9.1: Central Limit Theorem for Bernoulli Trials

The second fundamental theorem of probability is the *Central Limit Theorem.* This theorem says that if S_n is the sum of n mutually independent random variables, then the distribution function of S_n is well-approximated by a certain type of continuous function known as a normal density function, which is given by the formula

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} , \tag{9.1.1}$$

as we have seen in Chapter 5. In this section, we will deal only with the case that $\mu = 0$ and $\sigma = 1$. We will call this particular normal density function the standard normal density, and we will denote it by $\phi(x)$:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} . \tag{9.1.2}$$

A graph of this function is given in Figure [fig 9.1.1]. It can be shown that the area under any normal density equals 1.

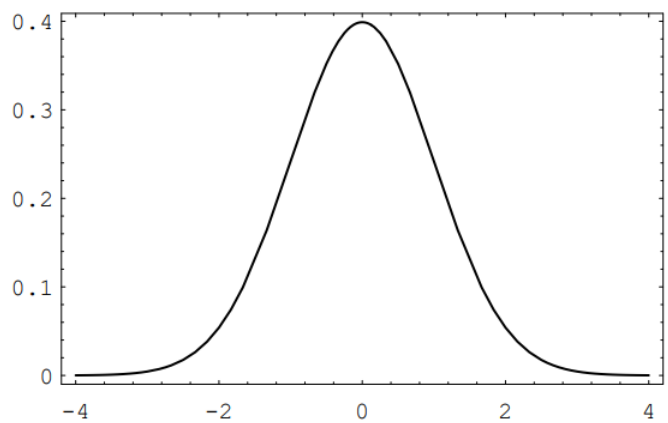


Figure 9.1.1: Standard normal density

The Central Limit Theorem tells us, quite generally, what happens when we have the sum of a large number of independent random variables each of which contributes a small amount to the total. In this section we shall discuss this theorem as it applies to the Bernoulli trials and in Section 1.2 we shall consider more general processes. We will discuss the theorem in the case that the individual random variables are identically distributed, but the theorem is true, under certain conditions, even if the individual random variables have different distributions.

Bernoulli Trials

Consider a Bernoulli trials process with probability p for success on each trial. Let $X_i = 1$ or 0 according as the ith outcome is a success or failure, and let $S_n = X_1 + X_2 + \cdots + X_n$. Then S_n is the number of successes in n trials. We know that S_n has as its distribution the binomial probabilities $b(n, p, j)$. In Section 3.2, we plotted these distributions for $p = .3$ and $p = .5$ for various values of n (see Figure [fig 3.8]).

We note that the maximum values of the distributions appeared near the expected value np, which causes their spike graphs to drift off to the right as n increased. Moreover, these maximum values approach 0 as n increased, which causes the spike graphs to flatten out.

Standardized Sums

We can prevent the drifting of these spike graphs by subtracting the expected number of successes np from S_n, obtaining the new random variable $S_n - np$. Now the maximum values of the distributions will always be near 0.

To prevent the spreading of these spike graphs, we can normalize $S_n - np$ to have variance 1 by dividing by its standard deviation $\sqrt{npq}$ (see Exercise 6.2.13 and Exercise 6.2.16

Definition: Term

The of S_n is given by

$$S_n^* = \frac{S_n - np}{\sqrt{npq}} \ . \tag{9.1.3}$$

S_n^* always has expected value 0 and variance 1.

Suppose we plot a spike graph with the spikes placed at the possible values of S_n^*: x_0, x_1, …, x_n, where

$$x_j = \frac{j - np}{\sqrt{npq}} \tag{9.1.4}$$

We make the height of the spike at x_j equal to the distribution value $b(n, p, j)$. An example of this standardized spike graph, with $n = 270$ and $p = .3$, is shown in Figure $9.1.2$. This graph is beautifully bell-shaped. We would like to fit a normal density to this spike graph. The obvious choice to try is the standard normal density, since it is centered at 0, just as the standardized spike graph is. In this figure, we have drawn this standard normal density. The reader will note that a horrible thing has occurred: Even though the shapes of the two graphs are the same, the heights are quite different.

