A. **Gamma Random Variable.** Consider again a **Poisson process** of arrival times, for example the arrival times of passengers to a metro stop, where the mean number of passengers that arrive per hour is λ. Recall that the **inter–arrival times** are independent exponential random variables with parameter λ: $T_1, T_2, T_3, \ldots$.

What is the time T until the 3rd passenger arrives?

$T =$

A sum of **iid** exponential random variables is a **gamma random variable**.

Recall "iid" stands for "independent and identically distributed".

Let's calculate the expected value of T:

$\mathbb{E}[T] =$

Recall the expectation and variance of $\mathrm{Exp}(\lambda)$ are:

$\mathbb{E} = \frac{1}{\lambda}$ and $\mathrm{Var} = \frac{1}{\lambda^2}$

And, since the summands are **indepedendent**, in this case variance preserves addition:

$\mathrm{Var}(T) =$

Finding the pdf of a sum is **not** a trivial task, even if the events are independent. Later we will discuss means for doing so. For now we state the pdf.

Gamma Random Variable. Let $T_1, T_2, \ldots$ be **iid** exponential random variable with parameter λ, for example inter–arrival times for a Poisson process. Then the sum T of the first n of these, i.e. the time until the nth arrival, is a **gamma random variable** with parameters n and λ.

$T = T_1 + \cdots + T_n \sim \Gamma(n, \lambda)$

Generally, a gamma random variable T with **shape parameter** $\alpha > 0$ and **rate parameter** $\lambda > 0$ is written:

α can be any positive number.

$T = \Gamma(\alpha, \lambda)$

and has pdf:

$$f(t) = \begin{cases} & \text{if } t \geqslant 0 \\ 0 & \text{if } t < 0 \end{cases}$$

As indicated, we will justify later in the course that this is in fact a sum of iid exponential random variables, when $\alpha = n \in \mathbb{N}$. If α is not in $\mathbb{N}$, then we cannot interpret this as a sum of iid exponential random variables. The reason for the gamma function in the denominator is to "normalize" the function, i.e. to ensure the total probability is 1.

It has expectation and variance:

$\mathbb{E}[T] =$

$\mathrm{Var}(T) =$

Example 1. Let's compute, using integrals for practice with the gamma function, the second moment of $T \sim \Gamma(\alpha, \lambda)$.

$\mathbb{E}[T^2] =$

Of course, we could easily calculate the second moment using the expectation on the variance. But I want practice with the gamma function here!

Recall: for $t \geqslant 0$ the pdf $T \sim \Gamma(\alpha, \lambda)$ is:

$$f(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^{\alpha-1}}{\Gamma(\alpha)}$$

Recall also the gamma function:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$$

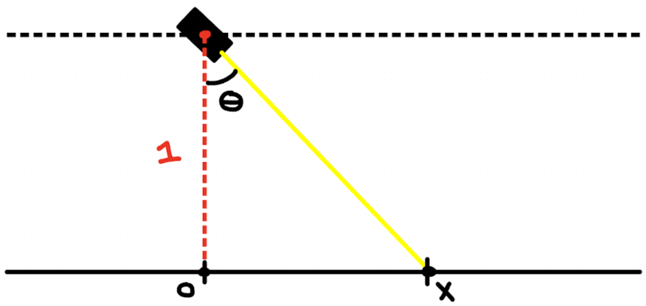
and its functional equation:

$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$

as long as $\alpha > 1$.

B. **Cauchy Random Variable.** Suppose that a narrow-beam flashlight is rotated left and right, then suddenly stopped, so that its angle Θ is a **uniform** random variable on $(-\frac{\pi}{2}, \frac{\pi}{2})$. Assume the flashlight is at position $x = 0$ and $y = 1$ in the xy–plane. Consider the point X at which the beam intersects the x-axis.

In other words, if you randomly stop the flashlight, where does the beam stike the wall (the x-axis)?



