jacobian matrix

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}$$

Also assume that the determinant of this Jacobian matrix is never 0. Then the joint PDF of \mathbf{Y} is

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) \cdot \left| \det \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right|, \quad \text{for } \mathbf{y} \in B_0.$$

Example (Bivariate Normal joint PDF). Let (Z, W) be BVN with $\mathcal{N}(0, 1)$ marginals and $\text{Corr}(Z, W) = \rho$. Assume that $-1 < \rho < 1$ since otherwise the distribution is degenerate (with Z and W perfectly correlated).

We can construct (Z, W) as

$$\begin{aligned} Z &= X \\ W &= \rho X + \tau Y \end{aligned}$$

with $\tau = \sqrt{1 - \rho^2}$ and X, Y i.i.d. $\mathcal{N}(0, 1)$.

We also need the inverse transformation. Solving $Z = X$ for X , we have $X = Z$. Plugging this into $W = \rho X + \tau Y$ and solving for Y , we have $X = Z$ and $Y = -\frac{\rho}{\tau}Z + \frac{1}{\tau}W$.

The Jacobian is

$$\frac{\partial(x, y)}{\partial(z, w)} = \begin{pmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{\rho}{\tau} & \frac{1}{\tau} \end{pmatrix}$$

which has determinant $1/\tau$.

So by the change of variables formula,

$$\begin{aligned} f_{Z,W}(z, w) &= f_{X,Y}(x, y) \cdot \left| \det \frac{\partial(x, y)}{\partial(z, w)} \right|, \\ &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \cdot \frac{1}{\tau} \\ &= \frac{1}{2\pi\tau} \exp\left(-\frac{1}{2}\left(z^2 + \left(-\frac{\rho}{\tau}z + \frac{1}{\tau}w\right)^2\right)\right) \\ &= \frac{1}{2\pi\tau} \exp\left(-\frac{1}{2\tau^2}(z^2 + w^2 - 2\rho zw)\right), \text{ for all real } z, w \end{aligned}$$

In the last step we multiplied things out and used the fact that $\rho^2 + \tau^2 = 1$.

Convolutions

A **convolution** is a sum of independent random variables. The objective in this case is to find the distribution of $T = X + Y$, where X and Y are independent r.v.s whose distributions are known.

We can solve these problems using MGFs, but what happens if a distribution we are working with doesn't have a defined MGF?

Let X and Y be independent r.v.s and $T = X + Y$ be their sum.

If X and Y are discrete, then the PMF of T is:

$$P(T = t) = \sum_x P(Y = t - x)P(X = x) = \sum_y P(X = t - y)P(Y = y).$$

If X and Y are continuous, then the PDF of T is:

$$f_T(t) = \int_{-\infty}^{\infty} f_Y(t - x)f_X(x)dx = \int_{-\infty}^{\infty} f_X(t - y)f_Y(y)dy.$$

NOTE that in both cases we are using LOTP.

Example (Poisson convolution) Let $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ be independent. Then

$$\begin{aligned} \mathbb{P}_{X+Y}(k) &= \sum_{i=-\infty}^{\infty} \mathbb{P}_X(k-i) \cdot \mathbb{P}_Y(i) = \sum_{i=0}^k e^{-\lambda} \frac{\lambda^{k-i}}{(k-i)!} \cdot e^{-\mu} \frac{\mu^i}{i!} \\ &= e^{-(\lambda+\mu)} \frac{1}{k!} \sum_{i=0}^k \binom{k}{i} \cdot \lambda^{k-i} \cdot \mu^i = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^k}{k!} \end{aligned}$$

from where we conclude $X + Y \sim \text{Pois}(\lambda + \mu)$.

Example (Binomial convolution) Let $X \sim \text{Binom}(n, p)$ and $Y \sim \text{Binom}(m, p)$ be independent. Then $X + Y \sim \text{Binom}(n + m, p)$:

$$\begin{aligned} p_{X+Y}(k) &= \sum_{i=0}^k \binom{n}{k-i} p^{k-i} (1-p)^{n-k+i} \cdot \binom{m}{i} p^i (1-p)^{m-i} \\ &= p^k \cdot (1-p)^{m+n-k} \underbrace{\sum_{i=0}^k \binom{n}{k-i} \cdot \binom{m}{i}}_{\binom{n+m}{k}} = \binom{n+m}{k} p^k \cdot (1-p)^{m+n-k} \end{aligned}$$

Example (Exponential convolution)

Let $X, Y \stackrel{\text{i.i.d.}}{\sim} \text{Expo}(\lambda)$. Find the distribution of $T = X + Y$.

This is known as the $\text{Gamma}(2, \lambda)$ distribution. We will introduce the Gamma distribution later.

Solution: For $t > 0$, the convolution formula gives

$$f_T(t) = \int_{-\infty}^{\infty} f_Y(t-x)f_X(x)dx = \int_0^t \lambda e^{-\lambda(t-x)} \lambda e^{-\lambda x} dx$$

where we restricted the integral to be from 0 to t since we need $t - x > 0$ and $x > 0$ for the PDFs inside the integral to be nonzero.

