Normal Distribution

Nikos Apostolakis

April 7, 2025

1 Normal Distribution

In Figure 1 we see normal curves, with the same mean $\mu = 0$ and different standard deviations $\sigma = 1$, $\sigma = 0.5$, and $\sigma = 2$.

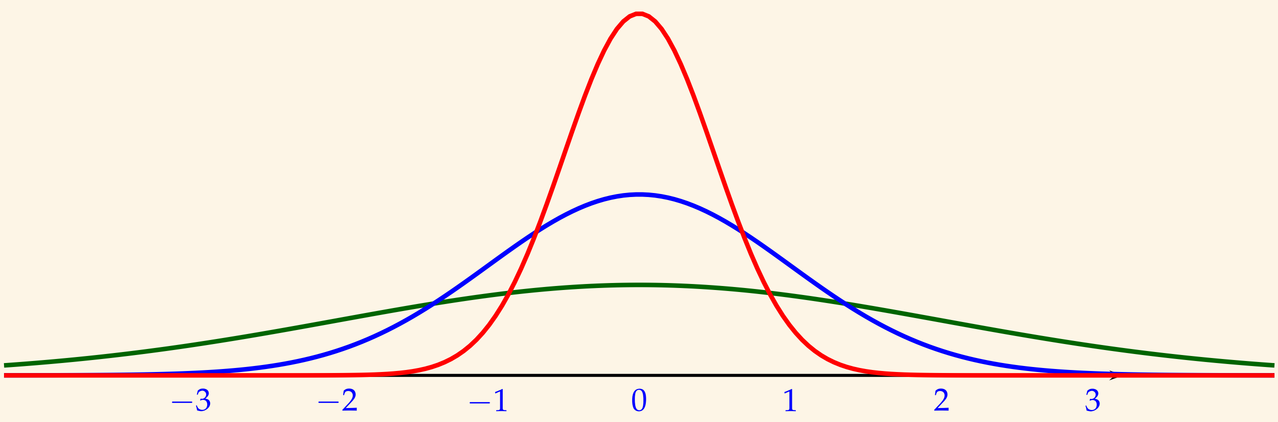

Figure 1: Normal distributions with the same μ and varying σ.

In Figure 2 we see normal curves, with the same standard deviation σ and different means $\mu = 2$, $\mu = 4$, and $\mu = -2$.

2 The Standard Normal Distribution

The *standard normal distribution* Z has $\mu = 0$ and $\sigma = 1$.

In class question

Use the tables to compute:

(a) $p(-3 < Z < 3)$.

(b) $p(-2 < Z < 2)$.

(c) $p(-1 < Z < 1)$.

