



MAKING SENSE OF DATA

Essentials series

THE NORMAL DISTRIBUTION

Copyright © 2012 by City of Bradford MDC

Prerequisites
Descriptive statistics
Charts and graphs
The normal distribution
Surveys and sampling
Correlation and regression

There are two major types of random variables: discrete and continuous. *Discrete* random variables basically count things (number of heads on 10 coin flips, and so on). The most well-known discrete random variable is the binomial. A *continuous* random variable measures things and takes on values within an interval, or they have so many possible values that they are deemed continuous (for example, exam scores).

Binomial probabilities

Binomial means “two names” and is associated with situations involving two outcomes: success or failure (for example, hitting a red light or not).

Characteristics of a binomial

A random variable has a binomial distribution if all of following conditions are met:

- 1) There are a fixed number of trials (n).
- 2) Each trial has two possible outcomes: success or failure.
- 3) The probability of success (p) is the same for each trial.
- 4) The trials are independent, meaning the outcome of one trial doesn't influence that of any other.

Let X equal the total number of successes in n trials. If all of the above conditions are met, X has a binomial distribution with probability of success equal to p .

Finding binomial probabilities

After confirming that X has a binomial distribution (the four conditions are met), probabilities can be calculated using the values of n and p unique to each problem.

Probabilities for a binomial random variable X can be found using this formula:

$$\binom{n}{x} p^x (1 - p)^{n-x}$$

Where

- n = the fixed number of trials
- x = the specified number of successes
- $n-x$ = the number of failures
- p = the probability of success on any given trial
- $1 - p$ = the probability of failure on any given trial

These probabilities hold for any value of X between 0 (lowest number of possible successes in n trials) and n (highest number of possible successes).



The number of ways to arrange x successes among n trials is called “ n choose x ” and the notation is $\binom{n}{x}$. For example, $\binom{3}{2}$ means “3 choose 2” and stands for the number of ways to get 2 successes in 3 trials. To calculate “ n choose x ” use the formula

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

The notation $n!$ stands for n -factorial, the number of ways to rearrange n items. To calculate $n!$ multiply $n(n-1)(n-2)\dots(2)(1)$. For example $3!$ is $3(2)(1)=6$; $2!$ is $2(1)=2$; and $1!=1$. By convention, $0!=1$.

To calculate “3 choose 2”

$$\binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{3 \times 2 \times 1}{(2 \times 1)(1!)} = \frac{6}{2 \times 1} = 3$$

Suppose a car crosses three traffic lights on the way to work, and the probability of each of them being red is 0.30. (Assume the lights are independent.) Let X be the number of red lights encountered.

We know p (probability of a red light) = 0.30; $1 - p$ (probability of a non-red light) = $1 - 0.30 = 0.70$; and the number of non-red lights is $3 - X$.

Question: What is the probability of the car hitting no red lights, and all three red lights?

Using the formula to obtain the probabilities P for $X=0$ and 3 red lights

$$\begin{aligned}P(X = 0) &= \binom{3}{0} 0.30^0 (1 - 0.30)^{3-0} \\&= \frac{3!}{0! (3 - 0)!} (0.30)^0 (0.70)^3 \\&= 1(1)(0.343) = 0.343\end{aligned}$$

$$\begin{aligned}P(X = 3) &= \binom{3}{3} 0.30^3 (1 - 0.30)^{3-3} \\&= \frac{3!}{3! (3 - 3)!} (0.30)^3 (0.70)^0 \\&= 1(0.027)(1) = 0.027\end{aligned}$$

Answer: The probability of the car hitting no red lights is 34.3%, and the chance of hitting all three red lights is 2.7%.

Expected value and variance of the binomial

The mean of a random variable is the long-term average of its possible values over the entire population of individuals (or trials). It's found by taking the weighted average of the x -values multiplied by their probabilities. The mean of a random variable is denoted by μ . For the binomial random variable the mean is $\mu = np$.

