COMP 585 | **Spring 2026**
Probabilistic Toolkit for Learning and Computing | *Instructor: Maryam Aliakbarpour*
February 10, 2026 | *Scriber: Maryam Aliakbarpour*

Lecture 9

Gaussian Random Variables

We want to study Gaussian random variables in more depth because they serve as the universal "attractor" for sums of independent random variables. We previously saw this in the context of the *Central Limit Theorem (CLT)*, which explains why the Gaussian distribution appears so frequently in nature and statistics. Today, we explore the specific mathematical properties of Gaussians.

Central Limit Theorem (CLT). To ground our discussion, recall the CLT. Informally, the sum of i.i.d. random variables acts like a Gaussian distribution. More formally, let $X_1, \ldots, X_m$ be m independent samples from a distribution with mean μ and bounded variance σ^2. If we define the sample mean as $\overline{X}_m = \frac{1}{m}\sum_{i=1}^m X_i$, then as m grows:

$$\frac{\sqrt{m}(\overline{X}_m - \mu)}{\sigma} \xrightarrow{d} N(0,1)$$

Equivalently, for any $t \in \mathbb{R}_{\geq 0}$:

$$\lim_{m\to\infty} \mathbf{Pr}\left[\frac{\sqrt{m}(\overline{X}_m - \mu)}{\sigma} \leq t\right] = \Phi(t)$$

This seemingly "weird" functional form for our statistic $\overline{X}_m$ is not arbitrary; it is effectively a *standardized version* of the empirical mean. We have carefully engineered it so that the mean and variance are 0 and 1, respectively. We verify the mean and the variance of $\overline{X}_m$ to ensure this normalization makes sense:

- **Expectation:** $\mathbf{E}\left[\overline{X}_m\right] = \frac{1}{m}\sum_{i=1}^m \mathbf{E}[X_i] = \mu$. Subtracting μ shifts the mean to zero.
- **Variance:** $\mathbf{Var}\left[\overline{X}_m\right] = \mathbf{Var}\left[\frac{1}{m}\sum_{i=1}^m X_i\right] = \frac{1}{m^2}\sum_{i=1}^m \mathbf{Var}[X_i] = \frac{\sigma^2}{m}$. The standard deviation of $\overline{X}_m$ is $\sigma/\sqrt{m}$. Dividing the shifted mean by this value makes the variance equal to 1.

Definitions. *Standard Normal (a.k.a. Gaussian)* random variable Z is a continuous variable with p.d.f.:

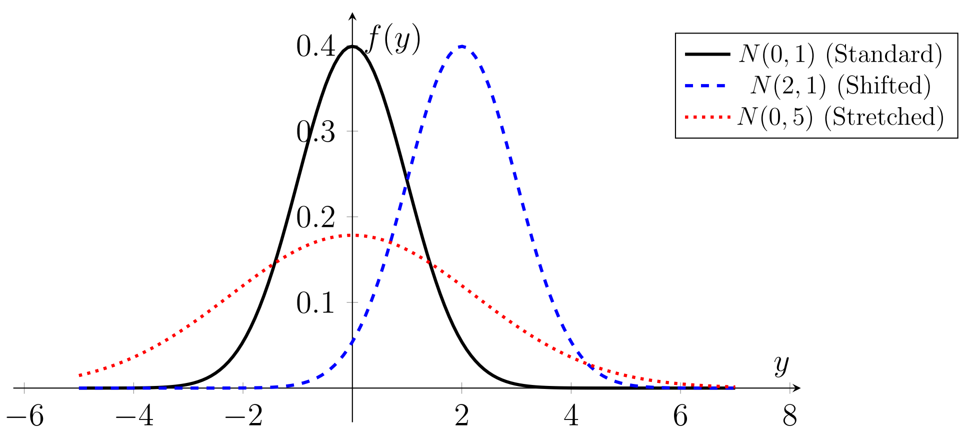
$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

The most critical component of this formula is the e^{-z^2} term, which represents the shape of the distribution and ensures the density drops toward zero extremely fast as we move away from the center. The denominator of 2 in the exponent serves to determine the width or "stretch" of the distribution, while the constant $1/\sqrt{2\pi}$ acts as a normalization term to ensure the total integral equals 1, making it a valid probability distribution.

For a *General Gaussian* $Y \sim N(\mu, \sigma^2)$, the p.d.f. is

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$

In this general form, the μ parameter simply acts as a shift that slides the entire distribution along the axis without changing its shape, and the σ^2 parameter modulates the stretch. A useful alternative definition is that any general Gaussian can be generated via the linear transformation $Y = \sigma Z + \mu$, where $Z \sim N(0, 1)$.



