Basic Statistics in One Hour

Evan Russenberger-Rosica

June 10, 2023

Contents

1 Introduction

Sometimes, we like to know stuff about things. If there are only a few things, it's pretty easy to examine them all. However, there are alot of things in the world, and it can be expensive or time consuming to look at them all. Wouldn't it be great too only look at some of the things but still make inferences about the whole group? This is the goal of statistics. When we only "look at some of the things" we call this taking a **sample**. The totality of the things from which we sampled is called the **population**.

Example 1.1. Suppose we want to know the average height of all adult human males. We could, theoretically, attempt to actually measure the heights of the entire population of several billion adult males and compute the average. However, it would be much more economical to take a sample of perhaps several thousand people and attempt to infer the population mean using statistics.

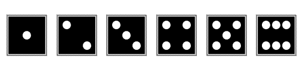
After reading the example, you might think that this seems like cheating. You argue, "the information we get from a sample of a few thousand can't be as good as the information from measuring the whole population of billions." And you would be right; taking a sample does introduce uncertainty into our measurements. Fortunatley, statistics provides rigorous ways to quantify this uncertainty.

2 Discrete Random Variables and Distributions

One way that statisticians quantify uncertainty is through the use of distributions. A **distribution** describes the probabilistic behavior of an unknown quantity known as a **random variable** X.[1] If X can only take on a countable[2] number of values or **outcomes** X is said to be a **discrete random variable**. The set of all outcomes is refered to as the **sample space**. Specifically, for each possible outcome x, the distribution tells us the probability that the random variable X has value x, i.e. the probability that $X = x$ or $\mathrm{P}(X = x)$.

Example 2.1. rDice Spin

The sample space of one dice spin is $\{1, 2, 3, 4, 5, 6\}$.



We don't know the result of the spin ahead of time, but we can model it as a random variable X, where every outcome has probability $\frac{1}{6}$. For instance, the

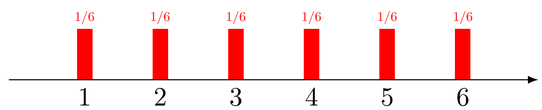
[1] It is traditional to use capital letters at the end of the alphabet to denote random variables and lower case letters to denote outcomes. As we will see later, it is traditional to use capital letters at the beginning of the alphabet to denote events.

[2] A set is countable if its cardinality is either finite or equal to $\aleph_0$, which is the cardinality of the natural numbers.

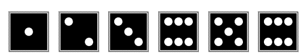
probability of rolling a one, i.e. is that $X = 1$, is $\frac{1}{6}$ and we write $P(X = 1) = \frac{1}{6}$. Since the probability of every event is $\frac{1}{6}$ we have:

$$P(X = 1) = P(X = 2) = \cdots = P(X = 6) = \frac{1}{6}$$

We can represent the distribution of the random variable X as a graph, where the outcomes are on the x-axis and the probabilities are on the y-axis. This graph shows probabilities as a function of the outcomes. Such a function is called a **probability mass function** or **pmf**. The probability mass function of X is shown below.



Exercise 2.2. Suppose that we add two more dots to the "four" side of the die, so that there are now two "six" sides. That is, the die looks like this:



Derive and plot the pmf of the new die.

3 Set Theory Review

Typically, we're interested not just in the probability of outcomes but **events**, which are sets of outcomes or equivalently subsets of the sample space. Typically we use capital letters from the beginning of the alphabet to denote events. Since events are just sets, we can manipulate them as such, i.e. by using the intersection, union, and complement operations. Thus, we briefly review set theory in this seciton.

A **set** is a collection of objects known as **elements** or **members**. If x is an an element of set S, we write $x \in S$, if x is not an element of S, we write $x \notin S$. Set A is a **subset** of set B if and only if every element of A is also an element of B. I.e. $A \subseteq B$ if $x \in A$ implies $x \in B$. If A is a subset of B, we write $A \subseteq B$. Observe that every set is a subset of itself, i.e. $A \subseteq A$. Two sets are **equal** if and only if they contain the same elements. We can show $A = B$ by showing $A \subseteq B$ and $B \subseteq A$.

3.1 Set Notation

We can denote a set using 2 methods. The simplest is the **roster method** in which we simply write the elements the set contains. When there is a clear pattern, we often write use ellipsis (...) to indicate that the pattern continues.

Example 3.1. The Roster Method

1. $A = \{1, 2, 3, 4\}$
2. $B = \{\ldots, -4, -2, 0, 2, 4, \ldots\}$

Definition 3.2. The Integers $\mathbb{Z}$

Some sets are so important that they have special names. One such set is the **integers**, i.e. the "whole numbers", and is written $\mathbb{Z}$. It is important to note that $\mathbb{Z}$ contains infinite elements.

$$\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

The roster method is fine for simple sets, however for large or complex sets, it is impractical. Thus we introduce **set builder notation**, which allows us to define a set by a rule which its elements must follow. Set builder notation has the following form $\{x \in S : P(x)\}$, which we read as "the set of all elements x in S such that some condition $P(x)$ is true.

Example 3.3. Set Builder Notation

1. $A = \{x \in \mathbb{Z} : 1 \leq x \leq 4\}$
2. $B = \{x \in \mathbb{Z} : x \text{ is even}\}$
3. $C = \{x : x = 2k, k \in \mathbb{Z}\}$

Definition 3.4. The Real Numbers $\mathbb{R}$

Another important set is the **Real Numbers**, written $\mathbb{R}$. Although a precise definition of $\mathbb{R}$ is outside of the scope of these notes, we will try to give an intuitive understanding. One can think of $\mathbb{R}$ as all the numbers that that might appear on an infitley long "number line" from $-\infty$ to ∞. $\mathbb{R}$ includes all integers, (so $\mathbb{Z} \subseteq \mathbb{R}$), fractions and irrational numbers (like π,e and $\sqrt{2}$), but not imaginary numbers like $3i + 2$. A critical property of the real numbers is that unlike $\mathbb{Z}$, there are no "gaps" in $\mathbb{R}$. That is to say, between every 2 real numbers, there is another real number. We call a set with this property a **dense set**.

