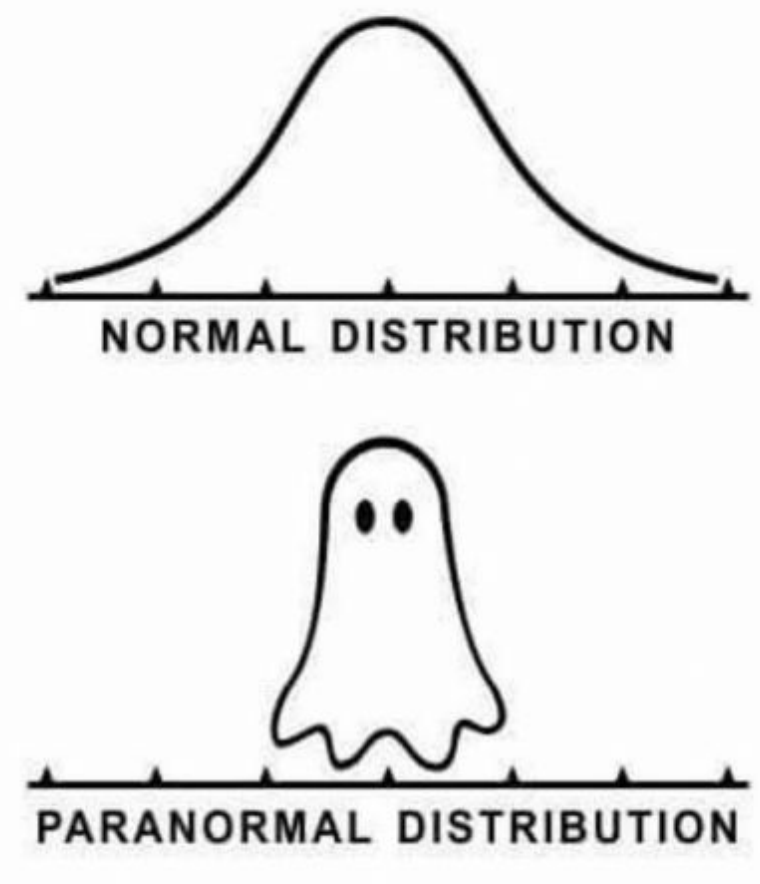
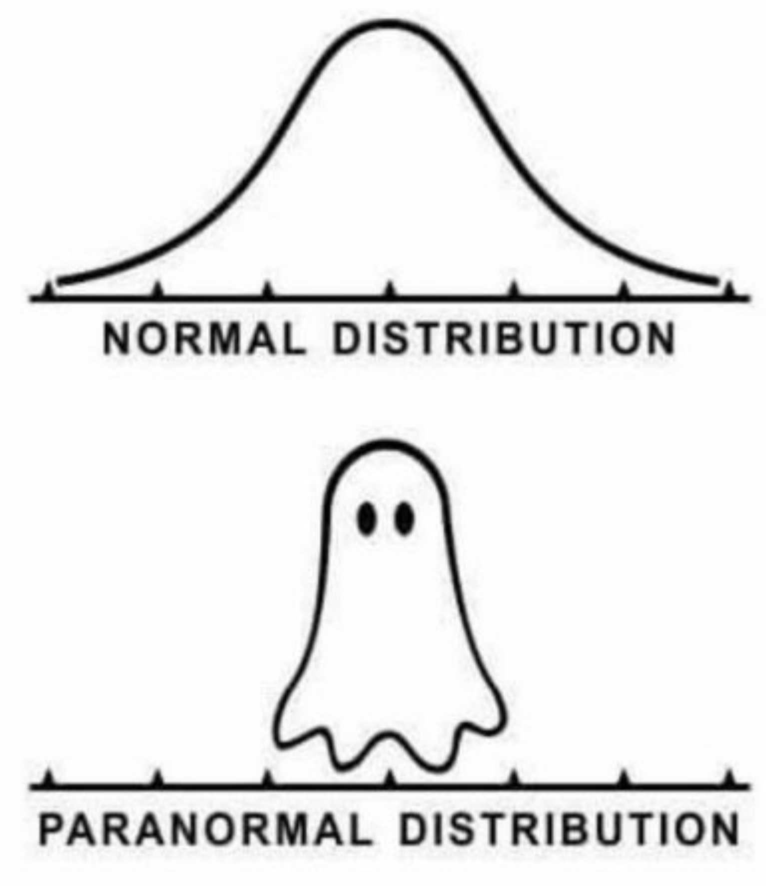
Lecture 12: The Normal Distribution



