Introduction to Probability
Joint distribution of random variables

Random vector

If X and Y are two discrete RVs on the same sample space Ω, i.e.

$$X : \Omega \to \mathbb{R}, \quad Y : \Omega \to \mathbb{R},$$

then the ordered pair (X, Y) is a function $(X, Y) : \Omega \to \mathbb{R}^2$, defined by

$$(X, Y)(\omega) = (X(\omega), Y(\omega)).$$

In general, we will call an ordered n-tuple of n random variables,

$$(X_1, X_2, ..., X_n) : \Omega \to \mathbb{R}^n,$$

a **random vector**.

Joint distributions of random variables

If $X_1, X_2, ..., X_n$ are all discrete RVs, then the **joint probability mass function** (joint PMF) of the discrete random vector $(X_1, X_2, ..., X_n)$ is defined by

$$p_{X_1,X_2,...,X_n}(k_1, k_2, ..., k_n) = P(X_1 = k_1, X_2 = k_2, ..., X_n = k_n).$$

First, note that joint PMF probabilities are, in fact, probabilities:

$$0 \leq p_{X_1,X_2,...,X_n}(k_1, k_2, ..., k_n) \leq 1$$

$$\sum \cdots \sum_{k_1,k_2,...,k_n} p_{X_1,X_2,...,X_n}(k_1, k_2, ..., k_n) = 1.$$

Expectation of a function of a random vector

If $g : \mathbb{R}^n \to \mathbb{R}$, then the **expectation** of the discrete function $g(X_1, X_2, ..., X_n)$ of the random vector with joint PMF p is

$$E(g(X_1, X_2, ..., X_n)) = \sum \cdots \sum_{k_1, k_2, ..., k_n} g(k_1, k_2, ..., k_n) p(k_1, k_2, ..., k_n)$$

if this sum is well defined.

Marginal distributions of random variables

From the joint PMF, we can recover each RV's individual PMF, called its **marginal PMF**, by summing over all possible values of the other RV.

For each $i = 1, 2, ..., n$, and fixed x, the marginal PMF of X_i is

$$p_{X_i}(x) = \sum \cdots \sum_{\substack{k_1, k_2, \ldots, k_{i-1}, \\ k_{i+1}, \ldots, k_n}} p(k_1, k_2, ..., k_{i-1}, x, k_{i+1}, ..., k_n).$$

The joint PMF of the first $m < n$ random variables is found using the same marginal summing technique:

$$p_{X_1, X_2, \ldots, X_m}(x_1, x_2, ..., x_m) = \sum \cdots \sum_{k_{m+1}, \ldots, k_n} p(x_1, x_2, ..., x_m, k_{m+1}, ..., k_n).$$

Joint distributions of random variables: example

An urn contains two red, one yellow, and three white marbles. Draw three marbles without replacement (all at once).

What is the probability you draw x red and y white marbles?

To answer this question, we'll build a table of the joint PMF $p_{X,Y}(x, y)$, where

- $X =$ number of red drawn
- $Y =$ number of white drawn.

Note that we can also say $Z =$ number of yellow drawn, but this value is *determined by X and Y*:

$$Z = 3 - X - Y.$$

Joint distributions of random variables: example

2 red $\implies X(\omega) = \{0, 1, 2\}$. 3 white $\implies Y(\omega) = \{0, 1, 2, 3\}$.

Certainly, $0 \leq X + Y \leq 3$.

There are two cases: $X + Y = 2$ if you draw the yellow marble, and $X + Y = 3$ if you don't draw the yellow marble.

We'll break these two cases down in the table below, noting that there are $|\Omega| = \binom{2+3+1}{3} = \binom{6}{3} = 20$ different draws possible.

	Y=0	Y=1	Y=2	Y=3	$p_X(x)$
X=0	0	0	$\frac{\binom{3}{2}\binom{1}{1}}{20} = \frac{3}{20}$	$\frac{\binom{3}{3}}{20} = \frac{1}{20}$	
X=1	0	$\frac{\binom{2}{1}\binom{3}{1}}{20} = \frac{6}{20}$	$\frac{\binom{2}{1}\binom{3}{2}}{20} = \frac{6}{20}$	0	
X=2	$\frac{\binom{2}{2}\binom{1}{1}}{20} = \frac{1}{20}$	$\frac{\binom{2}{2}\binom{3}{1}}{20} = \frac{3}{20}$	0	0	
$p_Y(y)$					

Joint distributions of random variables: example

2 red $\implies X(\omega) = \{0, 1, 2\}$. 3 white $\implies Y(\omega) = \{0, 1, 2, 3\}$.