Example 3.5. $\mathbb{Z}$ is not dense, because there is no element of $\mathbb{Z}$ between 3 and 4. However in $\mathbb{R}$ we can approach 3 arbitrarily closley (think $2.99\overline{9} \in \mathbb{R}$).

Definition 3.6. Open and Closed **Intervals**

An interval is another kind of set that is so commonly used that it has special notation. An interval is the set of all numbers in a set S which are between two endpoints a, b. If we include the endpoints in the interval, we call it a **closed interval** and write $[a, b]$. If we include the endpoints, we call this an **open interval**, and write (a, b). To be precise, we write:

$$[a, b] = \{x \in S : x \geq a \text{ and } x \leq b\}$$

$$(a, b) = \{x \in S : x > a \text{ and } x < b\}$$

Real intervals, i.e. where $S = \mathbb{R}$, are the most common type of intervals, although S can be any set where there is an ordering (i.e. where the idea of $a > b$ makes sense). Authors often write "Let $(a, b) \subseteq S$" so that it is clear which set is being discussed.

Exercise 3.7. Intervals of $\mathbb{Z}$

1. Let $Y = [3,8] \subseteq \mathbb{Z}$. What are the elements of Y?
2. Let $X = (3,8) \subseteq \mathbb{Z}$. What are the elements of X?
3. Let E be the even numbers, and $Z = (2,8) \subseteq E$. What are the elements of Z?

Exercise 3.8. Let $a, b \in \mathbb{R}$ and $a \neq b$. Find a number $c \in \mathbb{R}$ such that $c \in (a,b)$. If this is not possible, give a counterexample.

Exercise 3.9. Let $a, b \in \mathbb{Z}$ and $a \neq b$. Find a number $c \in \mathbb{Z}$ such that $c \in (a,b)$. If this is not possible, give a counterexample.

Exercise 3.10. (Hard) Find an infinite sequence that starts at 1 and approaches 0.

Definition 3.11. The Empty Set $\emptyset$

The **empty set $\emptyset$** is the set which contains no elements. This means for any x, the statement $x \in \emptyset$ is false. Additionally, the empty set is a subset of every set. I.e. for every set A, it is true that $\emptyset \subseteq A$.

3.2 Set Operations

We can form new sets from old sets using **set operations**. The most common set operations are **union**, **intersection**, and **complement**.

Definition 3.12. Set Operations

Union $A \cup B := \{x : x \in A \text{ or } x \in B\}$

Intersection $A \cap B := \{x : x \in A \text{ and } x \in B\}$

Relative Complement $A - B := \{x : x \in A \text{ and } x \notin B\}$

Absolute Compliment $A^{\complement} := \{x : x \notin A\}$

Example 3.13. Set Operations

Let $A = \{0, 1, 2\}$ and $B = \{2, 3, 4\}$. Then:

- $A \cup B = \{0, 1, 2, 3, 4\}$
- $A \cap B = \{2\}$
- $A - B = \{0, 1\}$
- $B - A = \{3, 4\}$
- $(A - B) \cap (A \cap B) = \emptyset$
- $\emptyset \cap A = \emptyset$
- $\emptyset \cup A = A$

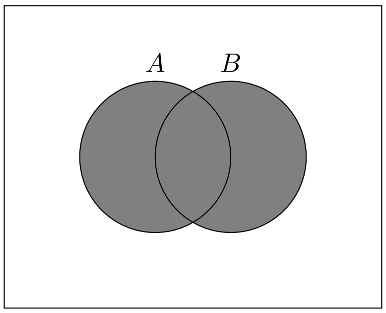


Table 1: $A \cup B$

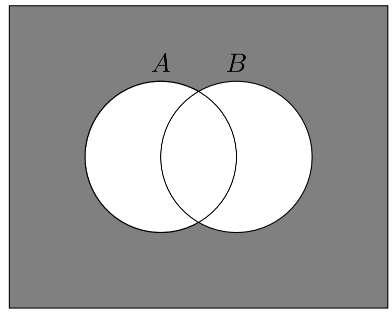


Table 2: $(A \cup B)^{\mathsf{C}}$

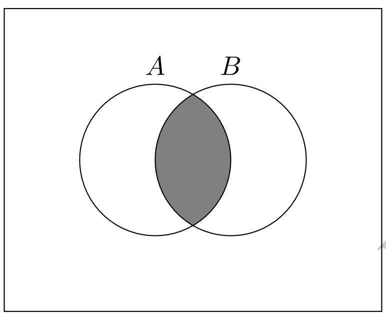


Table 3: $A \cap B$

rea

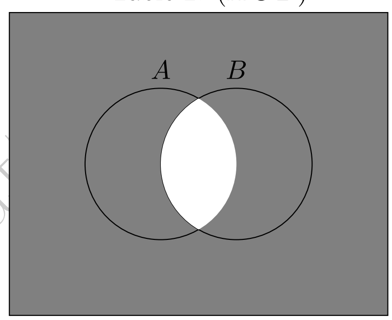


Table 4: $(A \cap B)^{\mathsf{C}}$

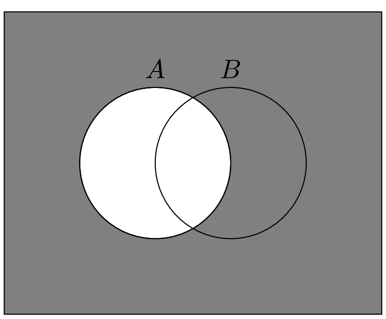


Table 5: A^{C}

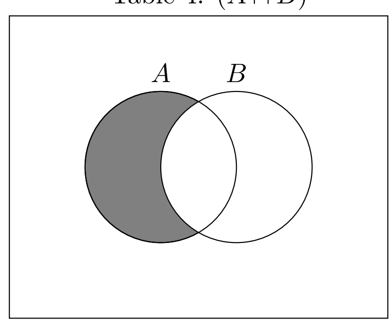


Table 6: $A - B$

Table 7: Set Unions, Intersections, and Complements

3.3 Visualization

4 Probability

Recall that since events are just sets, we can manipulate them using set operations. We define **disjoint events** as events which share no common elements. I.e. events A and B are disjoint if $A \cap B = \emptyset$. We are now in a position to state the three axioms of probability theory.[3]

Definition 4.1. Probability Axioms

1. Probabilites are non-negative. I.e, for any event A, $P(A) \geq 0$
2. All probabilities must sum to one. I.e, the probability of the sample space S is $P(S) = 1$
3. Countable additivity. This means that if $A_1, A_2, A_3, \ldots$ are disjoint events, then:

$$P\left(A_1 \cup A_2 \cup A_3 \cup \cdots\right) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

   I.e. the probability of the union of disjoint events is the sum of the individual probabilities.