Announcements and reminders

- Practicum 1 due Monday, 11:59 PM



Previously, on CSCI 3022…

Definition: A random variable X is **continuous** if for some function $f : \mathbb{R} \to \mathbb{R}$ and for any numbers *a* and *b* with $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)\ dx$$

The function *f* must satisfy:

1) $f(x) \geq 0$ for all x, and 2) $\int_{-\infty}^{\infty} f(x)\ dx = 1$

Definition: The **cumulative distribution** (or density) **function** of X is defined such that

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\ dt$$

The Normal distribution

The **normal distribution** (AKA, Gaussian distribution) is probably the most important and widely used distribution in probability and statistics.

Many populations have distributions well-approximated by a normal distribution.

It's **very important** to check that Normal is a good approximation though! And **justify**.

Examples:

- Height, Weight, Other physical attributes
- Scores on a test
- Time it takes to travel

Consider: Why might Normal be an issue?

The Normal distribution

Definition: A continuous random variable X has a **normal (or Gaussian) distribution** with parameters μ and σ^2 if its probability density function is given by

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

We say $X \sim N(\mu, \sigma^2)$

Let's play around with this distribution: https://academo.org/demos/gaussian-distribution/

The Standard Normal distribution

Definition: The normal distribution with parameter values $\mu = 0$ and $\sigma^2 = 1$ is called the **standard normal distribution**.

Question: What is the pdf of the standard normal distribution?

The Standard Normal distribution

Definition: The normal distribution with parameter values μ = 0 and $\sigma^2 = 1$ is called the **standard normal distribution**.

Question: What is the pdf of the standard normal distribution?

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

The Standard Normal distribution

Definition: The normal distribution with parameter values μ = 0 and $\sigma^2 = 1$ is called the **standard normal distribution**.

Question: What is the pdf of the standard normal distribution? $$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

A standard normal random variable is usually denoted Z

Recall: The normal distribution does not have a closed form cumulative distribution function

→ We use special notation to denote the cdf of the **standard** normal distribution:

$$\Phi(z) = P(Z \leq z)$$

→ And usually we just look up values for $\Phi(z)$ in a table

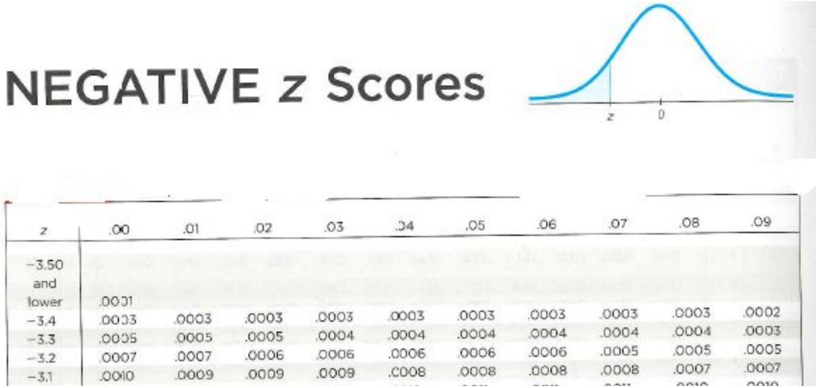
The Standard Normal distribution

The standard normal dist. **rarely** occurs in real life.

Instead, we take non-standard normal distributions, and **standardize** them using a simple transformation.