Recall: the cdf of a uniform random variable on (a, b) is, for $x \in (a, b)$:

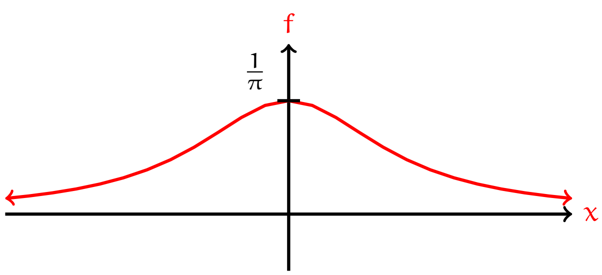
$$\mathbb{P}(X \leqslant x) = \frac{x - a}{b - a}$$

Let us calculate the cdf of X:

$F(x) = \mathbb{P}(X \leqslant x) =$

And thus its pdf is:

$f(x) =$



Let's calculate the expected distance of X from the origin:

$\mathbb{E}[|X|] =$

This showcases that expectation can **fail** to exist. Intuitively, this happens because the Cauchy distribution is **tail–heavy**, meaning that exceptionally large values are not exceptionally unlikely. The Cauchy distribution is also helpful for other situations with heavy tails, and symmetry about 0.

Cauchy Random Variable If Θ is uniform on an interval whose length is multiple of π, then $X = \tan\Theta$ is the **standard Cauchy** random variable.

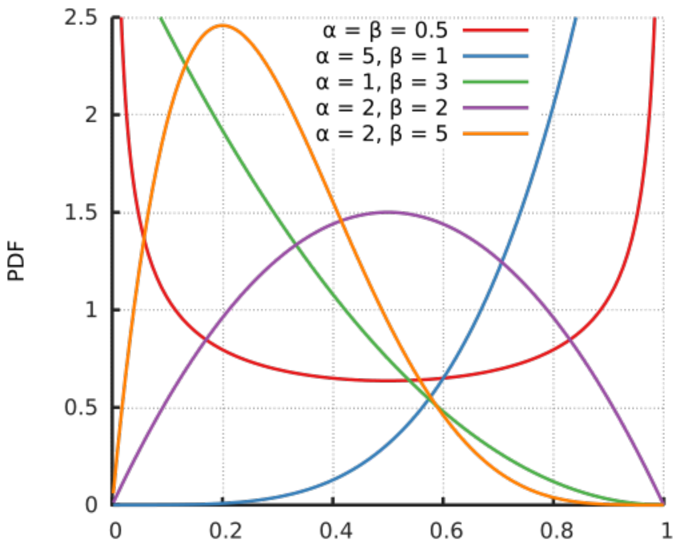
As we will show later, it can also be interpreted as a ratio of **independent** standard normal random variables:

A proof comes from a result that says a pair of independent standard normal variables X and Y can be written as $X = R\cos\Theta$ and $Y = R\sin\Theta$ where Θ is uniform on $[0, 2\pi)$ and R has a more complicated distribution. Then:

$$\frac{Y}{X} = \tan\Theta$$

which we have seen is precisely a Cauchy unifrom random variable.

C. **Beta Random Variable.** Sometimes, what we want is merely a random variable that is **flexible**, in that we can adjust parameters to better fit a situation. One of these is the **beta random variable** with **shape** parameters α and β. Here is a graph of its probability density function for various parameters. Click this Desmos link to play around with parameters.



As you can see, it only applies for random variables that have values only in the interval $[0, 1]$, like percents or fractions.

For example, the percent/fraction p of defective products in a manufacturing process comes with a bit of uncertainty, which we would model with a beta random variable. What makes it "flexible" is that there is a simple rule for updating parameters given new observations, for example as you take into account the number of defectives and nondefectives over the next month.

Without further explanation, here is the formula for its pdf.

Beta Random Variable. The **beta random variable** with shape parameters α and β is written $\mathrm{Beta}(\alpha, \beta)$ and has probability density function:

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

For example, suppose we make an initial estimate of our uncertainty about the fraction p of defectives in a manufacturing process using $\mathrm{Beta}(\alpha, \beta)$ for some choice of parameter α and β. This is called a **prior**. Then, let's say we observe x defectives and $n - x$ defectives. Due to properties of the beta distribution, we can update:

$$p \mid \text{data} \sim \mathrm{Beta}(\alpha + x, \beta + n - x)$$

The resulting new beta distribution is called a **posterior** and is a little tighter, as larger values of α and β tighten the curve near its peak. This is an example of a process called **Bayesian inference**.