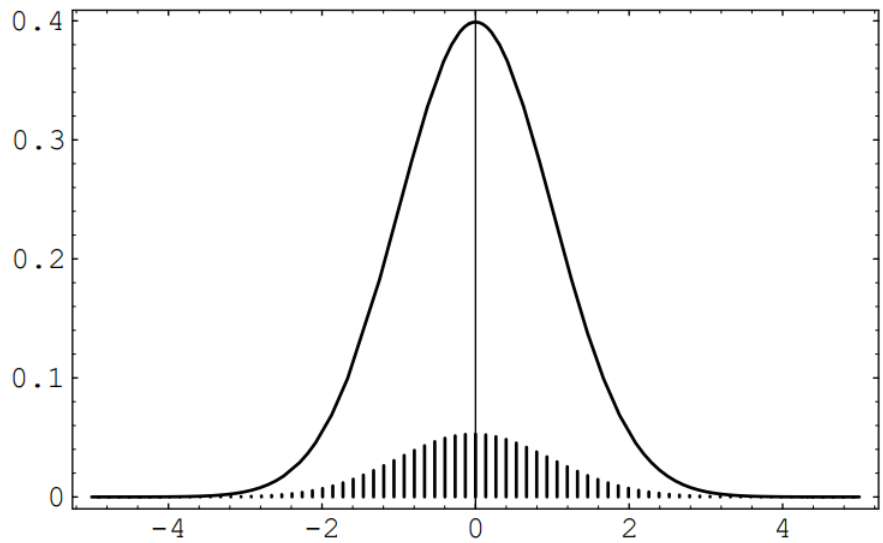


Figure $9.1.2$: Normalized binomial distribution and standard normal density

If we want the two graphs to fit each other, we must modify one of them; we choose to modify the spike graph. Since the shapes of the two graphs look fairly close, we will attempt to modify the spike graph without changing its shape. The reason for the differing heights is that the sum of the heights of the spikes equals 1, while the area under the standard normal density equals 1. If we were to draw a continuous curve through the top of the spikes, and find the area under this curve, we see that we would obtain, approximately, the sum of the heights of the spikes multiplied by the distance between consecutive spikes, which we will call ϵ. Since the sum of the heights of the spikes equals one, the area under this curve would be approximately ϵ. Thus, to change the spike graph so that the area under this curve has value 1, we need only multiply the heights of the spikes by $1/\epsilon$. It is easy to see from Equation $9.1.1$ that

$$\epsilon = \frac{1}{\sqrt{npq}} \ . \tag{9.1.5}$$

In Figure $9.1.2$ we show the standardized sum S_n^* for $n = 270$ and $p = .3$, after correcting the heights, together with the standard normal density. (This figure was produced with the program **CLTBernoulliPlot**.) The reader will note that the standard normal fits the height-corrected spike graph extremely well. In fact, one version of the Central Limit Theorem (see Theorem $9.1.1$) says that as n increases, the standard normal density will do an increasingly better job of approximating the height-corrected spike graphs corresponding to a Bernoulli trials process with n summands.

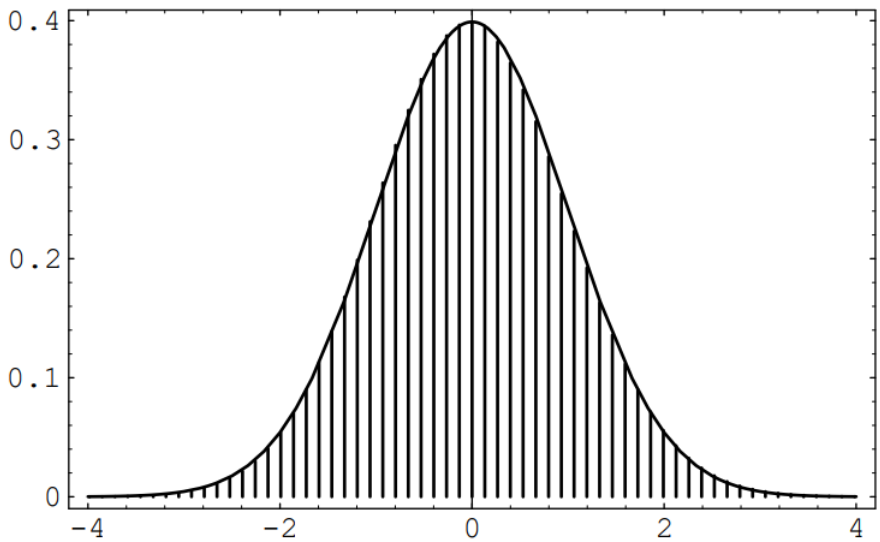


Figure 9.1.3: Corrected spike graph with standard normal density

Let us fix a value x on the x-axis and let n be a fixed positive integer. Then, using Equation [eq 9.1.1:], the point x_j that is closest to x has a subscript j given by the formula

$$j = \langle np + x\sqrt{npq}\rangle \ , \tag{9.1.6}$$

where $\langle a\rangle$ means the integer nearest to a. Thus the height of the spike above x_j will be

$$\sqrt{npq}\, b(n,p,j) = \sqrt{npq}\, b(n,p,\langle np + x_j\sqrt{npq}\rangle) \ . \tag{9.1.7}$$

For large n, we have seen that the height of the spike is very close to the height of the normal density at x. This suggests the following theorem.

Theorem 9.1.1

(Central Limit Theorem for Binomial Distributions) *For the binomial distribution $b(n,p,j)$ we have*

$$\lim_{n\to\infty} \sqrt{npq}\, b(n,p,\langle np + x\sqrt{npq}\rangle) = \phi(x) \ , \tag{9.1.8}$$

where $\phi(x)$ is the standard normal density.