Recall: For computing probabilities, having a cdf is just as good (or better!) as having a pdf

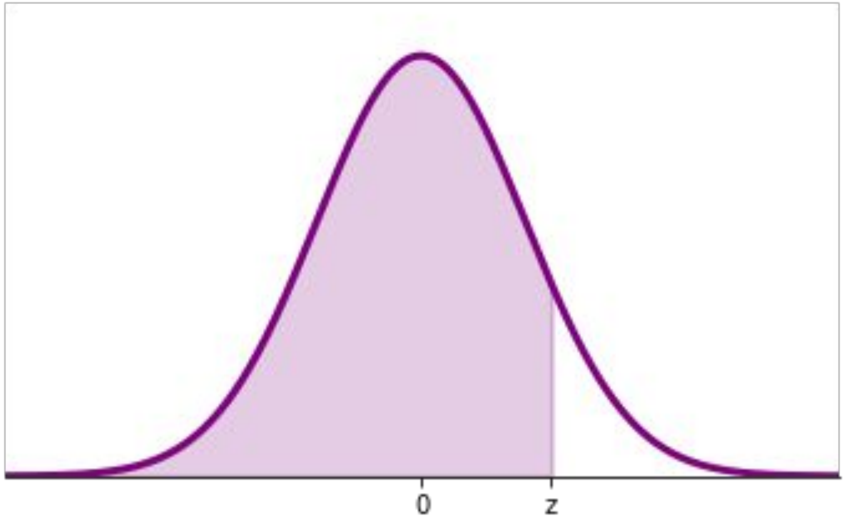
Back in MY day you had to look up values of the standard normal cdf in **tables** in the back of textbooks.


NEGATIVE z Scores

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007

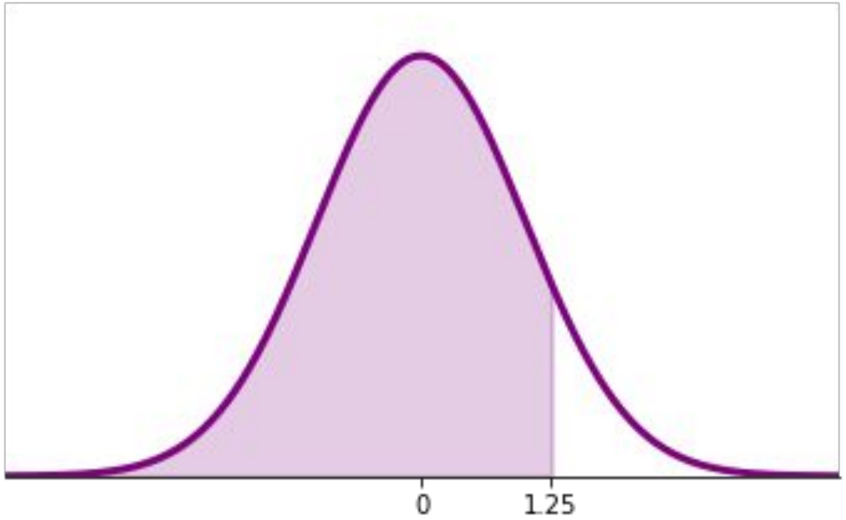
The Standard Normal distribution

$\Phi(z)$ = shaded area



The Standard Normal distribution

Example: What is $P(Z \leq 1.25)$?



The Standard Normal distribution

Example: What is $P(Z \geq 1.25)$?

Example: What is $P(Z \leq -1.25)$?

Example: How can we calculate $P(-0.38 \leq Z \leq 1.25)$?

Flip it *and* Reverse it

Example: What is the 99th percentile of $N(0, 1)$?

We have tables that tell us **area**… but we were given the area.

This is the **inverse** problem to $P(Z \leq z) = 0.99$

What about in Python?

- scipy.stats.norm.cdf
- scipy.stats.norm.pdf
- scipy.stats.norm.ppf

Flip it *and* Reverse it

Example: What is the 99th percentile of N(0, 1)?

We have tables that tell us **area**… but we were given the area.

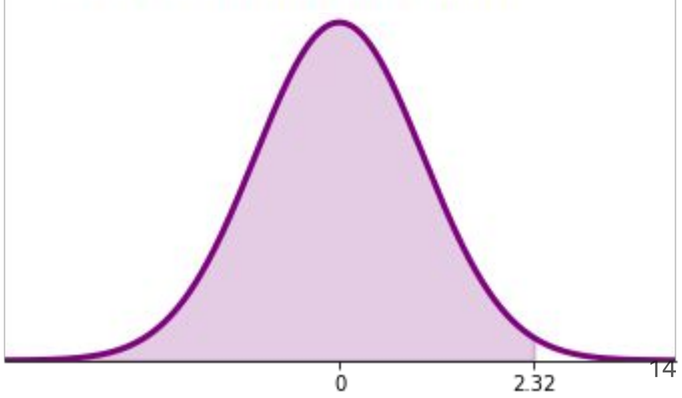
This is the **inverse** problem to $P(Z \leq z) = 0.99$

What about in Python?