This expression for $B(\alpha, \beta)$ ensures the total probability is 1. A proof that this is the correct normalizing factor can be found in Section 6.7 Example 7c from the textbook. That example shows that, if $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$ are independent gamma distributitons with the same rate parameter λ, then:

$$\mathrm{Beta}(\alpha,\beta) \sim \frac{X}{X+Y}$$

D. **Joint Mass Functions.** If we want to answer questions about probability involving combinations of random variables X and Y, for example their sum X and Y, we need to understand how their probability distributions depend on each other. First we address the **discrete** case.

Joint Mass. Let X and Y be **discrete** random variables on a common probability space. Their **joint mass** function is:

$p(x, y) =$

An important property of joint mass functions is:

Total Mass.

If you have a joint mass function, it is easy to recover the probability mass functions of X and Y, using the law of total probability.

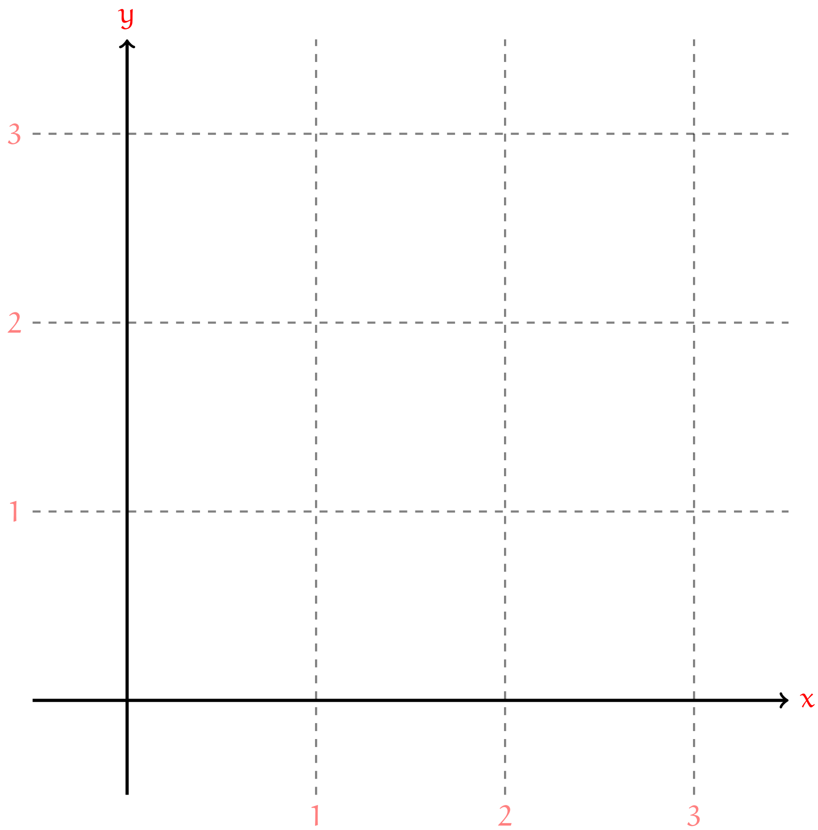
Marginal Mass. If $p(x, y)$ is the joint mass function, then the **marginal mass** functions are the probability mass functions of X and Y:

$p_X(x) =$

$p_Y(y) =$

Example 2. A box contains 2 red, 3 white, and 2 black balls. We draw 3 balls, without replacement, uniformly at random. Let X be the number of white balls drawn, and Y be the number of red balls drawn.

Find the joint mass function of X and Y, and indicate its **support** (the joint values of X and Y that occur with nonzero probability) in a grid.



Show how to use the grid to calculate:

$\mathbb{P}(X \leqslant 2, Y \leqslant 1) =$

$\mathbb{P}(X = 2) =$