Proof

The proof of this theorem can be carried out using Stirling's approximation from Section 3.1. We indicate this method of proof by considering the case $x = 0$. In this case, the theorem states that

$$\lim_{n\to\infty} \sqrt{npq}\, b(n,p,\langle np\rangle) = \frac{1}{\sqrt{2\pi}} = .3989\ldots \ . \tag{9.1.9}$$

In order to simplify the calculation, we assume that np is an integer, so that $\langle np\rangle = np$. Then

$$\sqrt{npq}\, b(n,p,np) = \sqrt{npq}\, p^{np} q^{nq} \frac{n!}{(np)!\,(nq)!} \ . \tag{9.1.10}$$

Recall that Stirling's formula (see Theorem 3.3) states that

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n} \qquad \text{as } \backslash,\backslash,\backslash, n \to \infty \ . \tag{9.1.11}$$

Using this, we have

$$\sqrt{npq}\, b(n,p,np) \sim \frac{\sqrt{npq}\, p^{np} q^{nq} \sqrt{2\pi n}\, n^n e^{-n}}{\sqrt{2\pi np}\sqrt{2\pi nq}\,(np)^{np}(nq)^{nq} e^{-np} e^{-nq}} \ , \tag{9.1.12}$$

which simplifies to $1/\sqrt{2\pi}$.

Approximating Binomial Distributions

We can use Theorem 9.1.1 to find approximations for the values of binomial distribution functions. If we wish to find an approximation for $b(n,p,j)$, we set

$$j = np + x\sqrt{npq} \tag{9.1.13}$$

and solve for x, obtaining

$$x = \frac{j-np}{\sqrt{npq}}\ . \tag{9.1.14}$$

Theorem 9.1.1 then says that

$$\sqrt{npq}, b(n,p,j) \tag{9.1.15}$$

is approximately equal to $\phi(x)$, so

$$b(n,p,j) \approx \frac{\phi(x)}{\sqrt{npq}} \tag{9.1.16}$$

$$= \frac{1}{\sqrt{npq}}\phi\left(\frac{j-np}{\sqrt{npq}}\right) \tag{9.1.17}$$

Example 9.1.1

Let us estimate the probability of exactly 55 heads in 100 tosses of a coin. For this case $np = 100\cdot 1/2 = 50$ and $\sqrt{npq} = \sqrt{100\cdot 1/2\cdot 1/2} = 5$. Thus $x_{55} = (55-50)/5 = 1$ and

$$P(S_{100} = 55) \sim \frac{\phi(1)}{5} = \frac{1}{5}\left(\frac{1}{\sqrt{2\pi}}e^{-1/2}\right) \tag{9.1.18}$$

$$= .0484 \tag{9.1.19}$$

To four decimal places, the actual value is .0485, and so the approximation is very good.

The program **CLTBernoulliLocal** illustrates this approximation for any choice of n, p, and j. We have run this program for two examples. The first is the probability of exactly 50 heads in 100 tosses of a coin; the estimate is .0798, while the actual value, to four decimal places, is .0796. The second example is the probability of exactly eight sixes in 36 rolls of a die; here the estimate is .1093, while the actual value, to four decimal places, is .1196.

The individual binomial probabilities tend to 0 as n tends to infinity. In most applications we are not interested in the probability that a specific outcome occurs, but rather in the probability that the outcome lies in a given interval, say the interval $[a,b]$. In order to find this probability, we add the heights of the spike graphs for values of j between a and b. This is the same as asking for the probability that the standardized sum S_n^* lies between a^* and b^*, where a^* and b^* are the standardized values of a and b. But as n tends to infinity the sum of these areas could be expected to approach the area under the standard normal density between a^* and b^*. The states that this does indeed happen.

Theorem 9.1.2

Central Limit Theorem for Bernoulli Trials) Let S_n be the number of successes in n Bernoulli trials with probability p for success, and let a and b be two fixed real numbers. Then

$$\lim_{n\to\infty} P\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) = \int_a^b \phi(x)\,dx\ . \tag{9.1.20}$$

Proof

This theorem can be proved by adding together the approximations to $b(n,p,k)$ given in Theorem 9.1.1.

We know from calculus that the integral on the right side of this equation is equal to the area under the graph of the standard normal density $\phi(x)$ between a and b. We denote this area by $NA(a^*, b^*)$. Unfortunately, there is no simple way to integrate the function $e^{-x^2/2}$, and so we must either use a table of values or else a numerical integration program. (See Figure [tabl 9.1] for values of $\backslash \mathrm{NA}(0, z)$. A more extensive table is given in Appendix A.)

It is clear from the symmetry of the standard normal density that areas such as that between -2 and 3 can be found from this table by adding the area from 0 to 2 (same as that from -2 to 0) to the area from 0 to 3.