Suppose you flip a fair coin 100 times and let X be the number of heads; this is a binomial random variable with $n = 100$ and $p = 0.50$. Its mean is $np = 100(0.50) = 50$.

The variance of a random variable X is the weighted average of the squared deviations (distances) from the mean. The variance of a random variable is denoted by σ^2 . The variance of the binomial distribution is $\sigma^2 = np(1 - p)$. The standard deviation of X is just the square root of the variance, which in this case is $\sigma = \sqrt{np(1 - p)}$.

Suppose you flip a fair coin 100 times and let X be the number of heads. The variance of X is $(1 - p) = 100(0.50)(1 - 0.50) = 25$, and the standard deviation is the square root, which is 5.

Basics of the normal distribution

A random variable X has a normal distribution if its values fall into a smooth (continuous) curve with a bell-shaped, symmetric pattern, meaning it looks the same on each side when cut down the middle. The total area under the curve is 1. Each normal distribution has its own mean, μ , and its own standard deviation, σ .

Figure 1 illustrates three different normal distributions with different means and standard deviations.

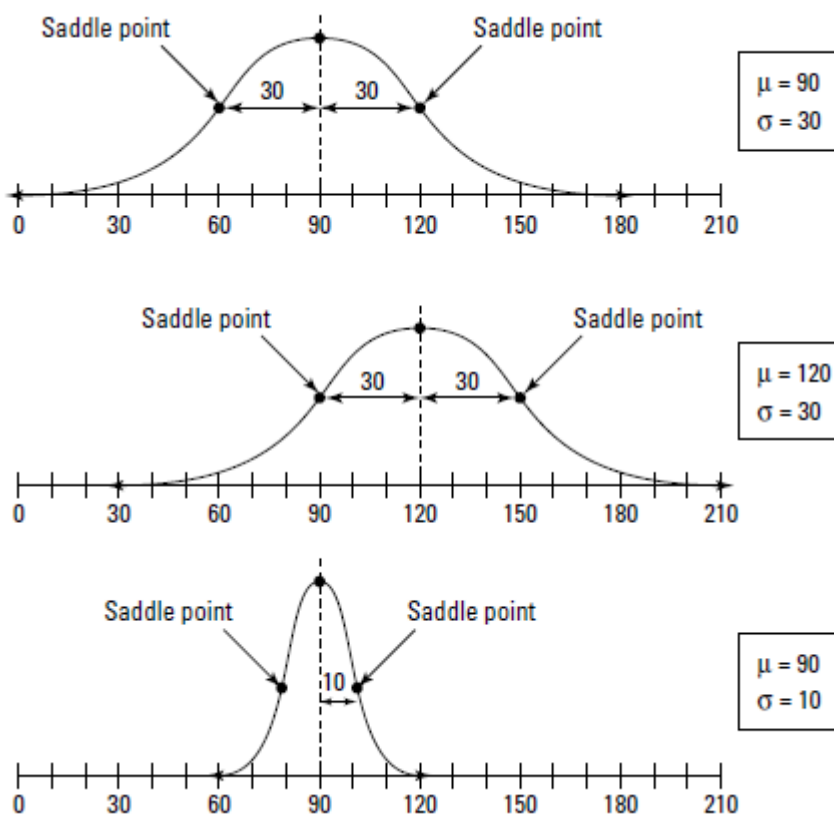


Figure 1: Three normal distributions

Note that the *saddle points* on each graph are where the graph changes from concave down to concave up. The distance from the mean out to either saddle point is equal to the standard deviation for the normal distribution. For any normal distribution, almost all its values lie within three standard deviations of the mean.

The standard normal (Z) distribution

One very special member of the normal distribution family is called the standard normal distribution, or *Z*-distribution. The *Z*-distribution is used to help find probabilities and solve other types of problems when working with any normal distribution.

The standard normal (*Z*) distribution has a mean of zero and a standard deviation of 1; its graph is shown in Figure 2. A value on the *Z*-distribution represents the number of standard deviations the data is above or below the mean; these are called *z*-scores (or *z*-values). For example, $z = 1$ on the *Z*-distribution represents a value that is 1 standard deviation above the mean.