Example 4.2. Events
Continuing from 2.1 on page 2, we define events A, B, C, D below.

A The event that a die toss results in an even number is the event $X \in \{2, 4, 6\}$

B The event that a die toss results in an odd number is the event $X \in \{1, 3, 5\}$

C The event that a die toss results in a perfect square is the event $X \in \{1, 4\}$ (1 is a perfect square because $1 = 1^2$. Likewise $2^2 = 4$).

D The event that a die toss results in a number less than 4 is the event $X \in \{1, 2, 3\}$

We are also interested in the probabilities of various combinations of events.

Theorem 4.3. *Law of Total Probability*
Let A be any event, and $A^{\complement}$ be it's complement (i.e. the event that A does not happen). Then $P(A^{\complement}) = 1 - P(A)$.

Example 4.4. Suppose 58% of Utah voters are Republicans. The probability that a Utah voter is not a Republican is $1 - .58 = 42\%$.

[3] An axiom is a mathematical statement which cannot be proved. Typically axioms are the starting point for the development of a mathematical theory.

Definition 4.5. Let A and B be two events.

- The event that A and B both occur is said to be the **intersection** of A and B and is written $A \cap B$. Mathematically, $A \cap B = \{x : x \in A \text{ and } x \in B\}$.
- The event that at least one of A or B occurs is the **union** of A and B and is written $A \cup B$. Mathematically, $A \cup B = \{x : x \in A \text{ or } x \in B\}$.

Thus $A \cap B$ and $A \cup B$ are just new events, i.e. sets of outcomes. Naturally, we may ask about the probability of $A \cap B$ and $A \cup B$, which we denote by $P(A \cap B)$ and $P(A \cup B)$ respectivley. With a descrete random variable, we can calulate these probabilities by simply computing the elments and computing $P(X = x)$ for each $x \in A \cap B$ or $A \cup B$.

Example 4.6. Computing Probabilities of Events

Let events A, B, C, D be defined as in 4.2 on the previous page.

- The event $A \cap C$ is the event that a die toss results in an even number and a perfect square. Since there is only one number (4) which is both even and perfect, $A \cap C = \{4\}$ and so $P(X \in A \cap C) = P(X \in \{4\}) = \frac{1}{6}$.
- The event $B \cap D$ is the event that a die toss results in an odd number which is less than 4. Thus $B \cap D = \{1, 3\}$ and so $P(B \cap D) = \frac{2}{6}$
- The event $A \cap B$ is the probability of a die toss that results in an odd number and an even number. There are no numbers which are both odd an even. Thus $P(A \cap B) = P(\emptyset) = 0$.

Exercise 4.7. Computing Probabilities of Events

Let A be the event that a die toss results in an odd number. Let B be the event that a die toss results in a number greater than 3. What is the probability

Theorem 4.8. *Probability Rules*

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

For instance, what the the probability of rolling a two or a four? In other words, what is $P(X \in \{2, 4\})$? We might correctly guess its $P(X = 2) + P(X = 4) = \frac{2}{6}$. We now consider a more challenging example:

Example 4.9. What is the probability of rolling an even number, or a number less than 3? I.e, what is $P(X \in \{2, 4, 6\} \text{ or } X \in \{1, 2\})$?

5 Continuous Random Variables and Distributions

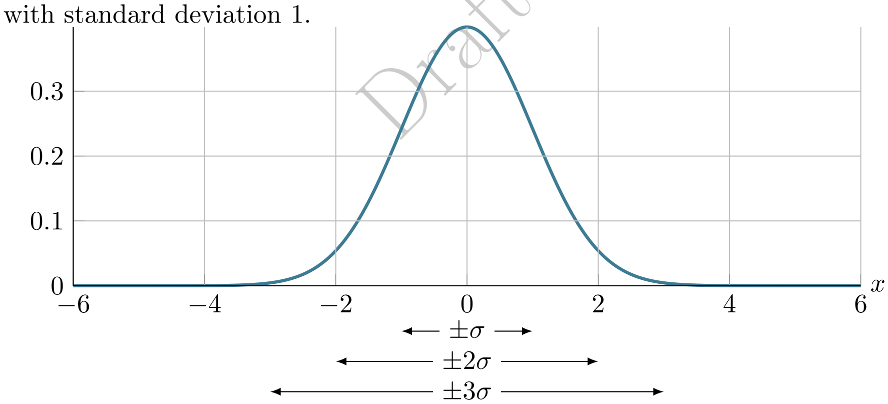
Recall that distribtuions describe the probabilistic behaivor of random variables. Also recall that a random variable X with a countable sample space is said to be a discrete random variable. A random variable which is not discrete is called a **continuous random variable**. Continuous random variable are represented by continuous functions with the property that the area under the curve and above the x-axis is 1.

The key property of a continous distribution is that the probability of a random variable X falling within a certain range or **interval** (a, b) is given by the area under the curve between points a and b. The probability of observing any single value is equal to 0, since the number of values which may be assumed by the random variable is infinite.

Continuous random variables are usually measurements. Examples include height, weight, the amount of sugar in an orange, the time required to run a mile.