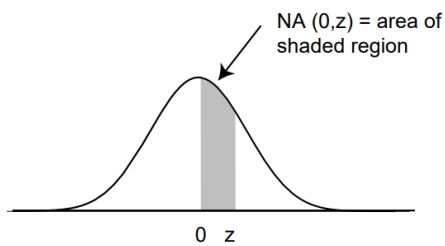
Approximation of Binomial Probabilities

Suppose that S_n is binomially distributed with parameters n and p. We have seen that the above theorem shows how to estimate a probability of the form

$$P(i \leq S_n \leq j)\,, \tag{9.1.21}$$

where i and j are integers between 0 and n. As we have seen, the binomial distribution can be represented as a spike graph, with spikes at the integers between 0 and n, and with the height of the kth spike given by $b(n, p, k)$. For moderate-sized values of n, if we standardize this spike graph, and change the heights of its spikes, in the manner described above, the sum of the heights of the spikes is approximated by the area under the standard normal density between i^* and j^*.

Table $9.1.1$ of values of NA(0, z), the normal area from 0 to z



Z	NA(z)	z	NA(z)	z	NA(z)	z	NA(z)
.0	.0000	1.0	.3413	2.0	.4772	3.0	.4987
.1	.0398	1.1	.3643	2.1	.4821	3.1	.4990
.2	.0793	1.2	.3849	2.2	.4861	3.2	.4993
.3	.1179	1.3	.4032	2.3	.4893	3.3	.4995
.4	.1554	1.4	.4192	2.4	.4918	3.4	.4997
.5	.1915	1.5	.4332	2.5	.4938	3.5	.4998
.6	.2257	1.6	.4452	2.6	.4953	3.6	.4998
.7	.2580	1.7	.4554	2.7	.4965	3.7	.4999
.8	.2881	1.8	.4641	2.8	.4974	3.8	.4999
.9	.3159	1.9	.4713	2.9	.4981	3.9	.5000

It turns out that a slightly more accurate approximation is afforded by the area under the standard normal density between the standardized values corresponding to $(i-1/2)$ and $(j+1/2)$; these values are

$$i^* = \frac{i - 1/2 - np}{\sqrt{npq}} \tag{9.1.22}$$

and

$$j^* = \frac{j+1/2-np}{\sqrt{npq}} \ . \tag{9.1.23}$$

Thus,

$$P(i \le S_n \le j) \approx \backslash\mathrm{NA}\left(\frac{i-\frac{1}{2}-np}{\sqrt{npq}}, \frac{j+\frac{1}{2}-np}{\sqrt{npq}}\right) \ . \tag{9.1.24}$$

It should be stressed that the approximations obtained by using the Central Limit Theorem are only approximations, and sometimes they are not very close to the actual values (see Exercise 9.2.111).

We now illustrate this idea with some examples.

Example 9.1.2

A coin is tossed 100 times. Estimate the probability that the number of heads lies between 40 and 60 (the word "between" in mathematics means inclusive of the endpoints). The expected number of heads is $100 \cdot 1/2 = 50$, and the standard deviation for the number of heads is $\sqrt{100 \cdot 1/2 \cdot 1/2} = 5$. Thus, since $n = 100$ is reasonably large, we have

$$\begin{aligned} P(40 \le S_n \le 60) \approx & \quad P\left(\frac{39.5-50}{5} \le S_n^* \le \frac{60.5-50}{5}\right) \\ = & \quad P(-2.1 \le S_n^* \le 2.1) \\ \approx & \quad \backslash\mathrm{NA}(-2.1, 2.1) \\ = & \quad 2\backslash\mathrm{NA}(0, 2.1) \\ \approx & \quad .9642 \ . \end{aligned}$$

The actual value is .96480, to five decimal places.

Note that in this case we are asking for the probability that the outcome will not deviate by more than two standard deviations from the expected value. Had we asked for the probability that the number of successes is between 35 and 65, this would have represented three standard deviations from the mean, and, using our 1/2 correction, our estimate would be the area under the standard normal curve between -3.1 and 3.1, or $2\backslash\mathrm{NA}(0, 3.1) = .9980$. The actual answer in this case, to five places, is .99821.

It is important to work a few problems by hand to understand the conversion from a given inequality to an inequality relating to the standardized variable. After this, one can then use a computer program that carries out this conversion, including the 1/2 correction. The program **CLTBernoulliGlobal** is such a program for estimating probabilities of the form $P(a \le S_n \le b)$.

Example 9.1.3

Dartmouth College would like to have 1050 freshmen. This college cannot accommodate more than 1060. Assume that each applicant accepts with probability .6 and that the acceptances can be modeled by Bernoulli trials. If the college accepts 1700, what is the probability that it will have too many acceptances?

If it accepts 1700 students, the expected number of students who matriculate is $.6 \cdot 1700 = 1020$. The standard deviation for the number that accept is $\sqrt{1700 \cdot .6 \cdot .4} \approx 20$. Thus we want to estimate the probability

$$\begin{aligned} P(S_{1700} > 1060) = & \quad P(S_{1700} \ge 1061) \\ = & \quad P\left(S_{1700}^* \ge \frac{1060.5-1020}{20}\right) \\ = & \quad P(S_{1700}^* \ge 2.025) \ . \end{aligned}$$

From Table [tabl 9.1], if we interpolate, we would estimate this probability to be $.5 - .4784 = .0216$. Thus, the college is fairly safe using this admission policy.