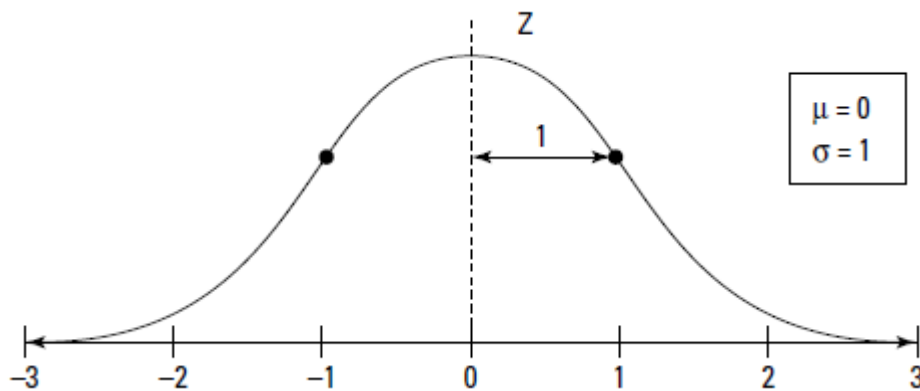


Figure 2: The *Z*-distribution has a mean of 0 and standard deviation of 1

Because probabilities for any normal distribution are difficult to calculate by hand, tables are used to find them. All the basic results to find probabilities for any normal distribution can be presented in one table based on the standard normal (*Z*) distribution. This table is called the *Z*-table and is found at the end of this topic. All you need is one formula to transform your normal distribution (*X*) to the standard normal (*Z*) distribution, and you can use the *Z*-table to find the probability needed.

The general formula for changing a value of X into a value of Z is

$$Z = \frac{X - \mu}{\sigma}$$

Take your x -value, subtract the mean, and divide by the standard deviation; this gives you its corresponding z -value.

For example, if the random variable X is a normal distribution with mean 16 and standard deviation 4, the value 20 on the X -distribution would transform into $20 - 16$ divided by 4, which equals 1. So, the value 20 on the X -distribution corresponds to the value 1 on the Z -distribution.

To use the Z -table at the end of this topic to find probabilities, do the following:

1. Go to the row that represents the leading digit of your z -value and the first digit after the decimal point.
2. Go to the column that represents the second digit after the decimal point of your z -value.
3. Intersect the row and column.

That number represents $P(Z < z)$.

For example, suppose we want to look at $P(Z < 2.13)$. Using the Z -table, find the row for 2.1 and the column for 0.03. Put 2.1 and 0.03 together as one three-digit number to get 2.13. Intersect that row and column to find the number: 0.9834. Therefore $P(Z < 2.13) = 0.9834$.

Finding probabilities for X



Here are the steps for finding a probability for X :

1. Draw a picture of the distribution.
2. Translate the problem into one of the following: $P(X < a)$, $P(X > b)$, $P(a < X < b)$. Shade in the area on your picture.
3. Transform a (and/or b) into a z -value, using the Z -formula:
$$Z = \frac{X - \mu}{\sigma}$$
4. Look up the transformed z -value on the Z -table and find its probability.
5. If you have
 - a. a less-than problem, you're done.
 - b. a greater-than problem, take one minus the result from Step 4.
 - c. a between-values problem, do Steps 1–4 for b (the larger of the two values) and then for a (the smaller of the two values), and subtract the results.

Suppose, for example, that you enter a fishing contest. The contest takes place in a pond where the fish lengths have a normal distribution with mean = 16 inches and standard deviation = 4 inches.

Problem 1: What's the chance of catching a small fish — say, less than 8 inches?

Problem 2: Suppose a prize is offered for any fish over 24 inches. What's the chance of catching a fish at least that size?

Problem 3: What's the chance of catching a fish between 16 and 24 inches?