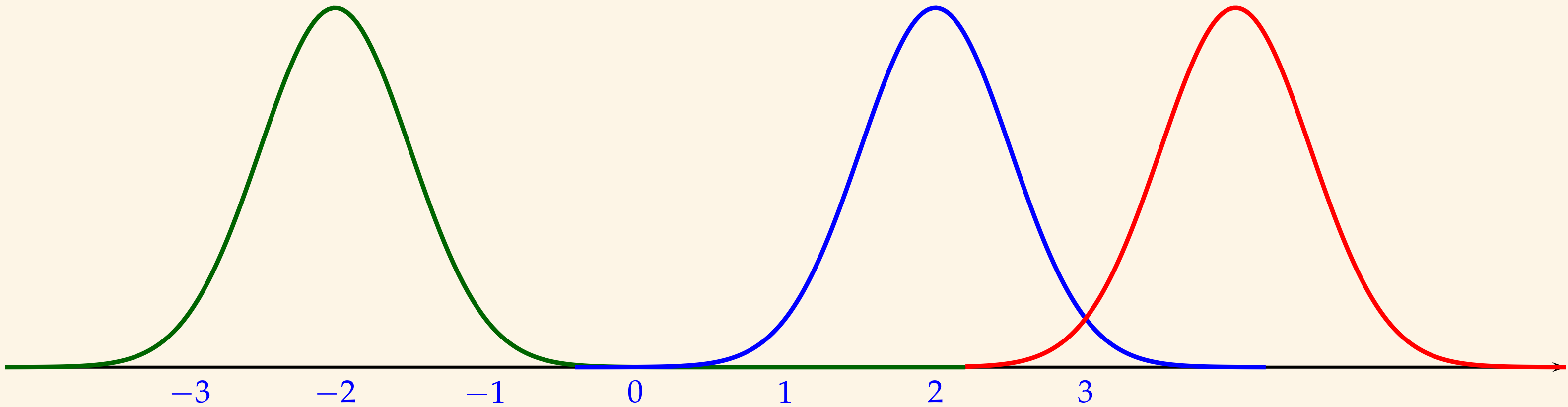

Figure 2: Normal distributions with varying μ and the same σ.

So the Empirical Rule holds exactly for the standard normal distribution, and actually all normal distributions.

Precise Empirical Rule

For Z have (see 3):

- 68.2% of the area is between -1 and 1.
- 95.4% of the area is between -2 and 2.
- 99.7% of the area is between -3 and 3.
- Only 0.135% of the area is outside $(-3, 3)$.

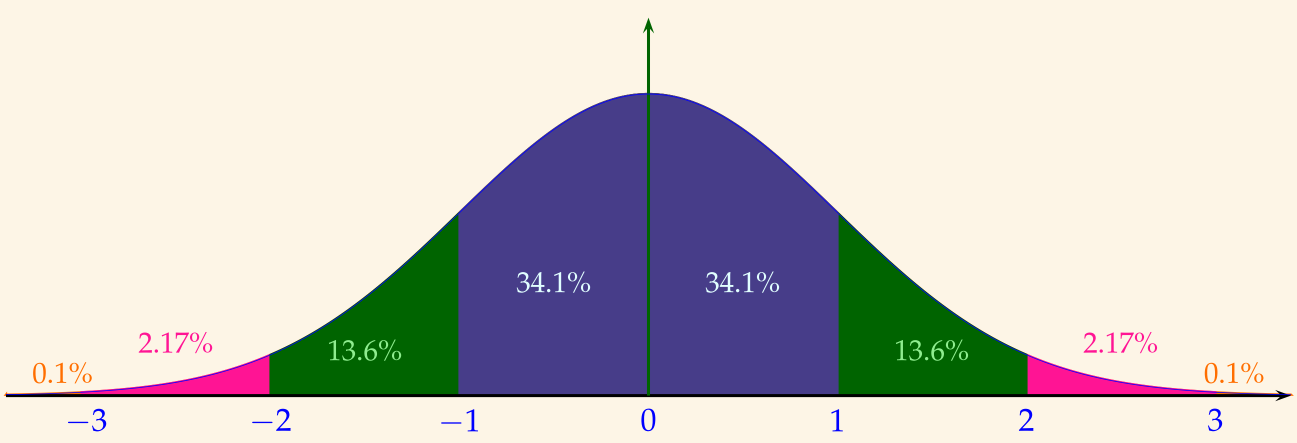

Figure 3: Empirical Rule holds precisely for normal distribution.

3 General Normal Distribution

Remember the formulas for going from raw scores x to z–scores and vice-versa.

If we know x (the raw score), to get z we use this formula

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

If we know the z–score to get the raw score x we use this formula

$$\boxed{x = \mu + z\sigma}$$

Example 1

Assume that the price of gold today is \$930 per ounce. The predicted price for gold in a month is normally distributed with a mean 930 and standard deviation 15. What is the probability that the price of gold in a month will be

(a) \$900 or below per ounce.

(b) Above \$950 per ounce.

(c) Between \$910 and \$960 per ounce?

(d) Find a range of prices that contains the price of one ounce of gold in a month with probability 99.7%.