Applications to Statistics

There are many important questions in the field of statistics that can be answered using the Central Limit Theorem for independent trials processes. The following example is one that is encountered quite frequently in the news. Another example of an application of the Central Limit Theorem to statistics is given in Section 1.2.

Example 9.1.4

One frequently reads that a poll has been taken to estimate the proportion of people in a certain population who favor one candidate over another in a race with two candidates. (This model also applies to races with more than two candidates A and B, and two ballot propositions.) Clearly, it is not possible for pollsters to ask everyone for their preference. What is done instead is to pick a subset of the population, called a sample, and ask everyone in the sample for their preference. Let p be the actual proportion of people in the population who are in favor of candidate A and let $q = 1 - p$. If we choose a sample of size n from the population, the preferences of the people in the sample can be represented by random variables $X_1,\ X_2,\ \ldots,\ X_n$, where $X_i = 1$ if person i is in favor of candidate A, and $X_i = 0$ if person i is in favor of candidate B. Let $S_n = X_1 + X_2 + \cdots + X_n$. If each subset of size n is chosen with the same probability, then S_n is hypergeometrically distributed. If n is small relative to the size of the population (which is typically true in practice), then S_n is approximately binomially distributed, with parameters n and p.

The pollster wants to estimate the value p. An estimate for p is provided by the value $\bar{p} = S_n/n$, which is the proportion of people in the sample who favor candidate B. The Central Limit Theorem says that the random variable $\bar{p}$ is approximately normally distributed. (In fact, our version of the Central Limit Theorem says that the distribution function of the random variable

$$S_n^* = \frac{S_n - np}{\sqrt{npq}} \tag{9.1.25}$$

is approximated by the standard normal density.) But we have

$$\bar{p} = \frac{S_n - np}{\sqrt{npq}} \sqrt{\frac{pq}{n}} + p\,, \tag{9.1.26}$$

i.e., $\bar{p}$ is just a linear function of S_n^*. Since the distribution of S_n^* is approximated by the standard normal density, the distribution of the random variable $\bar{p}$ must also be bell-shaped. We also know how to write the mean and standard deviation of $\bar{p}$ in terms of p and n. The mean of $\bar{p}$ is just p, and the standard deviation is

$$\sqrt{\frac{pq}{n}}\,. \tag{9.1.27}$$

Thus, it is easy to write down the standardized version of $\bar{p}$; it is

$$\bar{p}^* = \frac{\bar{p} - p}{\sqrt{pq/n}}\,. \tag{9.1.28}$$

Since the distribution of the standardized version of $\bar{p}$ is approximated by the standard normal density, we know, for example, that 95% of its values will lie within two standard deviations of its mean, and the same is true of $\bar{p}$. So we have

$$P\left(p - 2\sqrt{\frac{pq}{n}} < \bar{p} < p + 2\sqrt{\frac{pq}{n}}\right) \approx .954\,. \tag{9.1.29}$$

Now the pollster does not know p or q, but he can use $\bar{p}$ and $\bar{q} = 1 - \bar{p}$ in their place without too much danger. With this idea in mind, the above statement is equivalent to the statement

$$P\left(\bar{p} - 2\sqrt{\frac{\bar{p}\bar{q}}{n}} < p < \bar{p} + 2\sqrt{\frac{\bar{p}\bar{q}}{n}}\right) \approx .954\,. \tag{9.1.30}$$

The resulting interval

$$\left(\bar{p} - \frac{2\sqrt{\bar{p}\bar{q}}}{\sqrt{n}},\ \bar{p} + \frac{2\sqrt{\bar{p}\bar{q}}}{\sqrt{n}}\right) \tag{9.1.31}$$

is called the for the unknown value of p. The name is suggested by the fact that if we use this method to estimate p in a large number of samples we should expect that in about 95 percent of the samples the true value of p is contained in the confidence interval obtained from the sample. In Exercise $9.1.11$ you are asked to write a program to illustrate that this does indeed happen.

The pollster has control over the value of n. Thus, if he wants to create a 95% confidence interval with length 6%, then he should choose a value of n so that

$$\frac{2\sqrt{\bar{p}\bar{q}}}{\sqrt{n}} \leq .03 \ . \tag{9.1.32}$$

Using the fact that $\bar{p}\bar{q} \leq 1/4$, no matter what the value of $\bar{p}$ is, it is easy to show that if he chooses a value of n so that