To solve these problems, first draw a picture of the distribution. Figure 3 shows a picture of X 's distribution for fish lengths. You can see where each of the fish lengths mentioned in each of the three fish problems falls.

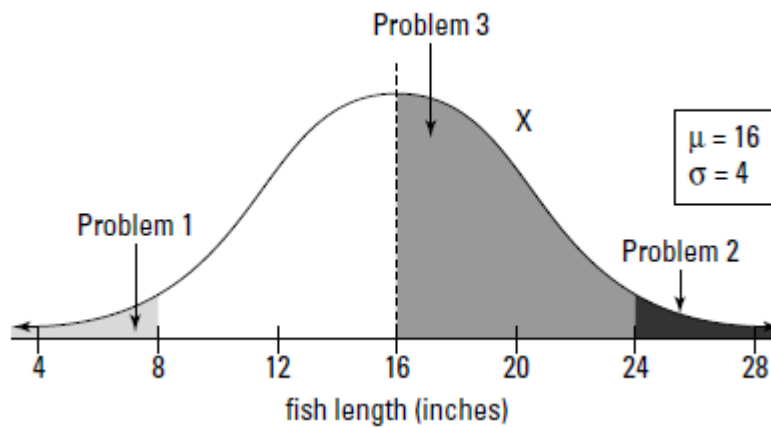


Figure 3: The distribution of fish lengths in a pond

Next, translate each problem into probability notation. Problem 1 means find $P(X < 8)$. For Problem 2, you want $P(X > 24)$. And Problem 3 is asking for $P(16 < X < 24)$.

Step 3 says change the x-values to z-values using the Z-formula, $Z = \frac{x - \mu}{\sigma}$. For Problem 1, you have $P(X < 8) = P\left(Z < \frac{8 - 16}{4}\right) = P(Z < -2)$. Similarly for Problem 2, $P(X > 24)$ becomes $P(Z > 2)$. Problem 3 translates from $P(16 < X < 24)$ to $P(0 < Z < 2)$. Figure 4 shows a comparison of the X-distribution and Z-distribution for the values $x = 8, 16,$ and 24 , which transform into $z = -2, 0,$ and $+2$, respectively.

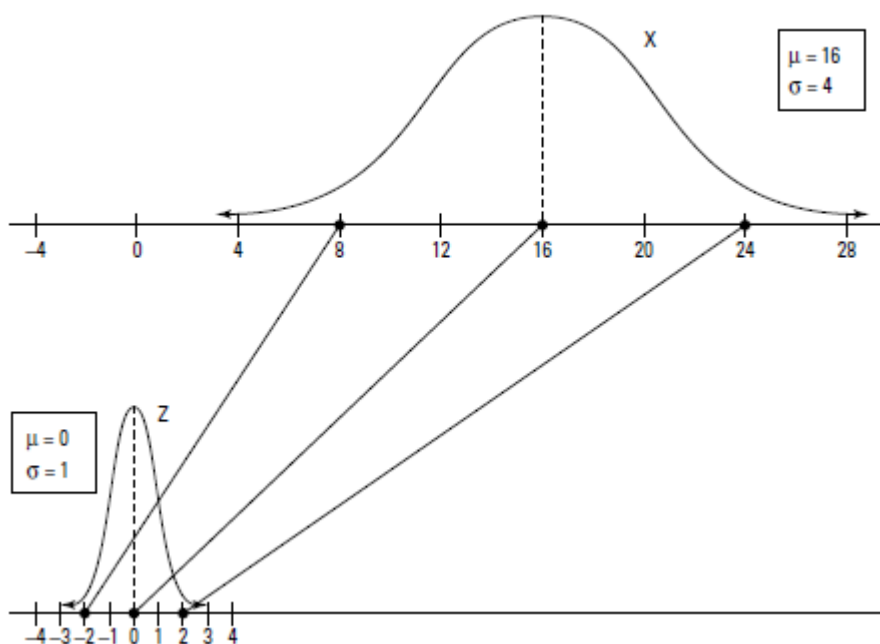


Figure 4: Transforming numbers on the normal distribution to numbers on the Z-distribution

Now that you have changed x -values to z -values, you move to Step 4 and find probabilities for those z -values using the Z -table (at the end of this topic). In Problem 1 of the fish example, you want $P(Z < -2)$; go to the Z -table and look at the row for -2.0 and the column for 0.00 , intersect them, and you find 0.0228 — according to Step 5a you're done. So, the chance of a fish being less than 8 inches is equal to 0.0228 .