Solution. (a) Let X be the price of an ounce of gold in a month. We need to find $p(X \leq 900)$. We convert the raw score into a $z - score$ and then compute the probability using the tables for Z.

$$x = 900 \implies z = \frac{900 - 930}{15} = -2.$$

So we have

$$\begin{aligned} p(X \leq 900) &= p(Z \leq -2) \\ &= 0.02275. \end{aligned}$$

(b) We are asked to find $p(X > 950)$. We calculate the z-score that corresponds to $x = 950$:

$$x = 900 \implies z = \frac{950 - 930}{15} = 1.33.$$

So $p(X > 950) = p(Z > 1.33)$. We need the *right tail*. There are two ways to find right tails:

1. **First way:** The right tail at $z = 1.33$ is the same as the left tail at $z = -1.33$. So

$$p(Z > 1.33) = p(Z < -1.33) = 0.09176.$$

2. **First way:** We can get the right tail at $z = 1.33$ by subtracting the left tail at $z = 1.33$ from 1. So

$$p(Z > 1.33) = 1 - p(Z < 1.33) = 1 - 0.90824 = 0.09176.$$

(c) We look for $p(910 < X < 960)$. We convert the raw scores $x = 910$ and $x = 960$ into z–scores:

$$x = 910 \implies z = \frac{910 - 930}{15} = -1.33, \quad x = 960 \implies z = \frac{960 - 930}{15} = 2.$$

Therefore

$$\begin{aligned} p(910 < X < 960) &= p(-1.33 < Z < 2) \\ &= p(Z < 2) - p(Z < -1.33) \\ &= 0.97725 - 0.09176 \qquad = 0.88549. \end{aligned}$$

(d) We know that $p(-3 < Z < 3) = 99.7\%$. So we need to find the raw scores x that correspond to $z = \pm 3$. Those are $x = \mu \pm 3\sigma$. Since $\sigma = 15$ we have $3\sigma = 45$, so

$$z = -3 \implies x = 930 - 45 = 885, \quad z = 3 \implies x = 930 + 45 = 975.$$

So the interval $(885, 975)$ contains the price of one ounce of gold in a month with probability 99.7%.

□

Example 2

Heights of adult men between 18 and 34 years of age are normally distributed with mean 69.1 inches and standard deviation 2.92 inches. One requirement for enlistment in the military is that men must stand between 60 and 80 inches tall. Find the probability that a randomly selected man from that age group meets the height requirement for military service.

Solution. Let X be the random variable that stands for the height of men between 18 and 34 years of age. We need to compute

$$P(60 < X < 80).$$

We will turn this into a question about the *Standard Normal Distribution* so we can use our tables. We first turn the raw scores $x = 60$ and $x = 80$ into z–scores.

$$x = 60 \implies z = \frac{60 - 69.1}{2.92} = -3.12, \quad x = 60 \implies z = \frac{80 - 69.1}{2.92} = 3.73.$$

So we have:

$$\begin{aligned} P(60 < X < 80) &= P(-3.12 < X < 3.73) \\ &= P(X < 3.73) - P(X < -3.12) \\ &= 0.99990 - 0.00090 \\ &= 0.999. \end{aligned}$$

□

Example 3

Lengths of time taken by students on an algebra proficiency exam (if not forced to stop before completing it) are normally distributed with mean 28 minutes and standard deviation 1.5 minutes.

(a) Find the proportion of students who will finish the exam if a 30–minute time limit is set.

(b) Six students are taking the exam today. Find the probability that all six will finish the exam within the 30–minute limit, assuming that times taken by students are independent.