Certainly, $0 \leq X + Y \leq 3$.

There are two cases: $X + Y = 2$ if you draw the yellow marble, and $X + Y = 3$ if you don't draw the yellow marble.

We'll break these two cases down in the table below, noting that there are $|\Omega| = \binom{2+3+1}{3} = \binom{6}{3} = 20$ different draws possible.

	Y=0	Y=1	Y=2	Y=3	$p_X(x)$
X=0	0	0	$\frac{\binom{3}{2}\binom{1}{1}}{20} = \frac{3}{20}$	$\frac{\binom{3}{3}}{20} = \frac{1}{20}$	$\frac{4}{20}$
X=1	0	$\frac{\binom{2}{1}\binom{3}{1}}{20} = \frac{6}{20}$	$\frac{\binom{2}{1}\binom{3}{2}}{20} = \frac{6}{20}$	0	$\frac{12}{20}$
X=2	$\frac{\binom{2}{2}\binom{1}{1}}{20} = \frac{1}{20}$	$\frac{\binom{2}{2}\binom{3}{1}}{20} = \frac{3}{20}$	0	0	$\frac{4}{20}$
$p_Y(y)$	$\frac{1}{20}$	$\frac{9}{20}$	$\frac{9}{20}$	$\frac{1}{20}$	1

Multinomial distribution

The random vector $(X_1, X_2, ..., X_r)$ is said to have the **multinomial distribution** with parameters $(n, r, p_1, p_2, ..., p_r)$ if the joint PMF p of the vector is

$$p(k_1, k_2, ..., k_r) = \binom{n}{k_1, k_2, ..., k_r} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r},$$

when $k_i \geq 0$ for $i = 1, 2, ..., r$ and $k_1 + k_2 + ... + k_r = n$.

Denote such a random vector by

$$(X_1, X_2, ..., X_r) \sim Mult(n, r, p_1, p_2, ..., p_r).$$

Jointly continuous random vector

n random variables $X_1, X_2, ..., X_n$ are called **jointly continuous** if there exists a **joint density function (joint PDF)** $f : \mathbb{R}^n \to \mathbb{R}$ such that*, for $B \subseteq \mathbb{R}^n$,

$$P((X_1, X_2, ..., X_n) \in B) = \int \cdots \int_B f(x_1, x_2, ..., x_n) dx_1 ... dx_n$$

such that

$$f(x_1, x_2, ..., x_n) \geq 0 \text{ and } \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, ..., x_n) dx_1 ... dx_n = 1.$$

*Again, Borel sets B, not *any* subset.

Expectation, marginal densities

Expectation of this random vector works as you might expect: if $g : \mathbb{R}^n \to \mathbb{R}$ and the integral are well defined, then $E(g(X_1, X_2, ..., X_n)) =$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, ..., x_n) f(x_1, x_2, ..., x_n) dx_1 ... dx_n.$$

The **marginal density function** of X_j is found in a similar fashion, by integrating the joint PDF along all variables except X_j: for fixed x, the marginal PDF of X_j at x is denoted $f_{X_j}(x)$ and equals

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, x_{j-1}, x, x_{j+1}, ..., x_n) dx_1 ... dx_{j-1} dx_{j+1} ... dx_n,$$

where the integral is over $n - 1$ variables.

Expectation, marginal densities

Also as in the discrete case, a joint density of $k < n$ of the variables can be found by integrating over the other $n-k$ variables:

$$f_{(X_1,X_2,\dots,X_k)}(x_1,x_2,\dots,x_k)$$
$$= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1,x_2,\dots,x_{k-1},x_k,y_{k+1},\dots,y_n)dy_{k+1}\dots dy_n.$$

Uniform distribution on a subset of $\mathbb{R}^2$

The **uniform distribution** over a finite-area subset $D \subseteq \mathbb{R}^2$ can have its probabilities measured by integrating over any event subset and dividing by the area of D.