$$\frac{1}{\sqrt{n}} \leq .03 \ , \tag{9.1.33}$$

he will be safe. This is equivalent to choosing

$$n \geq 1111 \ . \tag{9.1.34}$$

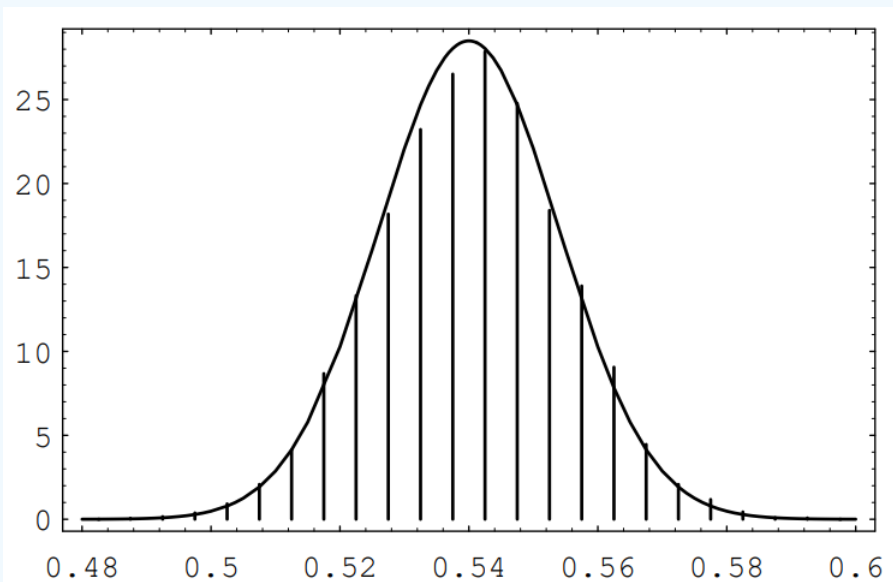


Figure $9.1.5$: Polling simulation

So if the pollster chooses n to be 1200, say, and calculates $\bar{p}$ using his sample of size 1200, then 19 times out of 20 (i.e., 95% of the time), his confidence interval, which is of length 6%, will contain the true value of p. This type of confidence interval is typically reported in the news as follows: this survey has a 3% margin of error. In fact, most of the surveys that one sees reported in the paper will have sample sizes around 1000. A somewhat surprising fact is that the size of the population has apparently no effect on the sample size needed to obtain a 95% confidence interval for p with a given margin of error. To see this, note that the value of n that was needed depended only on the number .03, which is the margin of error. In other words, whether the population is of size 100,000 or 100,000,000, the pollster needs only to choose a sample of size 1200 or so to get the same accuracy of estimate of p. (We did use the fact that the sample size was small relative to the population size in the statement that S_n is approximately binomially distributed.)

In Figure [fig 9.1.1], we show the results of simulating the polling process. The population is of size 100,000, and for the population, $p = .54$. The sample size was chosen to be 1200. The spike graph shows the distribution of $\bar{p}$ for 10,000 randomly chosen samples. For this simulation, the program kept track of the number of samples for which $\bar{p}$ was within 3% of .54. This number was 9648, which is close to 95% of the number of samples used.

Another way to see what the idea of confidence intervals means is shown in Figure [fig 9.1.6]. In this figure, we show 100 confidence intervals, obtained by computing $\bar{p}$ for 100 different samples of size 1200 from the same population as before. The reader can see that most of these confidence intervals (96, to be exact) contain the true value of p.

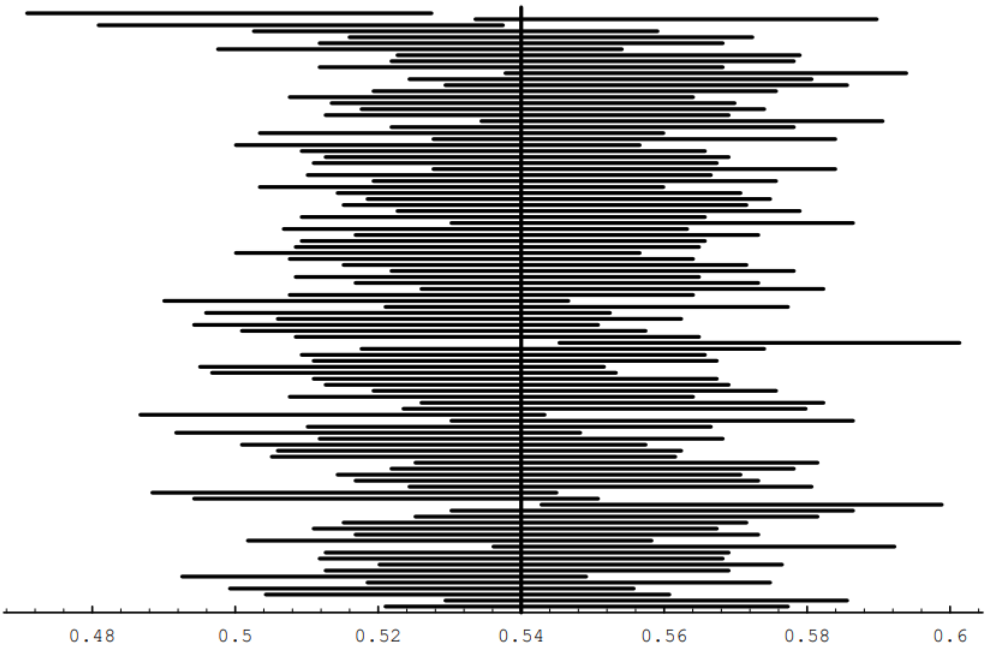


Figure 9.1.6: Confidence interval simulation.