Solution. (a) Let X be the random variable that stands for the length of time it takes a random student to finish the exam. We need to compute

$$p(X < 30).$$

We will turn this into a question about the *Standard Normal Distribution* so we can use our tables. We first turn the raw score $x = 30$ into a z–scores.

$$x = 60 \implies z = \frac{60 - 69.1}{2.92} = \frac{30 - 28}{1.5} = 1.33.$$

So we have:

$$\begin{aligned} p(X < 30) &= p(Z < 1.33) \\ &= 0.90824. \end{aligned}$$

(b) Since the times taken by students are independent, we can consider each student taking the exam as a trial of a Bernoulli experiment where "success" means the student finishes on time. If X is the number of students that finish in time then X is a binomial distribution with $n = 6$, and as we computed in Part (a), $p = 0.90$. We need to compute $p(X = 6)$. We use the PDF tables for the binomial distribution:

$$p(X = 6) = 0.5314.$$

So the probability that all six students will finish on time is 0.5414

□

Example 4

The scores in a standardized test are normally distributed with a mean of 530 and a standard deviation of 100. What is the probability that the score of a randomly selected student that took the exam is

(a) 780 or less?

(b) At least 400?

(c) Between 500 and 600?

(d) Jorge's score is 730. What is the percentile rank of Jorge's score?

Solution. (a) Let X stand for the score of the randomly selected student. The question asks for $p(X \leq 780)$. We start by calculating the z-score that corresponds to $x = 780$.

$$x = 780 \implies z = \frac{780 - 530}{100} = 2.5.$$

So we have

$$\begin{aligned} p(X \leq 780) &= p(Z \leq 2.5) \\ &= 0.97441. \end{aligned}$$

(b) The question asks for $p(X \geq 400)$.[1]. We have

$$x = 400 \implies z = \frac{400 - 530}{100} = -1.3.$$

[1] Note: this is a *right tail*

So we have

$$\begin{aligned} p(X \geq 400) &= p(Z \geq -1.3) \\ &= p(Z \leq 1.3) \\ &= 0.90320. \end{aligned}$$

(c) The question asks for $p(500 < X < 600)$. We have

$$x = 500 \implies z = \frac{500 - 530}{100} = -0.3, \quad x = 600 \implies z = \frac{600 - 530}{100} = 0.70.$$

So,

$$\begin{aligned} p(500 < X < 600) &= p(-0.30 < Z < 0.70) \\ &= 0.75804 - 0.38209 \\ &= 0.37595. \end{aligned}$$

(d) To find the percentile rank we need to know the percentage of scores that are less or equal to Jorge's score. But that percentage is exactly the left tail of $x = 730$ i.e. $p(X \leq 730)$.

We have

$$x = 730 \implies z = \frac{730 - 530}{100} = 2.00.$$

Therefore

$$\begin{aligned} p(X \leq 730) &= p(Z \leq 2) \\ &= 0.97725. \end{aligned}$$

Therefore the percentile rank of Jorge's score is 97.7.

□

Percentiles and left tails

- Percentile rank is a hundred times the left tail:

$$\boxed{\text{Percentile rank of } z = 100 \cdot p(Z \leq z).}$$

- Left tail is the Percentile rank divided by a hundred:

$$\boxed{p(Z \leq z) = \frac{\text{Percentile rank of } z}{100}.}$$

4 Inverse Normal Distribution

Often we know the probability we want and instead we want to find the value of X that gives this probability. Here is such an example

Example 1

The delivery time of packages delivered by a courier service is normally distributed with a mean of 14 hours and a standard deviation of 2 hours. What delivery time should the company *guarantee* if they want at least 95% of the packages will be delivered within the guaranteed time?

Solution. Now the company wants to find that particular time x^* that has left tail $p(X \leq x^*) = 0.95$. We will find z^* the z–score that has left tail 0.95 and then we will convert to a raw score x.

So now we look *inside* the table to find .9500. The exact value is not in the table (this will happen often) so we take the closest value 0.94950. This value corresponds to $z^* = 1.64$. Now we turn this in a row score

$$z^* = 1.64 \implies x^* = 14 + 1.64 \cdot 2 = 17.28.$$

So the company in order to have at least 95% of packages delivered within the guaranteed time the company should set the guaranteed time to 17.28 hours, that is about 17 hours and 17 minutes. □

Example 2

Heights of women are normally distributed with mean 63.7 inches and standard deviation 2.47 inches.