Example

Let $(X, Y) \sim \mathit{Unif}(D)$, where

$$D = \{(x, y) \in \mathbb{R}^2 : \ 0 \leq x \leq 2, \ 1 \leq y \leq 5\}.$$

What is $P(X \geq Y^2)$?

Note that the area of D is 8. Thus, the joint PDF of (X, Y) is

$$f(x, y) = \frac{1}{8} 1_D(x, y).$$

Uniform distribution on a subset of $\mathbb{R}^2$

To calculate $P(X \geq Y^2)$, we need to craft the variables x and y to be able to integrate over the event

$$E = \{(x, y) \in D : x \geq y^2\} = \{(x, y) \in D : \sqrt{x} \leq y\}.$$

That integral can be done in two ways: integrating over x first, or over y first. We must be careful to only integrate over the portion of D where the inequality holds.

$$P(E) = \int_1^{\sqrt{2}} \int_{y^2}^{2} \frac{1}{8} dx\, dy = \frac{4\sqrt{2} - 5}{24} \approx 0.0273689$$

or

$$P(E) = \int_1^{2} \int_1^{\sqrt{x}} \frac{1}{8} dy\, dx = \frac{4\sqrt{2} - 5}{24} \approx 0.0273689.$$

Non-uniform distribution on a subset of $\mathbb{R}^2$

A non-uniform distribution on a subset of $\mathbb{R}^n$ is handled similarly, with the joint pdf.

Example

Let the random pair (X, Y) have joint PDF

$$f(x, y) = cxe^{-2xy}1_{(0,10)}(x)1_{(0,\infty)}(y).$$

1. What is the normalizing constant c that makes f a joint pdf?
2. What is $P(X \leq 6)$?

Non-uniform distribution on a subset of $\mathbb{R}^2$

1. Integrating the PDF, we see that we require

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy\, dx = c \int_{0}^{10} \int_{0}^{\infty} x e^{-2xy} dy\, dx = 5c,$$

   so we conclude $c = \frac{1}{5}$.

2. What is $P(X \leq 6 \text{ and } Y > 2)$?

$$\frac{1}{5} \int_{0}^{6} \int_{2}^{\infty} x e^{-2xy} dy\, dx = \frac{1}{40}(1 - e^{-24}) \approx \frac{1}{40}.$$

Independent random variables

Recall, two events E and F are called **independent** (and use the notation $E \perp F$) if one's introduction as evidence does not affect the other's probability:

$$P(E \mid F) = P(E), \text{ or } P(F \mid E) = P(F).$$

Equivalently, this has an easier computational form:
the probability of the intersection EF equals the product of the individual probabilities:

$$P(EF) = P(E)P(F).$$

Independent random variables

Random variables work in a similar fashion.

Two discrete random variables, X and Y, are called **independent** and use the notation

$$X \perp Y$$

if evidence about one does not affect the other's probabilities:

$$P(X = x \mid Y = y) = P(X = x), \text{ or } P(Y = y \mid X = x) = P(Y = y).$$

(We'll discuss conditional distributions in detail at the end of the course, but preview them here.)

Conditional PMF

To use this definition, we define the *conditional PMF* of X, given $Y = y$, by the conditional probability definition for events (such that $P(Y = y) > 0$):

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

Note that this definition uses the joint PMF of X and Y and the marginal PMF of Y.

Independent RVs: checking joint vs marginal PMF

Thus, $X \perp Y$ means the joint PMF factors into the marginals: if

$$P(X = x \mid Y = y) = P(X = x),$$

then

$$\begin{aligned} p_{X,Y}(x, y) &= P(X = x, Y = y) \\ &= P(X = x)P(Y = y) = p_X(x)p_Y(y). \end{aligned}$$

We will often use this criterion to determine if two random variables are independent.

Independent RVs: checking joint vs marginal PMF

The experiment: flip a fair coin 4 times. Let

- $X =$ number of H flipped on all four flips and
- $Y =$ number of T flipped in the first three flips.

