

## 6.4: The Central Limit Theorem

The central limit theorem and the law of large numbers are the two *fundamental theorems* of probability. Roughly, the central limit theorem states that the distribution of the sum (or average) of a large number of independent, identically distributed variables will be approximately normal, regardless of the underlying distribution. The importance of the central limit theorem is hard to overstate; indeed it is the reason that many statistical procedures work.

### Partial Sum Processes

#### Definitions

Suppose that  $\mathbf{X} = (X_1, X_2, \dots)$  is a sequence of independent, identically distributed, real-valued random variables with common probability density function  $f$ , mean  $\mu$ , and variance  $\sigma^2$ . We assume that  $0 < \sigma < \infty$ , so that in particular, the random variables really are *random* and not constants. Let

$$Y_n = \sum_{i=1}^n X_i, \quad n \in \mathbb{N} \quad (6.4.1)$$

Note that by convention,  $Y_0 = 0$ , since the sum is over an empty index set. The random process  $\mathbf{Y} = (Y_0, Y_1, Y_2, \dots)$  is called the *partial sum process* associated with  $\mathbf{X}$ . Special types of partial sum processes have been studied in many places in this text; in particular see

- the binomial distribution in the setting of Bernoulli trials
- the negative binomial distribution in the setting of Bernoulli trials
- the gamma distribution in the Poisson process
- the arrival times in a general renewal process

Recall that in statistical terms, the sequence  $\mathbf{X}$  corresponds to sampling from the underlying distribution. In particular,  $(X_1, X_2, \dots, X_n)$  is a random sample of size  $n$  from the distribution, and the corresponding sample mean is

$$M_n = \frac{Y_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i \quad (6.4.2)$$

By the law of large numbers,  $M_n \rightarrow \mu$  as  $n \rightarrow \infty$  with probability 1.

#### Stationary, Independent Increments

The partial sum process corresponding to a sequence of independent, identically distributed variables has two important properties, and these properties essentially characterize such processes.

If  $m \leq n$  then  $Y_n - Y_m$  has the same distribution as  $Y_{n-m}$ . Thus the process  $\mathbf{Y}$  has *stationary increments*.

Proof

Note that  $Y_n - Y_m = \sum_{i=m+1}^n X_i$  and is the sum of  $n - m$  independent variables, each with the common distribution. Of course,  $Y_{n-m}$  is also the sum of  $n - m$  independent variables, each with the common distribution.

Note however that  $Y_n - Y_m$  and  $Y_{n-m}$  are very different random variables; the theorem simply states that they have the same *distribution*.

If  $n_1 \leq n_2 \leq n_3 \leq \dots$  then  $(Y_{n_1}, Y_{n_2} - Y_{n_1}, Y_{n_3} - Y_{n_2}, \dots)$  is a sequence of independent random variables. Thus the process  $\mathbf{Y}$  has *independent increments*.

Proof

The terms in the sequence of increments  $(Y_{n_1}, Y_{n_2} - Y_{n_1}, Y_{n_3} - Y_{n_2}, \dots)$  are sums over disjoint collections of terms in the sequence  $\mathbf{X}$ . Since the sequence  $\mathbf{X}$  is independent, so is the sequence of increments.

Conversely, suppose that  $\mathbf{V} = (V_0, V_1, V_2, \dots)$  is a random process with stationary, independent increments. Define  $U_i = V_i - V_{i-1}$  for  $i \in \mathbb{N}_+$ . Then  $\mathbf{U} = (U_1, U_2, \dots)$  is a sequence of independent, identically distributed variables and  $\mathbf{V}$  is

the partial sum process associated with  $U$ .

Thus, partial sum processes are the only discrete-time random processes that have stationary, independent increments. An interesting, and much harder problem, is to characterize the continuous-time processes that have stationary independent increments. The Poisson counting process has stationary independent increments, as does the Brownian motion process.

## Moments

If  $n \in \mathbb{N}$  then

1.  $\mathbb{E}(Y_n) = n\mu$
2.  $\text{var}(Y_n) = n\sigma^2$

Proof

The results follow from basic properties of expected value and variance. Expected value is a linear operation so  $\mathbb{E}(Y_n) = \sum_{i=1}^n \mathbb{E}(X_i) = n\mu$ . By independence,  $\text{var}(Y_n) = \sum_{i=1}^n \text{var}(X_i) = n\sigma^2$ .