- scipy.stats.norm.cdf
- scipy.stats.norm.pdf
- scipy.stats.norm.ppf

```python
from scipy import stats
stats.norm.ppf(.99)
```

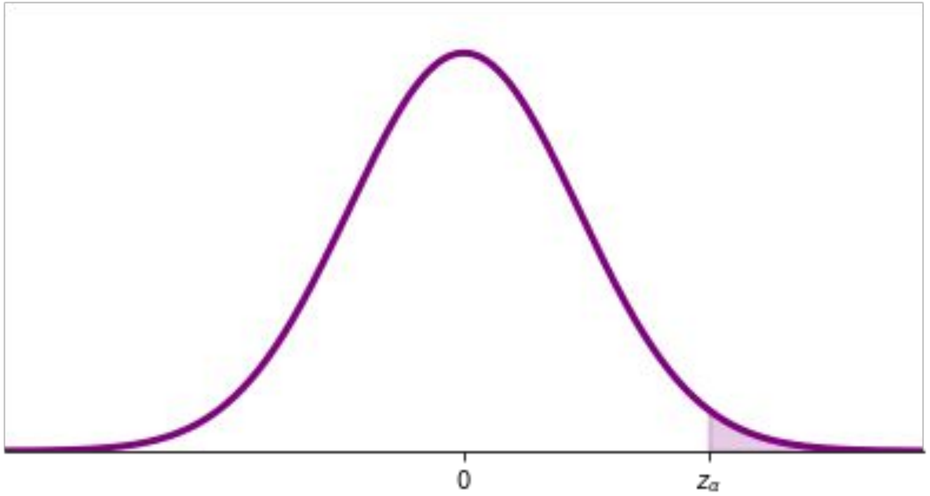
```
2.3263478740408408
```



The Critical Value

Notation: We say z_α is the **critical value** of Z under the standard normal distribution that gives a certain **tail area**. In particular, it is the Z value such that exactly α of the area under the curve lies to the **right** of z_α

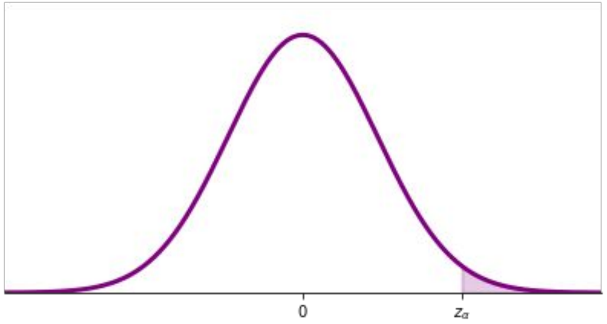
Note that other books/resources might use different conventions.
Be careful and **use sanity checks**!



The Critical Value

Notation: We say z_α is the **<u>critical value</u>** of Z under the standard normal distribution that gives a certain **tail area**. In particular, it is the Z value such that exactly α of the area under the curve lies to the **right** of z_α

Question: What is the relationship between z_α and the cdf?



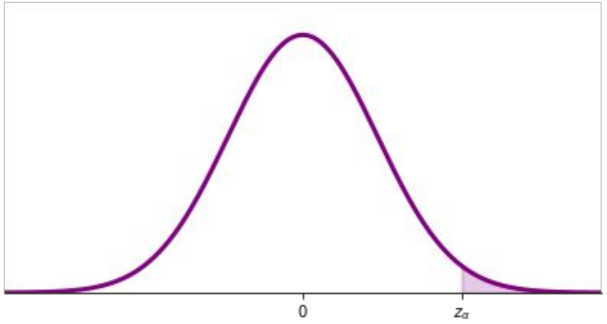
Question: What is the relationship between z_α and percentiles?

