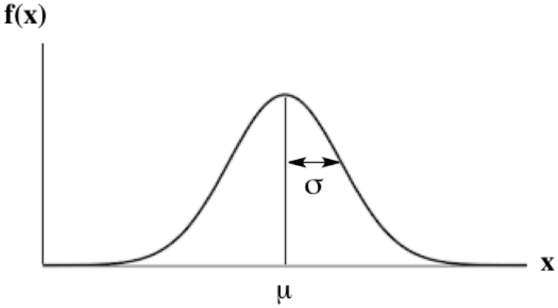
# CS 70 Discussion 12B

November 22, 2024

# Normal/Gaussian Distribution

**Problem**: What are we referring-to when we are talking about the "bell curve"?

**Solution**: We have a continuous distribution known as the **Gaussian distribution** or **normal distribution**. The PDF of a normal distribution of mean $\mu$ and **standard deviation** $\sigma$ looks like the following:



# Normal/Gaussian Distribution (Cont.)

For some Gaussian random variable $X$, the following always hold:

- Notation: $X$ being Gaussian with expectation $\mu$ and standard deviation $\sigma$ is denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$.
- Standard Normal Distribution: $X$ following a **standard normal distribution** is denoted as $X \sim \mathcal{N}(0, 1)$.
- PDF of $X$:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

- CDF: There is actually no closed-form CDF formula for a Gaussian distribution, and we only have approximations. We commonly denote the CDF for a standard Gaussian as $\Phi(x)$, and:

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

# Normal/Gaussian Distribution (Cont.)

For Gaussian random variable $X$, the following also always hold:

- $X \sim \mathcal{N}(\mu, \sigma^2)$ implies $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$
- For constants $a$ and $b$, $X \sim \mathcal{N}(\mu, \sigma^2)$ implies:

$$aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

- For constants $a$ and $b$, two independent Gaussians $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ follow:

$$aX + bY \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

# Central Limit Theorem (CLT)

**Problem**: When I have a distribution that is defined as the sum of $n$ **independent, identically-distributed (i.i.d.)** trials, why does it look like a Normal distribution?
**Solution**: Let's consider some random variable $S_n = \sum_{i=1}^{n} X_i$, where each $X_i$ are i.i.d. with mean $\mu$ and variance $\sigma^2$. The **Central Limit Theorem (CLT)** says that as $n \to \infty$:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1) \text{ and } \frac{\frac{1}{n}S_n - \mu}{\frac{\sigma}{\sqrt{n}}} \to \mathcal{N}(0,1)$$

This means that the **standardized** form of the distribution $S_n$ (i.e. $\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$) looks more and more like the standard normal distribution as $n$ grows larger. Therefore, for large $n$, we can use $\Phi$ to approximate probabilities of $S_n$ taking certain values.