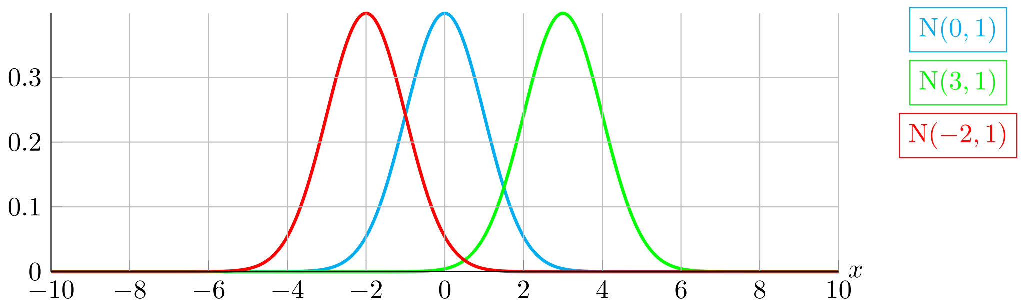
5.1 The Normal Distribution

The Normal Distribution is the most well known continuous distribution. It is parameterized by just two numbers, μ and σ called the **mean** and **standard deviation**. This means μ and σ alone determine the shape of the normal distribution. Thus, we denote a normal distribution with mean μ and standard deviation σ as $\mathrm{N}(\mu, \sigma)$, and write that $Z \sim \mathrm{N}(0, 1)$.[4] Below, we plot the **standard normal distribution** $\mathrm{N}(0, 1)$. This distribution is centered around mean 0 with standard deviation 1.

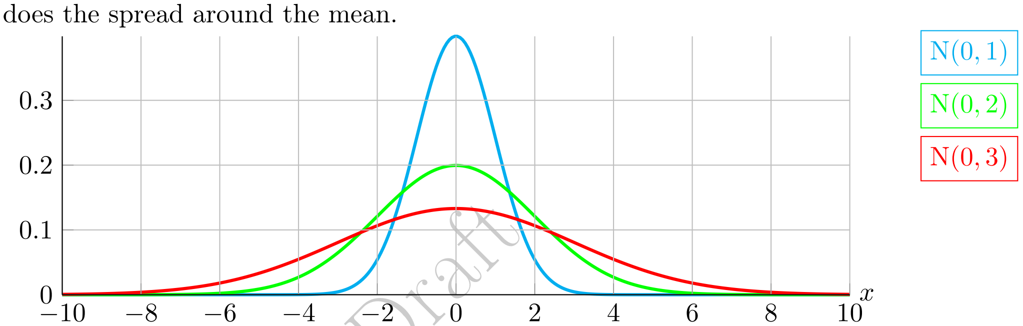


To see how the choice of μ affects the shape of the normal distribution, we overlay three normal distributions below, in cyan, we have $N(0, 1)$, in green, we have $N(3, 1)$ and in red we have $N(-2, 1)$. We observe that changing μ shifts the center of the distribution, but doesn't affect the spread at all.

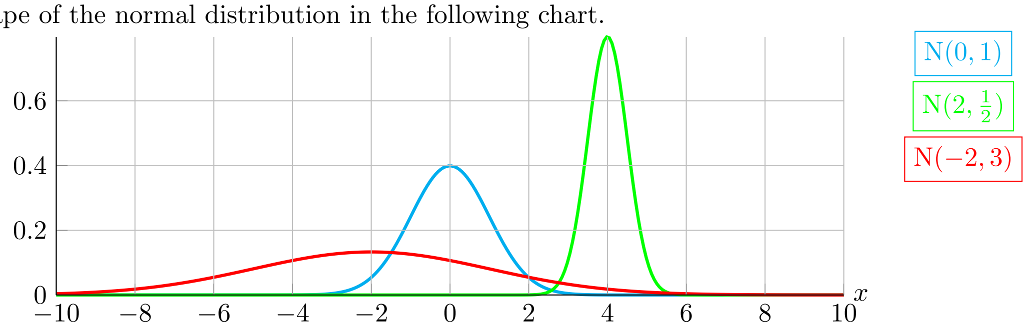
[4] Traditionally, it is common to represent a standard normal random variable with the letter Z



To see how the choice of σ affects the shape of the normal distribution, we overlay three normal distributions below, in cyan, we have $N(0, 1)$, in green, we have $N(0, 2)$ and in red we have $N(0, 3)$. We observe that as σ increases, the spread of the distribution around the mean increases. Likewise as σ decreases, so does the spread around the mean.

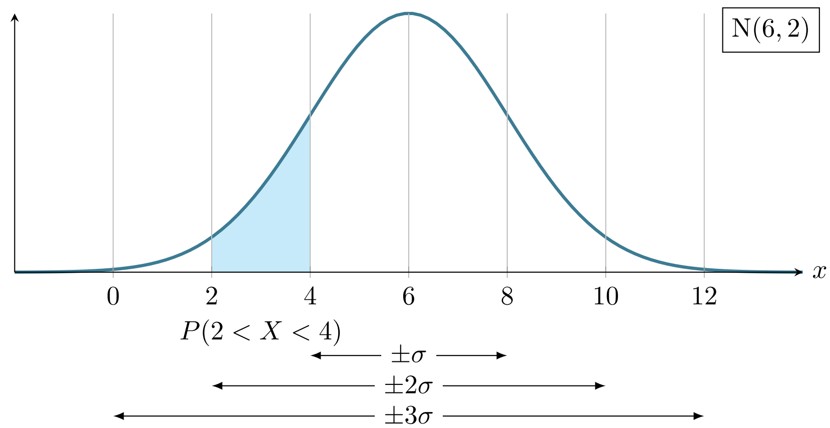


We deomnstrate how the choice of μ and σ can be combined to change the shape of the normal distribution in the following chart.