(a) What height is the 10th percentile?

(b) What height is the 80th percentile?

Solution. Let X be the random variable in question.

(a) The 10th percentile is that score x^* that has left tail $p(X < x^*) = 0.1$. Looking *inside* the tables we see that the closest left tail to 0.1 is 0.10027 for $z^* = -1.28$. We convert this to a raw score

$$z^* = -1.28 \implies x^* = 63.7 - 1.28 \cdot 2.47 = 60.5384.$$

So the 10th percentile is 60.54 inches.

(b) From the table we see that the closest left tail to .80 is 0.79955 at $z^* = 0.84$. The corresponding raw score is

$$x^* = 63.7 + 0.84 \cdot 2.47 = 65.7748.$$

So the 80th percentile is 65.77 inches.

□

Example 3

A company that makes 12-Volt batteries. Using data from many years of product testing the company knows that the life of their batteries follows a normal distribution with a mean of 45 months and a standard deviation of 8 months.

(a) The company guarantees a full refund on any battery that lasts less than 36 months. What percentage of batteries will the company have to refund?

(b) How long (to the nearest month) should the guarantee be If the company wishes to refund no more than 10% of its batteries?

Solution. Let X be lifetime, measured in months, of a randomly selected battery manufactured by the company. Then X is a normally distributed random variable with $\mu = 45$ and $\sigma = 8$.

(a) We are asked to compute the probability $p(X \leq 36)$. We have

$$x = 36 \implies z = \frac{36 - 45}{8} \approx -1.13$$

and therefore

$$p(X \leq 36) = p(Z \leq -1.13) = 0.12924.$$

Thus the company will refund 13% of its batteries.

(b) We want to find the value x^* such that

$$p(X \leq x^*) = 0.10.$$

Looking inside the table we see that the closest probability to 0.10 is 0.10027 for $z^* = -1.28$, in other words

$$p(Z \leq -1.28) = 0.10027.$$

Now

$$z^* = -1.28 \implies x^* = 45 - 1.28 \times 8 = 34.76.$$

Rounding to the closest integer we conclude that if the company guarantee lasts for at least 35 months, then it will refund no more than 10% of its batteries.

□

5 Questions

1. Men's shoe sizes nationwide are normally distributed with mean 10.5 and standard deviation of 1.2. What percentage of men have shoe size between 8.25 and 12.25?

2. Heights X of adult women are normally distributed with mean 63.7 inches and standard deviation 2.71 inches. Romeo, who is 69.25 inches tall, wishes to date only women who are shorter than he but within 4 inches of his height. Find the probability that the next woman he meets will have such a height.

3. Heights X of adult men are normally distributed with mean 69.1 inches and standard deviation 2.92 inches. Juliet, who is 63.25 inches tall, wishes to date only men who are taller than she but within 6 inches of her height. Find the probability that the next man she meets will have such a height.

4. The length of time required for a rodent to escape a noxious experimental situation was found to have a normal distribution with a mean of 6.3 and a standard deviation of 1.5 minutes.
   (a) What is the probability that a randomly selected rodent will take longer than 7 minutes to escape the noxious experimental situation?
   (b) Some rodents become so disoriented in the experimental situation that they don't escape in a reasonable amount of time. For how long should the researchers wait before they terminate the experiment if they wish to ensure that at least 90% of the rodents have enough time to escape?

5. The delay time on a subway line is normally distributed with a mean of 5 minutes and a standard deviation of 11 minutes. On a given day what is the probability that

(a) The train will be late? (In other words the delay time is more than 0 minutes).

(b) The train will be early?

(c) The train will be more than 5 minutes late?

(d) The train will be at least 10 minutes late?

6. Bags of M&M labeled 1.89 oz don't all have exactly the same weight. Their weight is normally distributed with a mean 1.89 oz and a standard deviation of 0.05 oz.