The Gallup Poll has used these polling techniques in every Presidential election since 1936 (and in innumerable other elections as well). Table [table 9.1.1][1] shows the results of their efforts. The reader will note that most of the approximations to p are within 3% of the actual value of p. The sample sizes for these polls were typically around 1500. (In the table, both the predicted and actual percentages for the winning candidate refer to the percentage of the vote among the “major" political parties. In most elections, there were two major parties, but in several elections, there were three.)

Table 9.1.1: Gallup Poll accuracy record.

Year	Winning	Gallup Final	Election	Deviation
	Candidate	Survey	Result	
1936	Roosevelt	55.7%	62.5%	6.8%
1940	Roosevelt	52.0%	55.0%	3.0%
1944	Roosevelt	51.5%	53.3%	1.8%
1948	Truman	44.5%	49.9%	5.4%
1952	Eisenhower	51.0%	55.4%	4.4%
1956	Eisenhower	59.5%	57.8%	1.7%
1960	Kennedy	51.0%	50.1%	0.9%
1964	Johnson	64.0%	61.3%	2.7%
1968	Nixon	43.0%	43.5%	0.5%
1972	Nixon	62.0%	61.8%	0.2%
1976	Carter	48.0%	50.0%	2.0%
1980	Reagan	47.0%	50.8%	3.8%
1984	Reagan	59.0%	59.1%	0.1%
1988	Bush	56.0%	53.9%	2.1%

1992	Clinton	49.0%	43.2%	5.8%
1996	Clinton	52.0%	50.1%	1.9%

This technique also plays an important role in the evaluation of the effectiveness of drugs in the medical profession. For example, it is sometimes desired to know what proportion of patients will be helped by a new drug. This proportion can be estimated by giving the drug to a subset of the patients, and determining the proportion of this sample who are helped by the drug.

Historical Remarks

The Central Limit Theorem for Bernoulli trials was first proved by Abrahamde Moivre and appeared in his book, first published in 1718.[2]

De Moivre spent his years from age 18 to 21 in prison in France because of his Protestant background. When he was released he left France for England, where he worked as a tutor to the sons of noblemen. Newton had presented a copy of his to the Earl of Devonshire. The story goes that, while de Moivre was tutoring at the Earl's house, he came upon Newton's work and found that it was beyond him. It is said that he then bought a copy of his own and tore it into separate pages, learning it page by page as he walked around London to his tutoring jobs. De Moivre frequented the coffeehouses in London, where he started his probability work by calculating odds for gamblers. He also met Newton at such a coffeehouse and they became fast friends. De Moivre dedicated his book to Newton.

Confidence interval simulation.provides the techniques for solving a wide variety of gambling problems. In the midst of these gambling problems de Moivre rather modestly introduces his proof of the Central Limit Theorem, writing

> *A Method of approximating the Sum of the Terms of the Binomial* $(a+b)^n$ *expanded into a Series, from whence are deduced some practical Rules to estimate the Degree of Assent which is to be given to Experiments.*[3]

De Moivre's proof used the approximation to factorials that we now call Stirling's formula. De Moivre states that he had obtained this formula before Stirling but without determining the exact value of the constant $\sqrt{2\pi}$. While he says it is not really necessary to know this exact value, he concedes that knowing it "has spread a singular Elegancy on the Solution."

The complete proof and an interesting discussion of the life of de Moivre can be found in the book by F. N. David.[4]

Exercises

Exercise 9.1.1:

Let S_{100} be the number of heads that turn up in 100 tosses of a fair coin. Use the Central Limit Theorem to estimate

a. $P(S_{100} \leq 45)$.
b. $P(45 < S_{100} < 55)$.
c. $P(S_{100} > 63)$.
d. $P(S_{100} < 57)$.

Exercise 9.1.2:

Let S_{200} be the number of heads that turn up in 200 tosses of a fair coin. Estimate

a. $P(S_{200} = 100)$.
b. $P(S_{200} = 90)$.
c. $P(S_{200} = 80)$.

Exercise 9.1.3:

A true-false examination has 48 questions. June has probability 3/4 of answering a question correctly. April just guesses on each question. A passing score is 30 or more correct answers. Compare the probability that June passes the exam with the probability that April passes it.

Exercise 9.1.4:

Let S be the number of heads in 1,000,000 tosses of a fair coin. Use (a) Chebyshev's inequality, and (b) the Central Limit Theorem, to estimate the probability that S lies between 499,500 and 500,500. Use the same two methods to estimate the probability that S lies between 499,000 and 501,000, and the probability that S lies between 498,500 and 501,500.

Exercise 9.1.5:

A rookie is brought to a baseball club on the assumption that he will have a .300 batting average. (Batting average is the ratio of the number of hits to the number of times at bat.) In the first year, he comes to bat 300 times and his batting average is .267. Assume that his at bats can be considered Bernoulli trials with probability .3 for success. Could such a low average be considered just bad luck or should he be sent back to the minor leagues? Comment on the assumption of Bernoulli trials in this situation.