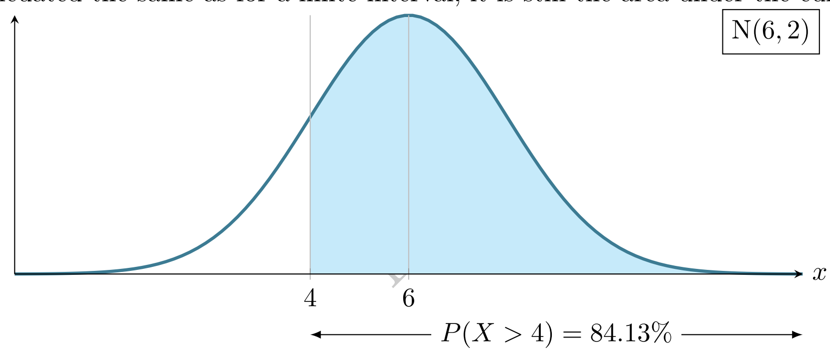


5.2 Probabilities in the Normal Distribution

To see how the normal distribution relates to probabilities, remember that the probability of a random variable X falling within a certain range or **interval** (a, b) is given by the area under the curve between points a and b. For instance, if $X \sim \mathrm{N}(6, 2)$ then $P(2 < X < 4) = 13.59\%$ is shown by the blue region below.



Frequently, we wish to know the probability of an infinite interval, for instance $P(4 < X) = P(4 < X < \infty) = 84.13\%$. In principal, this probability is calcuated the same as for a finite interval; it is still the area under the curve.



5.3 Computing Areas in the Normal Distribution

So how did we know that if $X \sim \mathrm{N}(6, 2)$ then $P(2 < X < 4) = 13.59\%$ and that $P(4 < X) = 84.13\%$? We simply computed the relevant areas (pictured above). But how did we do this? Well, it turns out that actually computing these areas is hard so typically, we use one of two methods.

5.3.1 Method 1: Use a Calculator

Modern computers are able to compute the area under a normal distrubtion in an interval (a, b) very quickly. Statistical programming languages like R and Python have such calculators built in, or easily accessable. There are also many websites which feature these calculators as well as computer programs like Microsoft Excel.

5.3.2 Method 2: Use a Z-table

Traditionally (i.e. before everyone had easy access to computers and the internet), normal probabilities were computed using a Z-table. The purpose of a Z-table is to allow users to calculate probabilities involving any $N(\mu, \sigma)$ distribution by first converting to the $N(0,1)$ distribution.[5] They do this by allowing the user to translate between a Z-score and a probability (and back).

Definition 5.1. Let x be an observation of $X \sim N(\mu, \sigma)$. Then x is referred to as a **raw score**. The **Z-score** $z(x)$, is simply the number of standard deviations x is above the mean μ.

Theorem 5.2. *Converting from a raw score to a Z-score.*

Given raw score x, we can compute the Z-score $z(x)$ as follows:[6]

$$z(x) = \frac{x - \mu}{\sigma}$$

Definition 5.3. A **Z-table** allows users to translate from Z-scores to probabilities (and back).

Example 5.4. Calculating Z-scores

Let $X \sim \mathrm{N}(6, 2)$. Then the standard deviation of X is $\sigma = 2$. What are the Z-scores of the following numbers: $4, 0, 6, 7$?

- $z(4) = \frac{4-6}{2} = \frac{-2}{2} = -1$
- $z(0) = \frac{0-6}{2} = \frac{-6}{2} = -3$
- $z(6) = \frac{6-6}{2} = \frac{0}{2} = 0$
- $z(7) = \frac{7-6}{2} = \frac{1}{2}$

These results should make intuitive sense. The Z-score of 4, written $z(4)$, is -1 because 4 is one standard deviation σ (of size 2) below the mean μ (which is 6). $z(6) = 0$ because $\mu = 6$ and 6 is 0 standard deviations above μ. $z(7) = \frac{1}{2}$ because 7 is $\frac{1}{2} = .5$ standard deviations (of size 2) above $\mu = 6$.

Now that we know how to calculate a Z-score, we can discuss how to use a Z-table. To reiterate, Z-tables allow users to translate from Z-scores to probabilities (and back). Specifically, given $z(x)$ the Z-table gives the the probability that $X < Z(x)$, i.e. it tells us $P(X < Z(x))$, which is often written simply $P(X < z)$. Using a z-table can be somewhat confusing at first, however it is not complicated. In the example below, we show how to compute Z-scores using a Z-table.

---

[5] This eliminates the need to have one table for each of the inifnitley many $N(\mu, \sigma)$ distributions!

[6] It can be shown that if $X \sim N(\mu, \sigma)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$. I.e. Thus we can think of calculating a Z-score as converting a $N(\mu, \sigma)$ random variable to a $N(0,1)$ random variable. Indeed, if we attempt to calculate the Z-score of an observation x from a $X \sim N(0,1)$ random variable variable, we will simply get x back. I.e. $z(x) = x$ because X is already $N(0,1)$. We can also go the other way, i.e. we can convert from a $Z \sim N(0,1)$ random variable to a $N(\mu, \sigma)$ random variable by computing $Z\sigma + \mu$.

Example 5.5. Using a Z-table

In figure 5.1, we show a small section of a Z-table which has been marked for clarity. The complete table is shown at the end of the text.

z	0.00	0.01	0.02	0.03	0.04
0.0	0.5000	0.5040	0.5080	0.5120	0.5160
0.1	0.5398	0.5438	0.5478	0.5517	0.5557
0.2	0.5793	0.5832	0.5871	0.5910	0.5948
0.3	0.6179	0.6217	0.6255	0.6293	0.6331
0.4	0.6554	0.6591	0.6628	0.6664	0.6700
0.5	0.6915	0.6950	0.6985	0.7019	0.7054

Figure 5.1: Using the z-table to look up $P(X < .52) = 69.85\%$

The yellow highlighted areas are indexes used to represent z. For instance, the first entry in the 6th row of the index is 0.5. This represents a Z-score of 0.5. To improve our precision, we need to input the second digit of the Z-score. This is the purpose of the columns. For instance, the 3rd colulmn in the index is .02, which represents the second digit of the Z-score. Thus, if we select the 6th row (.5) and the 3rd column (.02) (circled in red), the associated Z-score is 0.52. The value at the intersection of the selected row and column represents $P(X < z = 0.52) = P(X < 0.52) = 0.6985$. The area represented by $P(X < 0.52)$ is depicted in figure 5.2.