For Problem 2, find $P(Z > 2.00)$. Because it's a "greater-than" problem, this calls for Step 5b. To be able to use the Z -table you need to rewrite this in terms of a "less-than" statement. Because the entire probability for the Z -distribution equals 1, we know $P(Z > 2.00) = 1 - P(Z < 2.00) = 1 - 0.9772 = 0.0228$. So, the chance that a fish is greater than 24 inches is 0.0228 . (Note the answers to Problems 1 and 2 are the same because the Z -distribution is symmetric; see Figure 3.)

In Problem 3, you find $P(0 < Z < 2.00)$; this requires Step 5c. First find $P(Z < 2.00)$, which is 0.9772 from the Z -table, and then subtract off the part you don't want, which is $P(Z < 0) = 0.500$ from the Z -table. This gives you $0.9772 - 0.500 = 0.4772$. So the chance of a fish being between 16 and 24 inches is 0.4772 .

Normal approximation to the binomial

Suppose you flip a fair coin 100 times, and we let X equal the number of heads. What is the probability that X is greater than 60?

If n (number of trials) is large enough, then the skew of the distribution is not too great. In this case you can use the normal distribution to get an approximate answer. To determine whether n is large enough to use the normal approximation, two conditions must hold:

1. $(n \times p) \geq 10$
2. $n \times (1 - p) \geq 10$

In general, follow these steps to find the approximate probability for a binomial distribution when n is large:

1. **Verify whether n is large enough to use the normal approximation by checking the two conditions.**

For the coin-flipping question, the conditions are met since $n * p = 100 * 0.50 = 50$, and $n * (1 - p) = 100 * (1 - 0.50) = 50$, both of which are at least 10.

2. **Write down what you need to find as a probability statement about X .**

For the coin-flipping example, find $P(X > 60)$.

3. **Transform the x -value to a z -value, using the Z -formula, $Z = \frac{X - \mu}{\sigma}$.**

For the mean of the normal distribution, use $\mu = n * p$ (the mean of the binomial), and for the standard deviation σ , use $\sqrt{np(1 - p)}$ (the standard deviation of the binomial).

For the coin-flipping example, use $\mu = n * p = 100 * 0.50 = 50$ and $\sigma = \sqrt{np(1 - p)} = \sqrt{100 * 0.50(1 - 0.50)} = 5$.

Now put these values into the Z -formula to get $Z = (60 - 50) / 5 = 2$. Find the $P(Z > 2)$.

4. **Proceed as you would for any normal distribution. That is, do Steps 4 and 5 described in the earlier section "Finding Probabilities for X ".**

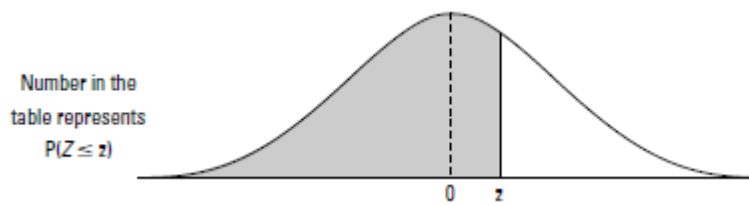
For the coin flips, $P(X > 60) = P(Z > 2.00) = 1 - 0.9772 = 0.0228$.

The chance of getting more than 60 heads in 100 flips of a coin is about 2.28 percent.

The Z-table



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247



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	0.7704	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