Exercise 9.1.6:

Once upon a time, there were two railway trains competing for the passenger traffic of 1000 people leaving from Chicago at the same hour and going to Los Angeles. Assume that passengers are equally likely to choose each train. How many seats must a train have to assure a probability of .99 or better of having a seat for each passenger?

Exercise 9.1.7:

Dartmouth admits 1750 students. What is the probability of too many acceptances?

Exercise 9.1.8:

A club serves dinner to members only. They are seated at 12-seat tables. The manager observes over a long period of time that 95 percent of the time there are between six and nine full tables of members, and the remainder of the time the numbers are equally likely to fall above or below this range. Assume that each member decides to come with a given probability p, and that the decisions are independent. How many members are there? What is p?

Exercise 9.1.9:

Let S_n be the number of successes in n Bernoulli trials with probability .8 for success on each trial. Let $A_n = S_n/n$ be the average number of successes. In each case give the value for the limit, and give a reason for your answer.

a. $\lim_{n\to\infty} P(A_n = .8)$.
b. $\lim_{n\to\infty} P(.7n < S_n < .9n)$.
c. $\lim_{n\to\infty} P(S_n < .8n + .8\sqrt{n})$.
d. $\lim_{n\to\infty} P(.79 < A_n < .81)$.

Exercise 9.1.10:

Find the probability that among 10,000 random digits the digit 3 appears not more than 931 times.

Exercise 9.1.11:

Write a computer program to simulate 10,000 Bernoulli trials with probability .3 for success on each trial. Have the program compute the 95 percent confidence interval for the probability of success based on the proportion of successes. Repeat the experiment 100 times and see how many times the true value of .3 is included within the confidence limits.

Exercise 9.1.12:

A balanced coin is flipped 400 times. Determine the number x such that the probability that the number of heads is between $200 - x$ and $200 + x$ is approximately .80.

Exercise 9.1.13:

A noodle machine in Spumoni's spaghetti factory makes about 5 percent defective noodles even when properly adjusted. The noodles are then packed in crates containing 1900 noodles each. A crate is examined and found to contain 115 defective noodles. What is the approximate probability of finding at least this many defective noodles if the machine is properly adjusted?

Exercise 9.1.14:

A restaurant feeds 400 customers per day. On the average 20 percent of the customers order apple pie.

a. Give a range (called a 95 percent confidence interval) for the number of pieces of apple pie ordered on a given day such that you can be 95 percent sure that the actual number will fall in this range.
b. How many customers must the restaurant have, on the average, to be at least 95 percent sure that the number of customers ordering pie on that day falls in the 19 to 21 percent range?

Exercise 9.1.15:

Recall that if X is a random variable, the of X is the function $F(x)$ defined by

$$F(x) = P(X \leq x) \,. \tag{9.1.35}$$

a. Let S_n be the number of successes in n Bernoulli trials with probability p for success. Write a program to plot the cumulative distribution for S_n.
b. Modify your program in (a) to plot the cumulative distribution $F_n^*(x)$ of the standardized random variable

$$S_n^* = \frac{S_n - np}{\sqrt{npq}} \,. \tag{9.1.36}$$

c. Define the $N(x)$ to be the area under the normal curve up to the value x. Modify your program in (b) to plot the normal distribution as well, and compare it with the cumulative distribution of S_n^*. Do this for $n = 10, 50$, and 100.

Exercise 9.1.16:

In Example 3.12, we were interested in testing the hypothesis that a new form of aspirin is effective 80 percent of the time rather than the 60 percent of the time as reported for standard aspirin. The new aspirin is given to n people. If it is effective in m or more cases, we accept the claim that the new drug is effective 80 percent of the time and if not we reject the claim. Using the Central Limit Theorem, show that you can choose the number of trials n and the critical value m so that the probability that we reject the hypothesis when it is true is less than .01 and the probability that we accept it when it is false is also less than .01. Find the smallest value of n that will suffice for this.

Exercise 9.1.17:

In an opinion poll it is assumed that an unknown proportion p of the people are in favor of a proposed new law and a proportion $1-p$ are against it. A sample of n people is taken to obtain their opinion. The proportion $\bar{p}$ in favor in the sample is taken as an estimate of p. Using the Central Limit Theorem, determine how large a sample will ensure that the estimate will, with probability .95, be correct to within .01.

Exercise 9.1.18:

A description of a poll in a certain newspaper says that one can be 95% confident that error due to sampling will be no more than plus or minus 3 percentage points. A poll in the New York Times taken in Iowa says that “according to statistical theory, in 19 out of 20 cases the results based on such samples will differ by no more than 3 percentage points in either direction from what would have been obtained by interviewing all adult Iowans." These are both attempts to explain the concept of confidence intervals. Do both statements say the same thing? If not, which do you think is the more accurate description?

This page titled 9.1: Central Limit Theorem for Bernoulli Trials is shared under a GNU Free Documentation License 1.3 license and was authored, remixed, and/or curated by Charles M. Grinstead & J. Laurie Snell (American Mathematical Society) via source content that was edited to the style and standards of the LibreTexts platform.