In the following, we examine several fundamental properties of Gaussian random variables

Fact 1: Rotational Symmetry

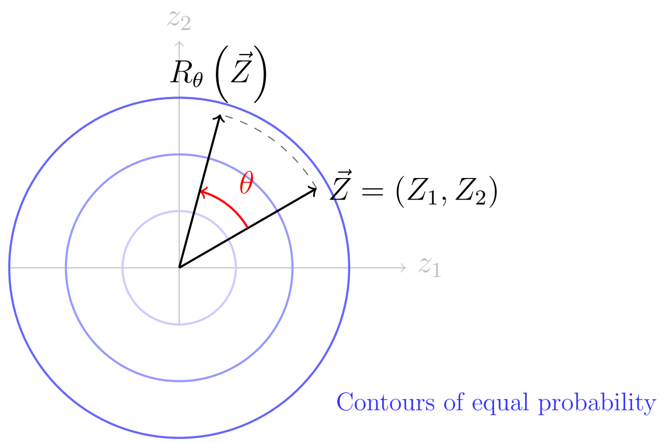
In higher dimensions, let $\vec{Z} = (Z_1, \dots, Z_d) \in \mathbb{R}^d$ be a vector of i.i.d. standard Gaussians. The joint p.d.f. is:

$$pdf(\vec{z}) = \phi(z_1)\dots\phi(z_d) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{||\vec{z}||_2^2}{2}}$$

Because the density depends only on the L_2-norm, the distribution is *rotationally symmetric*. That is, rotating $\vec{Z}$ by an angle θ results in a vector with the exact same distribution.

Another way to state this fact is that the following two random processes are identical:

1. We sample a vector $\vec{Z}$ directly by drawing d Gaussian random variables in every coordinate.
2. We sample $\vec{Z}$ in the same way, then apply a rotation R_θ to it, and output the result.



The Standard Normal distribution is very special because the joint density is perfectly symmetric in every direction. Because the contours of the Gaussian PDF are perfect circles centered at the origin, "spinning" the vector does not change the probability of where it lands. To an observer looking only at the output, there is no statistical test that could determine whether the rotation ever took place.

Fact 2: Closure Under Summation

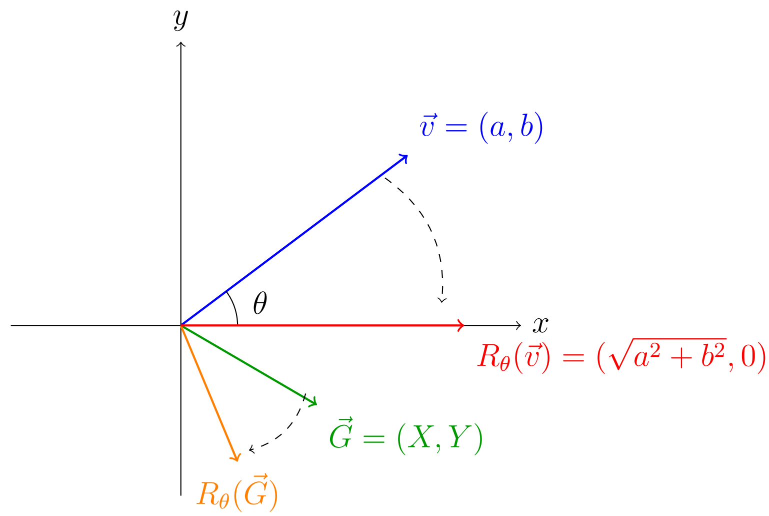
Our next property is the following: the sum of independent Gaussian random variables is itself a Gaussian. This is a characteristic already hinted at by the Central Limit Theorem (CLT). Recall the informal version of the CLT: the sum of m i.i.d. random variables acts like a Gaussian distribution. Thus, if we have a set of m samples $X_1, \ldots, X_m$, their sum $\sum_{i=1}^{m} X_i$ is a Gaussian. Similarly, for a second set of m samples $X_{m+1}, \ldots, X_{2m}$, their sum $\sum_{j=m+1}^{2m} X_j$ is also a Gaussian. If we look at the combined set of $2m$ samples, the total sum $\sum_{k=1}^{2m} X_k$ must also be a Gaussian according to the CLT. Since this total sum is simply the sum of two individual Gaussian variables, it implies that the sum of two Gaussians must itself be a Gaussian.