   (a) What is the probability that a randomly selected M&M bag labeled 1.89 oz actually weighs between 1.79 and 1.99 oz?

   (b) What is the probability that a randomly selected M&M bag labeled 1.89 oz will actually weigh less that 1.73 oz?

   (c) Find an interval (a range) of weights that contains the weight of 95.2% of M&M bag labeled 1.69 oz.

   (d) Find an interval (a range) of weights that contains the weight of 99.7% of M&M bag labeled 1.69 oz.

7. The *International Tennis Federation* defines the official diameter of a tennis ball as 2.57–2.70 inches. The diameters of tennis balls produced by a particular machine are normally distributed with mean 2.64 inches and a standard deviation of 0.045 inches.

   (a) What percentage of tennis balls produced by this machine have diameters within the limits set by the International Tennis Federation?

   (b) In a trial run of the machine 25 tennis balls were produced. What is the probability that *at least* 20 of those balls have a diameter within the limits set by the International Tennis Federation?

8. The watches manufactured by a company have a normally distributed lifetime with mean 28 months and standard deviation of 5 months.

   (a) If the company guarantees a full refund for any of its watches that last less than 2 years, what percentage of watches should it expect to replace?

   (b) If the company wishes to refund no more than 12% of its watches how long should the guarantee period be (to the nearest month)?

9. Tests of a new light bulb led to an estimated mean life of $1,321$ hours and standard deviation of 106 hours. The manufacturer will advertise the lifetime of the bulb using the largest value for which it is expected that 90% of the bulbs will last at least that long. Assuming bulb life is normally distributed, find that advertised value.

6 More challenging problems

1. Refer to Question 5. MTA wants to adjust the scheduled arrival time so that that line appears to operate more efficiently. The want to adjust the scheduled arrival time so that only 10% of the trains would arrive late, that is later than the scheduled time. How many minutes should they add to the current scheduled time to accomplish this?

2. On the final exam, a statistics instructor always gives A to the highest 10%, B to the next 20%, C to the next 40%, *D* to the next 20%, and *F* to the lower 10%. The final exam was normally distributed with mean 70 and standard deviation 10. What ranges of scores will get A, B, C, D, and F?

3. A vending machine dispenses coffee into 6–ounce cups in such a way that the actual amount of coffee dispensed is normally distributed with a standard deviation of .07 ounce. The vending machine has a dial in the interior so that the mean amount of coffee dispensed can be set at any desired level. At what level should they set the mean so that the 6–ounce cup will overflow no more of 2% of time?

4. The levels of sugar in the urine of people with diabetes are normally distributed with a mean of 56 and a standard deviation of 3, while for non diabetic people it is normally distributed with a mean of 46 and a standard deviation of 10. A hospital delivers a preliminary test to patients and those having urine sugar level above 50.5 are scheduled for further tests, while those with levels less than 50.5 are deemed non–diabetic.
   (a) What percentage of patients that are not actually diabetic is unnecessarily scheduled for further tests?
   (b) What percentage of diabetic patients are incorrectly judged deemed to not be diabetic?

Writing Assignment

In Question 4 we have two different types of errors in a medical test:

Type I	**False Positive:** A person who does not have the disease is deemed diabetic.
Type II	**False Negative:** A person who actually has the disease is deemed healthy.

This dichotomy occurs in many other areas; for example in a trial an innocent person may be convicted or a guilty person may be acquitted.

It is an unfortunate fact of nature that any procedure that decreases the probability of one type of error, increases the probability of the other type of error. So often we need to decide which type of error is more serious in a given situation and minimize the probability of committing that type of error, and live with the fact that by doing so we have increased the probability of the other type of error.

(a) Write a paragraph stating your opinion, with reason, of what type of error is more serious in the case of a medical examinations.

(b) Write a paragraph stating your opinion, with reason, of what type of error is more serious in the case of a trial.

There is no right or wrong answer in this question. Just state your opinion an explain the reasons that you hold it.