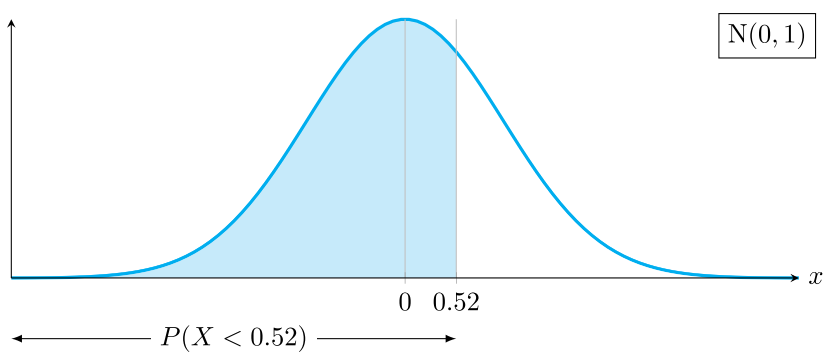


Figure 5.2: $P(X < 0.52) = 0.6985$

Exercise 5.6. Computing z-scores

Let $X \sim N(6, 2)$. Compute the z-scores and the associated probabilities $P(X < z)$ for each of the following values. Use a z-table and then check your answers using a calculator.

1. $x = 6.34$
2. $x = 3.78$
3. $x = 8$

5.4 p-values and Z-tests

Since quantifying uncertainty is a central concern of statistics, we often wish to know if our results are plausible. For instance, let X represent the average height of an adult male human. Suppose your friend claims that $X \sim \mathrm{N}(7, \frac{1}{2})$, where the units are feet. You discover that a $\frac{1}{2}$foot (or 6 inches) is the correct population standard deviation, but the mean seems incorrect. To test your hypothesis that the population $\mu \neq 7$, you set up a statistical experiment. You take a random sample of one person.[7] You find that this person has a height of 5.9 feet. How unlikley is observing such a value, given that the population mean μ really is 7? To answer this question, we introduce the idea of a p-value.

Definition 5.7. A ***p*-value** is the probability of observing a result as or more extreme than that of our sample given that the null hypothesis is true.

Z-test

A **Z-test** is

In our case the **null hypothesis** (H_0) is that $\mu = 7$. The **alternitive hypothesis** is that $\mu < 7$. When we say "probability of observing a result as or more extreme than that of our sample", we mean the probability of observing an x further away from μ (in both positive and negative distance) than our x. We measure this distance in terms of standard deviations above/below the mean. This is exactly what a Z-score tells us, and from the Z-score, we can calculate the relevant probabilities. Since the distance between 5.9 and μ is $|5.9 - \mu| = 1.1$, we need to find the probability that the distance between X and μ is greater than 1.1.

$$\begin{aligned}
P\left(|X - \mu| > 1.1\right) &= P\left(X < (\mu - 1.1)\right) + P\left(X > (\mu + 1.1)\right) \\
&= P\left(X < 5.9\right) + P\left(X > 8.1\right) \\
&= 2 \cdot P\left(X < 5.9\right)
\end{aligned}$$

where the last line is by the summetry of the normal distribution about μ. We can visualize this as below:

[7]Of course, taking a sample of size one isn't realistic, however it is easiest to start with the simplest case. Next we will consider a more realistic example.

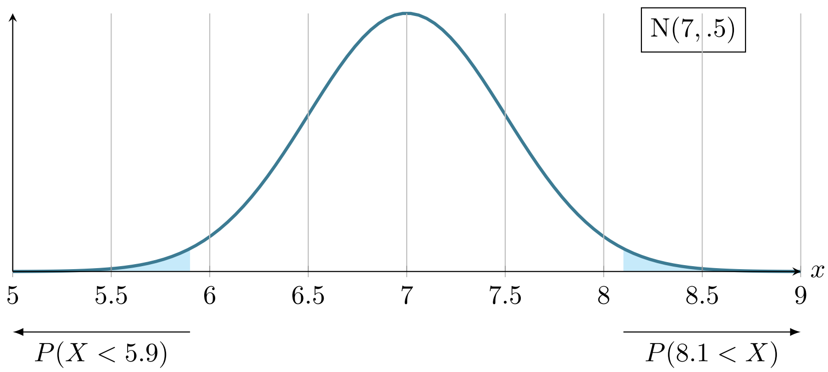


Figure 5.3: $P(|X - \mu| < 1.1) = 0.6985$

Thus to find the p-value, all we need to do is compute $2 \cdot P(X < 5.9)$. We do this using either a calculator or a z-table. Using a z-table, we have:

$$\begin{aligned} z(5.9) &= \frac{5.9 - 7}{.5} \\ &= \frac{-1.1}{.5} \\ &= -2.2 \end{aligned}$$

Using the z-table, we can find $P(Z < -2.2)$. If your z-table only has positive numbers, use the symmetry of the normal distribution and the fact that probabilities must add up to 1 to conclude that:

$$\begin{aligned} P(X < 5.9) &= P(Z < -2.2) \\ &= P(Z > 2.2) \\ &= 1 - P(Z > 2.2) \\ &= 1 - 0.9861 \\ &= .0139 \end{aligned}$$

Thus, the probability of observing this result given that the null hypothesis is true is

$$2 \cdot .0139 = 0.0278$$

which is a very unlikley result. Typically, in order to draw a conclusion, we compare our calculated p-value to some reference number α called the **significance level**. If our p-value is less than the significiance level, we conclude that our result is so improbable that we we cannot accept H_0 as truth. In this case, we must **reject the null hypothesis** H_0, and find that it cannot be the case that the average height is 7 feet tall.

6 Estimators and Estimates

Let us return to the problem of trying to identify the mean height of the adult male population based on a sample. To do this, we need to guage how likely a given sample is to occur given that it came from a given distribution, say $\mathrm{N}(\mu, \sigma)$. Instead of a sample of size 1 as before, let us consider the more realistic general case of a sample of size n. To estimate the true population mean, we might use the sample mean $\bar{x}$, which is an **estimator** of the population mean.