A more formal statement of this property is the following: let $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ be two independent Gaussian random variables. Let $Z = aX + bY$. Then:

$$Z \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

Using the linearity of expectation and properties of variance for independent variables, we can verify the moments. First, $\mathbf{E}[Z] = a\mathbf{E}[X] + b\mathbf{E}[Y] = a\mu_1 + b\mu_2$, which matches our target mean. Second, $\mathbf{Var}[Z] = a^2\mathbf{Var}[X] + b^2\mathbf{Var}[Y] = a^2\sigma_1^2 + b^2\sigma_2^2$, which matches our target variance.

Geometric Proof via Rotational Symmetry: Without loss of generality, assume $X, Y \sim N(0, 1)$. (Otherwise, we can use the fact that any non-standard Gaussian drawn from $N(\mu, \sigma)$ can be written as $X = \mu + \sigma X'$ where X is a standard Gaussian distribution.) We can view the linear combination $Z = aX + bY$ as a dot product between the constant vector $\vec{v} = (a, b)$ and the random vector $\vec{G} = (X, Y)$.



The key insight is that since $\vec{G}$ is rotationally symmetric, its distribution is invariant under any rotation R_θ (Fact 1):

$$R_\theta(\vec{G}) \overset{d}{=} \vec{G}\,.$$

Because the dot product is a geometric operation that is preserved under rotation, we have:

$$Z = \vec{v} \cdot \vec{G} = R_\theta(\vec{v}) \cdot R_\theta(\vec{G}) \overset{d}{=} R_\theta(\vec{v}) \cdot \vec{G}\,.$$

By choosing R_θ such that $R_\theta(\vec{v})$ aligns with the x-axis, the vector becomes $(\sqrt{a^2 + b^2}, 0)$. Substituting this back, we get:

$$Z \overset{d}{=} R_\theta(\vec{v}) \cdot \vec{G} = (\sqrt{a^2 + b^2}, 0) \cdot (X, Y) = \sqrt{a^2 + b^2} \cdot X\,.$$

Thus, Z has a Gaussian distribution as well.

Fact 3: The Berry-Esseen Theorem

While the Central Limit Theorem (CLT) ensures that the distribution of a normalized sum of independent random variables converges to a Gaussian in the limit, it does not provide information about the error for a finite number of samples m. The Berry-Esseen Theorem fills this gap by quantifying the rate of convergence, telling us exactly how quickly the distribution "settles" into its Gaussian shape.

Theorem 1. *Let $X_1, \ldots, X_m$ be i.i.d. random variables with mean μ, variance σ^2, and a finite third moment $\mathbf{E}_X[|X-\mu|^3] = \rho$. Let Y_m be the standard normalized sum:*

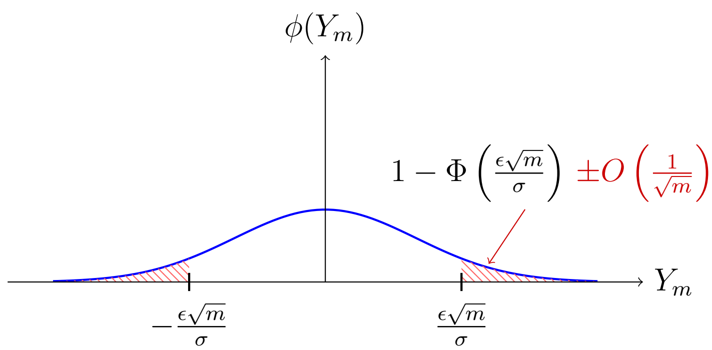
$$Y_m = \frac{\sqrt{m}(\overline{X}_m - \mu)}{\sigma}.$$

Then, the maximum difference between the cumulative distribution function (CDF) of Y_m and the standard normal CDF $\Phi(t)$ is bounded as follows:

$$\sup_{t\in\mathbb{R}} |\mathbf{Pr}[Y_m \le t] - \Phi(t)| \le \frac{C \cdot \rho}{\sigma^3\sqrt{m}}$$

where C is a universal constant (approximately less than 0.4748).

The term ρ/σ^3 is a measure of the "skewness" of the distribution. Distributions that are highly asymmetric or have "heavy" near-tails (high third moments) require a much larger m before they look Gaussian.