$X \sim Bin(4, \frac{1}{2})$, and $Y \sim Bin(3, \frac{1}{2})$, so we know the marginal PMFs.

X and Y are clearly dependent since they consider the same flips.

We will build the joint PMF for X and Y to formally verify this dependence.

Independent RVs: checking joint vs marginal PMF

	Y=0	Y=1	Y=2	Y=3	$p_X(x)$
X=0					
X=1					
X=2					
X=3					
X=4					
$p_Y(y)$					

To fill out the chart, first fill in the marginal PDFs on the edges.

Independent RVs: checking joint vs marginal PMF

	Y=0	Y=1	Y=2	Y=3	$p_X(x)$
X=0					$\frac{1}{16}$
X=1					$\frac{4}{16}$
X=2					$\frac{6}{16}$
X=3					$\frac{4}{16}$
X=4					$\frac{1}{16}$
$p_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Next, think about how X and Y are related.

For example, where does the flip sequence HTHT go in this chart?

Once all the sequences are placed, count to verify the marginals.

Independent RVs: checking joint vs marginal PMF

	Y=0	Y=1	Y=2	Y=3	$p_X(x)$
X=0	0	0	0	TTTT	$\frac{1}{16}$
X=1	0	0	TTHT, THTT, HTTT	TTTH	$\frac{4}{16}$
X=2	0	THHT, HTHT, HHTT	TTHH, THTH, HTTH	0	$\frac{6}{16}$
X=3	HHHT	THHH, HTHH, HHTH	0	0	$\frac{4}{16}$
X=4	HHHH	0	0	0	$\frac{1}{16}$
$p_Y(y)$	$\frac{2}{16}=\frac{1}{8}$	$\frac{6}{16}=\frac{3}{8}$	$\frac{6}{16}=\frac{3}{8}$	$\frac{2}{16}=\frac{1}{8}$	1

Now replace the sequences with their counts to get the joint PMF.

Independent RVs: checking joint vs marginal PMF

	Y=0	Y=1	Y=2	Y=3	$p_X(x)$
X=0	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$
X=1	0	0	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{4}{16}$
X=2	0	$\frac{3}{16}$	$\frac{3}{16}$	0	$\frac{6}{16}$
X=3	$\frac{1}{16}$	$\frac{3}{16}$	0	0	$\frac{4}{16}$
X=4	$\frac{1}{16}$	0	0	0	$\frac{1}{16}$
$p_Y(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	1

Conditional probabilities with random variables: example

- $X =$ number of H in 4 fair coin flips,
- $Y =$ number of T in the first 3 of those 4 flips,

We can see that X and Y are not independent, since, for example,

$$p_{X,Y}(0,0) = 0 \neq \frac{1}{16} \cdot \frac{1}{8} = p_X(0)p_Y(0).$$

Independent random variables: example

Let X = the number of fair die rolls it takes for a 1 to appear, and Y = the number of rolls after X it takes for an even to appear.

$$X \sim geom\left(\frac{1}{6}\right) \text{ and } Y \sim geom\left(\frac{1}{2}\right)$$

and their (marginal) CDFs are, for $a, b \in \{1, 2, 3, ...\}$,

$$F_X(a) = P(X \leq a) = \sum_{j=1}^{a} P(X = j) = \sum_{j=1}^{a} \left(\frac{5}{6}\right)^{j-1}\left(\frac{1}{6}\right) = 1 - \left(\frac{5}{6}\right)^{a}$$

$$F_Y(b) = P(Y \leq b) = 1 - \left(\frac{1}{2}\right)^{b}.$$

Independent random variables: example

Since X and Y consider independent die rolls, $X \perp Y$.

Thus, the joint CDF of the pair (X, Y) is the product of the marginal CDFs:

$$F_{X,Y}(a, b) = \left[1 - \left(\frac{5}{6}\right)^{a}\right] \left[1 - \left(\frac{1}{2}\right)^{b}\right] = F_X(a)F_Y(b).$$

Independent random variables implies expectation factors

If two random variables, X and Y, are independent, then the expected value of their product splits:

$$X \perp Y \implies E(XY) = E(X)E(Y).$$

THE CONVERSE IS NOT TRUE IN GENERAL!

$$E(XY) = E(X)E(Y) \text{ does NOT imply that } X \perp Y.$$

Independent random variables implies expectation factors

Example

Let X be a random variable, and $Y = X$. Clearly, X and Y are dependent on each other. In fact, the only way we get

$$E(XY) = E(X^2) = E(X)E(Y) = E(X)^2$$

is if X is a constant. (Why?)

