The Standard Normal Density Function

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

\[ E[X] = \mu \quad \text{Var}[X] = \sigma^2 \]

**X** is a normal (aka Gaussian) random variable  \( X \sim N(\mu, \sigma^2) \)
changing $\mu, \sigma$

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

density at $\mu$ is $\approx 0.399/\sigma$
X is a normal random variable  \( X \sim N(\mu, \sigma^2) \)

\[
Y = aX + b
\]

\[
E[Y] = E[aX+b] =
\]

\[
\text{Var}[Y] = \text{Var}[aX+b] =
\]

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

\[
X = a_1X_1 + a_2X_2 + \ldots + a_nX_n
\]

\( a_i \)'s const

\( X_i \)'s are indep.

\[
E(X) = a_1\mu_1 + a_2\mu_2 + \ldots + a_n\mu_n
\]

\[
\text{Var}(X) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \ldots + a_n^2\sigma_n^2
\]

\[
X \sim N\left( \sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2 \right)
\]
X is a normal random variable \( X \sim N(\mu, \sigma^2) \)

\[
Y = aX + b \\
E[Y] = E[aX+b] = a\mu + b \\
\text{Var}[Y] = \text{Var}[aX+b] = a^2\sigma^2 \\
Y \sim N(a\mu + b, a^2\sigma^2)
\]

Important special case: \( Z = (X-\mu)/\sigma \sim N(0,1) \)

\[
f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}
\]

\( Z \sim N(0,1) \) “standard (or unit) normal”

Use \( \Phi(z) \) to denote CDF, i.e.

\[
\Phi(z) = \Pr(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx
\]

no closed form 😞
The Standard Normal Density Function

\[ f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \]

\( \mu = 0 \)

\( \sigma = 1 \)

\( \mu \pm \sigma \)
If $Z \sim N(\mu, \sigma^2)$ what is $P(\mu - \sigma < Z < \mu + \sigma)$?

$P(\mu - \sigma < Z < \mu + \sigma) = \Phi(1) - \Phi(-1) \approx 68\%$

$P(\mu - 2\sigma < Z < \mu + 2\sigma) = \Phi(2) - \Phi(-2) \approx 95\%$

$P(\mu - 3\sigma < Z < \mu + 3\sigma) = \Phi(3) - \Phi(-3) \approx 99\%$

Why?

$\mu - k\sigma < Z < \mu + k\sigma \iff -k < \frac{(Z-\mu)}{\sigma} < +k$
Consider i.i.d. (independent, identically distributed) random vars $X_1, X_2, X_3, \ldots$

$X_i$ has $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \text{Var}[X_i]$.

Consider random variables

$$X_1 + X_2 + \ldots + X_n$$

and

$$\frac{1}{n} \sum_{i=1}^{n} X_i$$
the central limit theorem (CLT)

Consider i.i.d. (independent, identically distributed) random vars $X_1, X_2, X_3, \ldots$

$X_i$ has $\mu = E[X_i]$ and $\sigma^2 = \text{Var}[X_i]$

As $n \to \infty$,

$$\frac{X_1 + X_2 + \cdots + X_n - n \mu}{\sigma \sqrt{n}} \to N(0, 1)$$

Restated: As $n \to \infty$,

$$M_n = \frac{1}{n} \sum_{i=1}^{n} X_i \to N \left( \mu, \frac{\sigma^2}{n} \right)$$
CLT applies even to even wacky distributions
CLT in the real world

CLT is the reason many things appear normally distributed
Many quantities = sums of (roughly) independent random vars

Exam scores: sums of individual problems
People’s heights: sum of many genetic & environmental factors
Measurements: sums of various small instrument errors

...