Definition 6.1. The **sample mean** $\bar{x}$ is an estimator of the population mean. If we take n observations $x_1, x_2, \ldots, x_n$ from some poplulation, the sample mean is simply the sum of the n observations divided by the number n of observations. That is:

$$\bar{x} = \frac{1}{n} \cdot (x_1 + x_2 + \cdots + x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

[8]

Definition 6.2. Observe that the sample mean is a function of random variables. We call such a function a **statistic**. Formally, if $X_1, X_2, \ldots, X_n$ are random variables, and T is some function, a statistic is given by:

$$T(X_1, X_2, \ldots, X_n)$$

Observe that since $X_1, X_2, \ldots, X_n$ are random variables which cannot be predicted with certainty, any statistic based on them is also subject to variability. This means that a statistic is also a random variable with its own mean, variance, distribution, etc. We often use statistics as estimators for population parameters which are unknown. For instance, if we took a sample of mens heights, we might use the sample mean $\bar{x}$ as an estimate for the true (but unknown) population mean μ.

Definition 6.3. Let $x_1, x_2, \ldots, x_n$ be n observations from some poplulation. Then the **sample variance** s^2 is a statistic given by:

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

This definition may look complicated, however, we can simplify it making the substitution $e_i = (x_i - \bar{x})^2$. We can interpret e_i as the squared deviation from the mean for observation i. Then the sample variance becomes:

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} e_i$$

Referring back to 6.1, we see that the sample variance s^2 is simply the average of the squared deviations.

[8]When we write $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, we are using *summation* or *sigma notation*. It is very useful for writing down large sums compactly. The meaning is as follows: $\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$

7 Sampling Distributions

When discussing the 6.2, we wrote that "a statistic is also a random variable with its own mean, variance, distribution, etc." In general the distribution of a static is known as the **sampling distiribution**. In this section, we will discuss the properties of sampling distributions, with a focus on the distribution of the sample mean $\bar{x}$.

Suppose that we were to take multiple samples of from a $N(0, 1)$ distribution and compute the sample means $\bar{x}_i$ of each sample. Because the sample mean is a function of random variables, which are of course random, we would expect the sample means to be different. Below we run this experiment with two samples of size 3 from a $N(0, 1)$ distribution. Note that each sample mean $\bar{x}_i$ is an observation from the sampling distrubition of $\bar{x}$.

	x_1	x_2	x_3	$\bar{x}$
Sample 1	-0.6	0.2	-0.8	-0.4
Sample 2	1.6	0.3	-0.8	.37

Table 8: Computing $\bar{x}$ from two random samples of size 3

Now suppose we were to take many samples from the same distribution and compute the sample mean for each of them. What might that look like? In other words, what does the sampling distribution of $\bar{x}$ look like? In figure 7.1, we visualize the sampling distribution of $\bar{x}$ via historgrams of the means of 10, 100, and 1000 samples (each same of size 10). E.g. for the 1000 samples histogram, we took 1000 samples, each of size 10, (so 10,000 observations total), and computed $\bar{x}$ for each of the 1000 samples. Observe that as the number of samples of $\bar{x}$ increased, the sampling distribution of $\bar{x}$ looks more and more normal. How can this be?

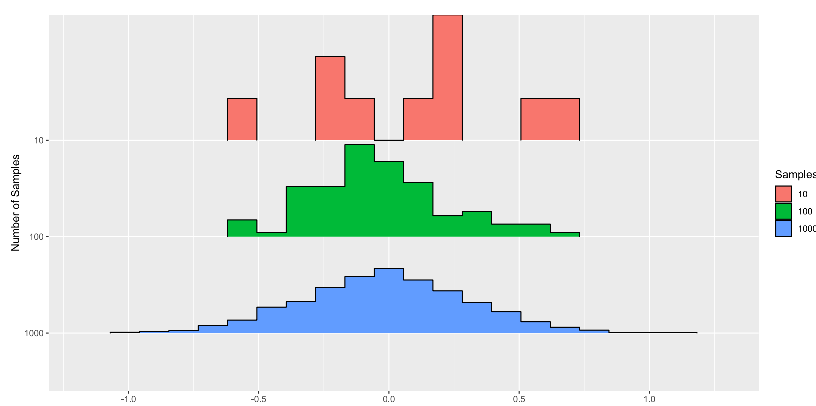


Figure 7.1: Distribution of $\bar{x}$ based on 10, 100, and 1000 samples

7.1 The Central Limit Theorem

The central limit theorem is an important and beautiful theorem in statistics. It tells us that as the number of samples n becomes large (i.e. as n grows towards infinity), the sampling distribution of $\bar{x}$ tends towards the normal distribution. Incredibly, this is the true even if the population distribution is not normal. In other words, no matter what the population distribution is, the sampling distribution of $\bar{x}$ is normal[9].

Theorem 7.1. *The Central Limit Theorem*
TODO

Knowing the distribution of the sample mean is good, but we should also want to know about the mean and variance as well.

Theorem 7.2. *The Mean of the Sampling Distribution of $\bar{x}$ is μ*

Proof. Let $X_1, X_2, \ldots, X_n$ be n i.i.d random variables. Then:

$$\begin{aligned}
\mathrm{E}\left[\bar{x}\right] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mu \\
&= \frac{n\mu}{n} \\
&= \mu
\end{aligned}$$

□

Theorem 7.3. *The Variance of the Sampling Distribution of $\bar{x}$ is $\frac{\sigma^2}{n}$*

Proof. Let $X_1, X_2, \ldots, X_n$ be n i.i.d random variables. Then:

[9] Provided that population distribution has finite variance

$$
\begin{aligned}
\operatorname{Var}[\bar{x}] &= \operatorname{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\operatorname{Var}\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 \\
&= \frac{1}{n^2} n\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

□

Definition 7.4. The standard deviation of the sampling distribution of a statistic T is called the **Standard Error** of T. If the statistic is $\bar{x}$, its standard deviation is called the **Standard Error of the Mean (SEM)** and written $\sigma_{\bar{x}}$.