Functions of random variables

A function of two random variables is also a random variable.

If $W = g(X, Y)$, then we can calculate the expectation, variance, etc. of W with the joint PMF of X and Y:

$$E(W) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} g(x, y) p_{X,Y}(x, y)$$

$$Var(W) = E(W^2) - E(W)^2$$

Functions of random variables

Example

Let $X \sim Bin(2, \frac{1}{3})$ and $Y \sim Bern(\frac{1}{5})$ be independent.

What is the expected value of $W = X - Y$?

$$E(W) = E(X - Y) = E(X) - E(Y) = \frac{3}{2} - \frac{1}{5} = \frac{13}{10}.$$

Joint cumulative distribution function (joint CDF)

The **joint CDF** of the random vector $(X_1, X_2, ..., X_n)$ is the function such that for any $s_1, s_2, ..., s_n \in \mathbb{R}$,

$$F(s_1, s_2, ..., s_n) = P(X_1 \leq s_1, X_2 \leq s_2, ..., X_n \leq s_n).$$

This joint CDF can be attained by integrating a joint PDF:

$$F(s_1, s_2, ..., s_n) = \int_{-\infty}^{s_n} \cdots \int_{-\infty}^{s_1} f(x_1, x_2, ..., x_n) dx_1 ... dx_n.$$

Joint cumulative distribution function (joint CDF)

Some properties of joint and marginal CDFs:

$$\lim_{a \to -\infty} F_{X,Y}(a,b) = \lim_{b \to -\infty} F_{X,Y}(a,b) = 0$$

$$\lim_{a \to \infty} F_{X,Y}(a,b) = F_{X,Y}(\infty, b) = F_Y(b)$$

$$\lim_{b \to \infty} F_{X,Y}(a,b) = F_{X,Y}(a, \infty) = F_X(a)$$

$$\lim_{b \to \infty} \lim_{a \to \infty} F_{X,Y}(a,b) = 1$$

Joint cumulative distribution function (joint CDF)

To get the probability of a RV X being in an interval, take the difference of its CDF at those two values (note the difference in the inequalities):

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

Joint cumulative distribution function (joint CDF)

The joint probability of two RVs, X and Y, both being in intervals, is calculated similarly:

$$
\begin{aligned}
&P(a < X \leq b,\, c < Y \leq d) \\
&\qquad = P(X \leq b, c < Y \leq d) - P(X \leq a, c < Y \leq d) \\
&\qquad = [P(X \leq b, Y \leq d) - P(X \leq b, Y \leq c)] \\
&\qquad\quad - [P(X \leq a, Y \leq d) - P(X \leq a, Y \leq c)] \\
&\qquad = [F_{X,Y}(b,d) - F_{X,Y}(b,c)] - [F_{X,Y}(a,d) - F_{X,Y}(a,c)].
\end{aligned}
$$

Joint CDF into joint PDF

We can, as in the one-dimensional version, recover the joint PDF from the joint CDF by differentiating along every variable.

If $F_{(X_1,X_2,...,X_n)}(x_1, x_2, ..., x_n)$ is the joint CDF of $(X_1, X_2, ..., X_n)$, and the joint PDF $f_{(X_1,X_2,...,X_n)}(x_1, x_2, ..., x_n)$ exists, then

$$f_{(X_1,X_2,...,X_n)}(x_1, x_2, ..., x_n) = \frac{\partial^n F}{\partial x_1 \partial x_2 \cdots \partial x_n}(x_1, x_2, ..., x_n).$$