The Critical Value

Notation: We say z_α is the **critical value** of Z under the standard normal distribution that gives a certain **tail area**. In particular, it is the Z value such that exactly α of the area under the curve lies to the **right** of z_α

Question: What is the relationship between z_α and the cdf?

$$P(Z \geq z_\alpha) = \alpha = 1 - P(Z \leq z_\alpha) = 1 - \Phi(z_\alpha)$$



Question: What is the relationship between z_α and percentiles?

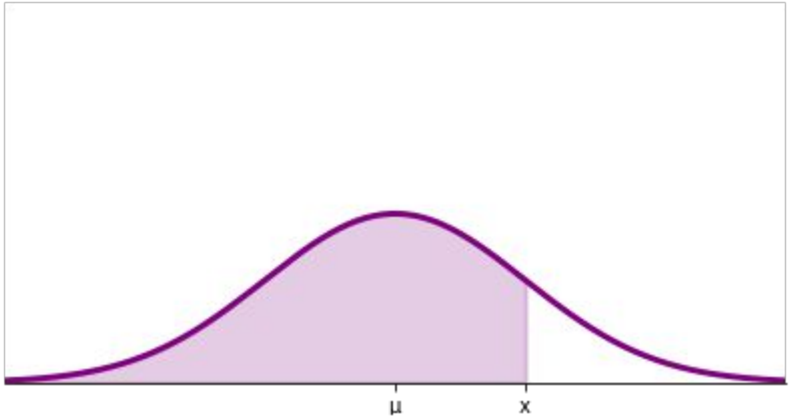
z_α is the $100(1-\alpha)^{th}$ percentile

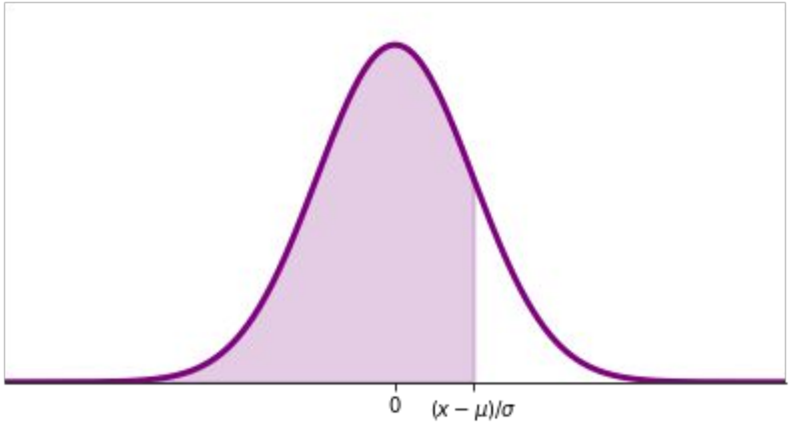
Non-standard Normal Distributions

Non-standard normal distributions can be turned into standard normals really easily

Proposition: If X is a normally distributed random variable with mean μ and standard deviation σ, then Z follows a standard normal distribution if we define:

$$Z = \frac{X - \mu}{\sigma} \qquad \text{and} \qquad X = \sigma Z + \mu$$





Brake lights!

Example: The time it takes a driver to react to brake lights on a decelerating vehicle is important for not getting into rear-end collisions.

The article "Fast-Rise Brake Lamp as a Collision Prevention Device" (linked here) suggests that reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled as a normal distribution having mean value 1.25 s and standard deviation 0.46 s.

Question: What is the probability that a reaction time is between 1.0 s and 1.75 s?

Brake lights!

Example: The time it takes a driver to react to brake lights on a decelerating vehicle is important for not getting into rear-end collisions.

The article "Fast-Rise Brake Lamp as a Collision Prevention Device" (linked here) suggests that reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled as a normal distribution having mean value 1.25 s and standard deviation 0.46 s.

Question: What is the probability that a reaction time is between 1.0 s and 1.75 s?

Follow-up question: What might be a potential problem with using a normal distribution? How can we check if it is much of an issue?

What just happened?

- We learned about the normal distribution!
- And the **standard normal distribution**
- And how to take any ol' normal random variable and **standardize it**