A. **Weak Law of Large Numbers.** The weak law of large numbers intuitively says, if we take **many independent** samples of a random variable X, then, with high likelihoood, the average of those samples, what we call the **empirical average**, will be near $\mathbb{E}[X] = \mu$. Let's visualize this using Desmos.

Weak Law of Large Numbers. Let X be a random variable with $\mathbb{E}[X] = \mu$. The **empirical average** of n independent samples of X is:

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

where the X_i are independent and identically distributed to X. For any $\epsilon > 0$:

$$\lim_{n \to \infty}$$

We provide a proof in the case that $\text{Var}(X) = \sigma^2$ is finite.

The Cauchy random variable is an example where variance is infinite.

B. **Central Limit Theorem.** When we discussed the weak law of large numbers, we saw that if $X_1, X_2, \ldots$ were independent samplings of a random varable X, then the emperical average of these samplings converges to $\mathbb{E}[X] = \mu$ in a **weak** sense. This is useful in that it confirms that "expectation" means what it **should** mean. But it does not really help us understand the shape of the distribution of our samplings.

There is an official definition for **weak convergence**, that we will not state here.

The next limit theorem states that, after normalizing by measuring the emperical averages in units of standard deviation away from the mean, the scaling of these normalized averages by $\sqrt{n}$ approaches a standard **normal** distribution. Let's explore in Desmos.

Central Limit Theorem. Let X be a random variable with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$. Let $\overline{X}_n$ be the empirical average of n independent samples of X. Then:

$$\text{distribution}\left[\mathcal{N}(0,1)\right] = \lim_{n\to\infty} \text{distribution}\left[\sqrt{n}\cdot\frac{\overline{X}_n - \mu}{\sigma}\right]$$

More precisely, we will phrase this in terms of cdfs:

$$\lim_{n\to\infty} \mathbb{P}\left(\sqrt{n}\cdot\frac{\overline{X}_n - \mu}{\sigma} \leqslant a\right) =$$

We will establish this theorem in the special case that X has a moment–generating function, using the non–trivial fact that cdf's converge in the manner above if and only if moment–generating functions converge, provided the random variables involved have moment–generating functions.

The Cauchy random variable is an example of a function that does not have a moment–generating function.

Further, for simplicity, we will assume $\mu = 0$ and $\sigma = 1$.

Therefore, if we let $Z_n = \sqrt{n}\cdot\overline{X}_n$ then we must show:

Do you remember the moment–generating function of the standard normal random variable?