$X \perp Y \iff$ joint CDF, PMF (or PDF), MGF factor

Theorem

The following are equivalent statements:

$$X \perp Y \iff \forall a, b \in \mathbb{R}, F_{X,Y}(a,b) = F_X(a)F_Y(b)$$

$$\iff \forall x, y,\ p_{X,Y}(x,y) = p_X(x)p_Y(y) \text{ (discrete)},$$
$$\text{or } f_{X,Y}(x,y) = f_X(x)f_Y(y) \text{ (continuous)}$$

$$\iff m_{X,Y}(t,s) = E(e^{tX+sY}) = E(e^{tX})E(e^{sY}) = m_X(t)m_Y(s).$$

$X \perp Y \iff$ joint CDF, PMF (or PDF), MGF factor

This is NOT true if only $E(XY) = E(X)E(Y)$!

$$E(XY) = E(X)E(Y) \implies \rho(X, Y) = 0,$$

i.e. X and Y are *uncorrelated*, but not necessarily independent.

(However, $\rho(X, Y) \neq 0$ does mean X and Y are dependent.)

Transformation of a joint PDF

How do we transform a joint PDF in

$$(X_1, X_2, ..., X_n)$$

into a joint PDF in

$$(Y_1, Y_2, ..., Y_n)$$

where

$$y = g(x)?$$

If this is possible, we can use the **Jacobian matrix** to do so.

Transformation of a joint PDF

Recall that, if $y = g(x)$ is an invertible function, then we can write

$$x = g^{-1}(y).$$

Generalizing the 1-dimensional transform the PDF of a continuous RV X into $Y = g(X)$,

$$f_Y(y) = f_x(g^{-1}(y))|(g^{-1})'(y)| = \frac{f_x(g^{-1}(y))}{|g'(g^{-1}(y))|},$$

the Jacobian is a matrix that holds all partial derivatives between the transformations. We'll examine the 2×2 case.

Jacobian matrix, determinant

The **Jacobian matrix** J of a coordinate transformation is the matrix that holds all partial derivatives of the transformation.

If we want to transform the RV (X_1, X_2) into $(Y_1, Y_2) = g(X_1, X_2)$, then we examine the coordinate functions and inverses

$$y_1 = g_1(x_1, x_2),\ y_2 = g_2(x_1, x_2);\ x_1 = h_1(y_1, y_2),\ x_2 = h_2(y_1, y_2).$$

If this is possible, we can use the **Jacobian determinant** (the determinant of the Jacobian matrix) to transform (X_1, X_2) into (Y_1, Y_2). The joint PDF of (Y_1, Y_2) is

$$f_{(Y_1,Y_2)}(y_1, y_2) = f_{(X_1,X_2)}(h_1(y_1, y_2), h_2(y_1, y_2)) \cdot \left| \begin{matrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} \end{matrix} \right|.$$

Rectangular and Polar Coordinates

We usually write points on the plane in

rectangular coordinates (x, y)

(our "usual" Cartesian coordinate plane).

However, we can transform these rectangular coordinates into

polar coordinates (r, θ),

where

- $0 \leq \theta < 2\pi$ is the angle between the positive x-axis and the vector (x, y), and
- $r \geq 0$ is the length of the vector (and the radius of the circle centered at the origin on which (x, y) is a point).

Rectangular and Polar Coordinates

From polar to rectangular:

$$x = r\cos\theta, \;\; y = r\sin\theta$$

From rectangular to polar:

$$r = \sqrt{x^2 + y^2}, \;\; \theta = \arctan\left(\frac{y}{x}\right) \; (x \neq 0).$$

Example: Box-Muller transformation

The **Box-Muller transformation** converts two independent uniforms into independent normals:

$$U_1, U_2 \sim Unif(0,1), U_1 \perp U_2 \longrightarrow X, Y \sim N(0,1), X \perp Y.$$

The transformation is

$$X = \sqrt{-2 \ln U_1} \cos(2\pi U_2),$$

$$Y = \sqrt{-2 \ln U_1} \sin(2\pi U_2).$$

Example: Box-Muller transformation

Using the polar coordinate transform

$$X = R\cos\Theta, \quad Y = R\sin\Theta,$$

we can find the distributions of R and Θ in terms of U_1, U_2:

$$R = \sqrt{-2\ln U_1}, \;\; \Theta = 2\pi U_2,$$

which yields

$$R \sim \chi^2(2) \sim Exp\left(\lambda = \frac{1}{2}\right), \;\; \Theta \sim Unif(0, 2\pi).$$