Example: Coin Bias Estimation. Consider m independent coin flips where $X_i \sim \mathbf{Ber}(1/2)$. If we want to know the probability that our sample mean $\overline{X}_m$ deviates from the true mean $1/2$ by more than ϵ, CLT says:

$$\mathbf{Pr}\left[\left|\overline{X}_m - \frac{1}{2}\right| > \epsilon\right] \approx 2\left(1 - \Phi\left(\frac{\epsilon\sqrt{m}}{\sigma}\right)\right)$$

The Berry-Esseen Theorem guarantees that our error in using this Gaussian tail to approximate the binomial probability is at most $O(1/\sqrt{m})$:

$$\begin{aligned}
\mathbf{Pr}\left[\left|\overline{X}_m - \frac{1}{2}\right| > \epsilon\right] &= 2\,\mathbf{Pr}\left[\overline{X}_m - \frac{1}{2} > \epsilon\right] && \text{(by symmetry of distribution)} \\
&= 2\,\mathbf{Pr}\left[\frac{\overline{X}_m - 1/2}{\sigma/\sqrt{m}} > \frac{\epsilon\sqrt{m}}{\sigma}\right] \\
&\leq 2\left(1 - \Phi\left(\frac{\epsilon\sqrt{m}}{\sigma}\right) + \frac{C\cdot\rho}{\sigma^3\sqrt{m}}\right) \\
&\leq 2\left(1 - \Phi\left(\frac{\epsilon\sqrt{m}}{\sigma}\right)\right) + \frac{1}{\sqrt{m}}\,. && \text{(using } \sigma^2 = 1/4, \rho = 1/8.\text{)}
\end{aligned}$$

Fact 4: Approximating the Tail Probability

Since the Gaussian CDF $\Phi(t) = \mathbf{Pr}[X \leq t]$ does not have a closed-form expression, we rely on tight analytic bounds to understand the behavior of the "tails"—the probability that a variable takes a value far from the mean. For a standard Gaussian $Z \sim N(0,1)$, we want to bound the tail probability:

$$\mathbf{Pr}[Z > t] = 1 - \Phi(t) = \int_t^\infty \phi(z)dz\,.$$

For $t > 0$, the tail probability is bounded as:

$$\frac{t}{t^2+1}\phi(t) \leq 1 - \Phi(t) \leq \frac{1}{t}\phi(t)\,.$$

As $t \to \infty$, the upper and lower bounds converge to each other. This implies that for large t:

$$1 - \Phi(t) \approx \frac{\phi(t)}{t} = \frac{e^{-t^2/2}}{t\sqrt{2\pi}}\,.$$

Proof of the Upper Bound: To prove the upper bound, we observe that for $z \geq t$, the inequality $1 \leq z/t$ holds. We can then rewrite the integral:

$$1 - \Phi(t) = \int_t^\infty \phi(z)dz \leq \int_t^\infty \frac{z}{t}\phi(z)dz\,.$$

Recalling that the derivative of the density is $\phi'(z) = -z\phi(z)$, we can evaluate the integral directly:

$$\frac{1}{t}\int_t^\infty z\phi(z)dz = \frac{1}{t}\left[-\phi(z)\right]_t^\infty = \frac{1}{t}(0 - (-\phi(t))) = \frac{1}{t}\phi(t)\,.$$

Proof of the Lower Bound: To show the lower bound, we consider the derivative of the function $g(t) = 1 - \Phi(t) - \frac{t}{t^2+1}\phi(t)$. Taking the derivative with respect to t:

$$\frac{d}{dt}\left(1 - \Phi(t) - \frac{t}{t^2+1}\phi(t)\right) = -\phi(t) - \left[\phi'(t)\frac{t}{t^2+1} + \phi(t)\frac{(t^2+1) - 2t^2}{(t^2+1)^2}\right] .$$

Substituting $\phi'(t) = -t\phi(t)$ and simplifying:

$$= -\phi(t)\left[1 - \frac{t^2}{t^2+1} + \frac{1-t^2}{(t^2+1)^2}\right] = -\phi(t)\left[\frac{(t^2+1)^2 - t^2(t^2+1) + (1-t^2)}{(t^2+1)^2}\right]$$

$$= -\phi(t)\frac{2}{(t^2+1)^2} .$$

Substituting our expression for $g'(x)$:

$$g(t) = -\left(0 - g(t)\right) = -\int_t^\infty g'(x)\,dx = \int_t^\infty \left(-g'(x)\right)\,dx .$$

Since the integrand $-g'(x)$ is strictly positive for all $x \geq t$, the integral from t to ∞ must also be strictly positive (> 0). Thus, $g(t) > 0$ for all $t \geq 0$, which yields the desired lower bound:

$$1 - \Phi(t) > \frac{t}{t^2+1}\phi(t) .$$

Bibliographic Note: The content of this lecture was partially adapted from the "CS Theory Toolkit" course notes and the accompanying lecture video by Ryan O'Donnell.