Simplifying, we have

$$f_T(t) = \lambda^2 \int_0^t e^{-\lambda t} dx = \lambda^2 e^{-\lambda t} \int_0^t 1 dx = \lambda^2 e^{-\lambda t} x \Big|_0^t = \lambda^2 e^{-\lambda t} t, \text{ for } t > 0$$

Beta Distribution

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

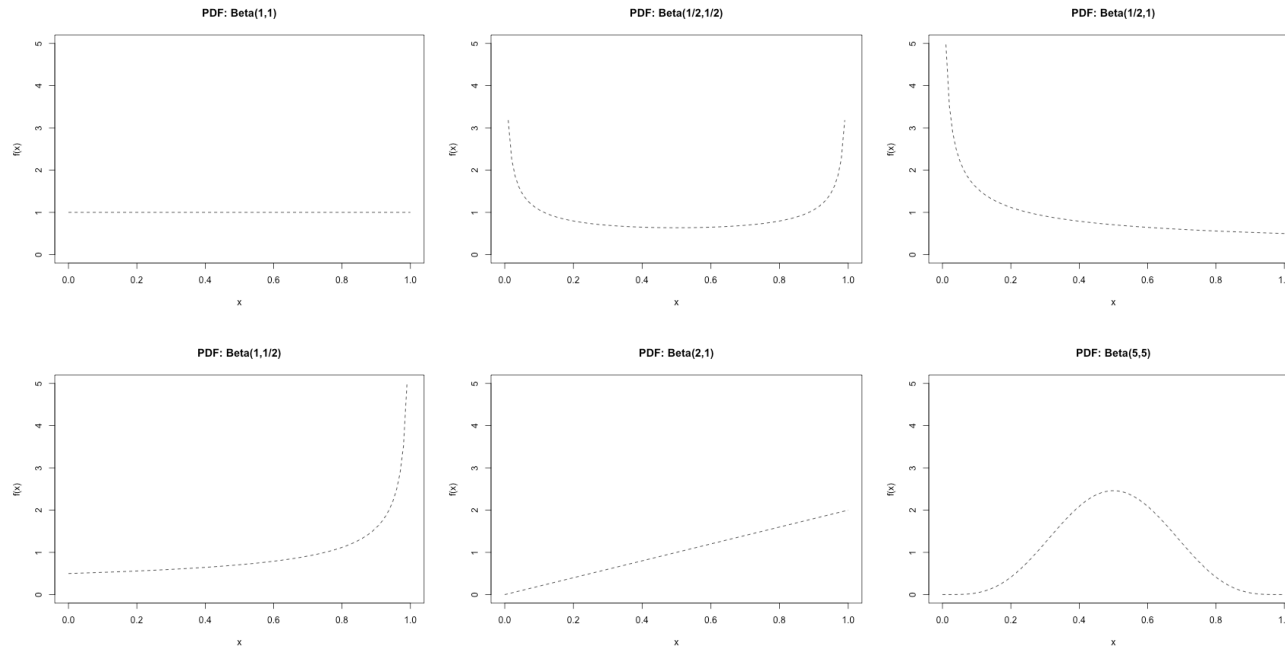
An r.v. X is said to have the **Beta Distribution** with parameters a and b , where $a > 0$, and $b > 0$, if its PDF is:

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

for $0 < x < 1$, where the constant $\beta(a, b)$ is chosen to make the PDF integrate 1. We write this as $X \sim \text{Beta}(a, b)$.

By definition then: $\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$ (beta integral)

[Beta Wiki](#)



- Taking $a = b = 1$, the Beta(1,1) PDF is constant on $(0, 1)$, so the Beta(1,1) and Unif(0,1) distributions are the same.
- If $a < 1$ and $b < 1$, the PDF is U-shaped and opens upward. If $a > 1$ and $b > 1$, the PDF opens down.
- 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 $1/2$.

Gamma Distribution

While the Exponential r.v. represents the waiting time for the first success under conditions of memorylessness, a Gamma r.v. represents the total waiting time for multiple successes.

An r.v. Y is said to have the **Gamma Distribution** with parameters a and λ , where $a > 0$, and $\lambda > 0$, if its PDF is:

$$f(y) = \frac{\lambda^a}{\Gamma(a)} y^{a-1} e^{-\lambda y}$$

for $y > 0$. We write this as $X \sim \text{Gamma}(a, \lambda)$.

Note that if $a = 1$ we have an *Expo*(λ).

[Gamma Dist Wiki](#)

Gamma Function

What is $\Gamma(a)$? Is the Gamma function != Gamma Distribution.

The Gamma function Γ is defined by:

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

for real numbers $a > 0$.

Important properties:

- $\Gamma(a + 1) = a\Gamma(a)$ for $a > 0$.
- $\Gamma(n) = (n - 1)!$ if n is a positive integer
- $\Gamma(1/2) = \sqrt{\pi}$
- $\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$
- $\int_0^{\infty} x^{a-1} e^{-\lambda x} dx = \frac{\Gamma(a)}{\lambda^a}$ for $\lambda > 0$.

Additional Practice Problems

1. Let X and Y be i.i.d. $Unif(0, 1)$. Compute the covariance of $X + Y$ and $X - Y$.
2. Let $X \sim \text{Exp}(\lambda)$, and $Y \sim \text{Gamma}(2, \lambda)$ be independent. Find the PDF of $X + Y$, $f_{X+Y}(t)$. How is $X + Y$ distributed?
2. Use a convolution integral to show that if $X \sim \mathcal{N}(\mu_1, \sigma^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma^2)$ are independent, then $T = X + Y \sim \mathcal{N}(\mu_1 + \mu_2, 2\sigma^2)$ (to simplify the calculation, we are assuming that the variances are equal). You can use a standardization (location-scale) idea to reduce to the standard Normal case before setting up the integral.

Hint: Complete the square.