Theorem 7.5. *The Standard Error of $\bar{x}$, written $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$*

Proof. Since we showed the variance of the sampling distribution of $\bar{x}$ is $\frac{\sigma^2}{n}$ in theorem 7.3, we can compute the standard deviation of the mean $\sigma_{\bar{x}}$ by taking square roots:

$$
\begin{aligned}
\operatorname{Var}[\bar{x}] &= \frac{\sigma^2}{n} \\
\sqrt{\operatorname{Var}[\bar{x}]} &= \sqrt{\frac{\sigma^2}{n}} \\
\sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}}
\end{aligned}
$$

□

8 Hypothesis Testing

Now that we know the sampling distribution of $\bar{x}$ is normal, as well its mean and standard deviation, we can answer questions involving $\bar{x}$ the same way as we did about a single observation x of some random variable. For example, in section §5, we studied questions like; if $X \sim N(3, 5)$, what is $P(X > 8)$? We saw that we could simply compute the Z-score, and consult the Z-table to find the correct probability. Now we are able to answer questions about $\bar{x}$ in the same way. The only difference is that the standard deiviation of the distribution of the mean is $\frac{\sigma}{\sqrt{n}}$, so we must use this as the standard deviation when calculating the mean. I.e, to get the Z-score, we compute:

$$z(x) = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

To reiterate, when we compute a Z-score, we are still converting to a $N(0, 1)$ distribution, i.e. we are still finding out the number of standard deivations above the mean $\bar{x}$ is. The only thing that has changed is that the standard deviation of the sampling distribution of $\bar{x}$ (i.e. the SEM) is given by $\frac{\sigma}{\sqrt{n}}$.

Example 8.1. Suppose that the lifespan of a human has some distribution with $\mu = 80, \sigma = 10$, with units in years. A random sample of 144 death cirtificates are examined. Calculate:

- The standard error of the mean
  - Since $\sigma = 10$, we have that the SEM is $\frac{10}{\sqrt{144}} = \frac{10}{12}$
- The probability of living to be greater than 100 years old.
  - This is $P(X > 100)$ in a $N(80, 10)$ RV.
- The probability of living to be less than 70 or more than 100 years old.
  - This is $P(X > 100) \cup (X < 70)$ in a $N(80, 10)$ RV.
- You take a random sample of 144 death cirtificates. You find $\bar{x} = 78$. What is the probability of observing this result given that $\mu = 80, \sigma = 10$? Is the assumption that $\mu = 80, \sigma = 10$ credible?
  - compute the Z-score as $z = \frac{78-80}{10/12} = -2.4$. The one sided p-value is .0082 or .82%. This is a unlikley event, which makes us reject the null hypothesis that $\mu = 80$.

8.1 The T-test

In previous examples, we usually assumed that the population standard deviation was known. However, this is unrealistic. In practice, we often need to use an estimate for the population standard deviation, such as the sample standard deviation s. We then calculate the **T-statistic** as:

$$T(x) = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Note the similarity to the caluation of a Z-score. However, it can be shown that $T(x)$ has a **T-distribution** and hence to evaluate the probabilities of events involving a T-statistic, we must use a T-table, which is simple the T-distribution analogue of a Z-table. In practice, the T-distribution looks similar to the normal, except that it assigns more higher probabilites to events further away from the mean. In other words, under a T-distribution, we are more likley to observe results further away from the mean. For this reason, we say that it is a **fat-tailed** distribution.

9 Confidence Intervals

When we estimate a population parameter like μ with a statistic like $\bar{x}$, there is always inherent uncertanty in our estimate. One way of quantifying our uncertainty is to provide a region of values in which the true parameter is likley to be. Such a region is called a confidence interval.

Definition 9.1. Let θ be an unknown population parameter, and T be an estimate of θ. We say the interval (a, b) is a $(1-\alpha)\,\%$ **confidence interval** for θ if

$$1-\alpha = P\left(\theta \in (a, b)\right)$$

I.e., the probability that the interval (a, b) includes the true parameter θ is $1-\alpha$. The idea is that if we were to repeatedly create say M independent confidence intervals, then we would expect $(1-\alpha)\cdot M$ of them to actually contain θ. Thus we are $(1-\alpha)100$ confident that the true value of θ lies in the interval (α, β). We refer to α as the **significance level** of the confidence interval and interpret it as the probability of committing a **Type 1 Error**, i.e. the probability of mistakenly rejecting the null hypothesis when it is true.

So how does one actually construct a confidence interval for some statistic T? Does it depend on the distribution of T, or the value of α? These are good questions, and it turns out that there is a general formula. However, for simplicity we will begin with a common special case; a confidence interval for the population mean μ.

9.0.1 Confidence Intervals for μ

Our confidence interval for μ is centered on our estimate $\bar{x}$ for μ. This should make sense since $\bar{x}$ is our "best guess" for μ. Of course, it is unlikley that $\mu = \bar{x}$, so we need an interval around $\bar{x}$ to capture μ more reliably. [10]

Theorem 9.2. *Confidence Interval Formula*

Let T be an estimate of θ. Then an $(1-\alpha)\,\%$ confidence interval for θ is given by:

$$T\pm$$

[10] In fact, if μ is the mean of a contiuous distribution, $P\left(\mu = \bar{x}\right) = 0$.

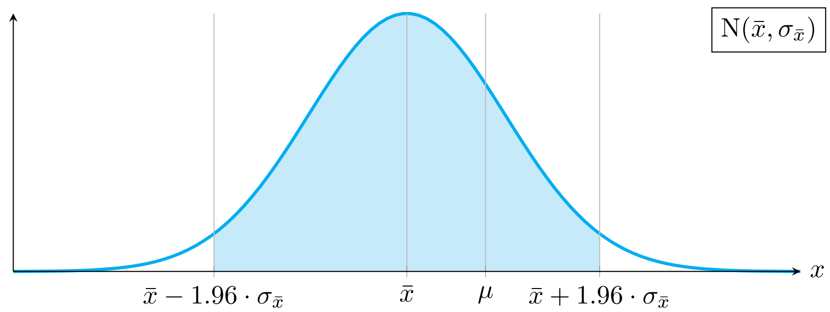


Figure 9.1: A 95% Confidence Interval for μ

NORMAL CUMULATIVE DISTRIBUTION FUNCTION

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000