If  $n \in \mathbb{N}_+$  and  $m \in \mathbb{N}$  with  $m \leq n$  then

1.  $\text{cov}(Y_m, Y_n) = m\sigma^2$
2.  $\text{cor}(Y_m, Y_n) = \sqrt{\frac{m}{n}}$
3.  $\mathbb{E}(Y_m Y_n) = m\sigma^2 + mn\mu^2$

Proof

1. Note that  $Y_n = Y_m + (Y_n - Y_m)$ . This follows from basic properties of covariance, and [Theorem 1](#) and [Theorem 2](#):

$$\text{cov}(Y_m, Y_n) = \text{cov}(Y_m, Y_m) + \text{cov}(Y_m, Y_n - Y_m) = \text{var}(Y_m) + 0 = m\sigma^2 \quad (6.4.3)$$

2. This result follows from part (a) and [Theorem 4](#)

$$\text{cor}(Y_m, Y_n) = \frac{\text{cov}(Y_m, Y_n)}{\text{sd}(Y_m)\text{sd}(Y_n)} = \frac{m\sigma^2}{\sqrt{m\sigma^2}\sqrt{n\sigma^2}} = \sqrt{\frac{m}{n}} \quad (6.4.4)$$

3. This result also follows from part (a) and [Theorem 4](#):  $\mathbb{E}(Y_m Y_n) = \text{cov}(Y_m, Y_n) + \mathbb{E}(Y_m)\mathbb{E}(Y_n) = m\sigma^2 + m\mu n\mu$

If  $X$  has moment generating function  $G$  then  $Y_n$  has moment generating function  $G^n$ .

Proof

This follows from a basic property of generating functions: the generating function of a sum of independent variables is the product of the generating functions of the terms.

## Distributions

Suppose that  $X$  has either a discrete distribution or a continuous distribution with probability density function  $f$ . Then the probability density function of  $Y_n$  is  $f^{*n} = f * f * \dots * f$ , the convolution power of  $f$  of order  $n$ .

Proof

This follows from a basic property of PDFs: the pdf of a sum of independent variables is the convolution of the PDFs of the terms.

More generally, we can use the stationary and independence properties to find the joint distributions of the partial sum process:

If  $n_1 < n_2 < \dots < n_k$  then  $(Y_{n_1}, Y_{n_2}, \dots, Y_{n_k})$  has joint probability density function

$$f_{n_1, n_2, \dots, n_k}(y_1, y_2, \dots, y_k) = f^{*n_1}(y_1) f^{*(n_2 - n_1)}(y_2 - y_1) \dots f^{*(n_k - n_{k-1})}(y_k - y_{k-1}), \quad (y_1, y_2, \dots, y_k) \in \mathbb{R}^k \quad (6.4.5)$$

Proof

This follows from the multivariate change of variables theorem.

## The Central Limit Theorem

First, let's make the central limit theorem more precise. From [Theorem 4](#), we cannot expect  $Y_n$  itself to have a limiting distribution. Note that  $\text{var}(Y_n) \rightarrow \infty$  as  $n \rightarrow \infty$  since  $\sigma > 0$ , and  $\mathbb{E}(Y_n) \rightarrow \infty$  as  $n \rightarrow \infty$  if  $\mu > 0$  while  $\mathbb{E}(Y_n) \rightarrow -\infty$  as  $n \rightarrow \infty$  if  $\mu < 0$ . Similarly, we know that  $M_n \rightarrow \mu$  as  $n \rightarrow \infty$  with probability 1, so the limiting distribution of the sample mean is degenerate. Thus, to obtain a limiting distribution of  $Y_n$  or  $M_n$  that is not degenerate, we need to consider, not these variables themselves, but rather the common *standard score*. Thus, let

$$Z_n = \frac{Y_n - n\mu}{\sqrt{n}\sigma} = \frac{M_n - \mu}{\sigma/\sqrt{n}} \quad (6.4.6)$$

$Z_n$  has mean 0 and variance 1.

1.  $\mathbb{E}(Z_n) = 0$
2.  $\text{var}(Z_n) = 1$

Proof

These results follow from basic properties of expected value and variance, and are true for the standard score associated with any random variable. Recall also that the standard score of a variable is invariant under linear transformations with positive slope. The fact that the standard score of  $Y_n$  and the standard score of  $M_n$  are the same is a special case of this.

The precise statement of the central limit theorem is that the distribution of the standard score  $Z_n$  converges to the standard normal distribution as  $n \rightarrow \infty$ . Recall that the standard normal distribution has probability density function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad z \in \mathbb{R} \quad (6.4.7)$$

and is studied in more detail in the chapter on special distributions. A special case of the central limit theorem (to Bernoulli trials), dates to Abraham De Moivre. The term *central limit theorem* was coined by George Pólya in 1920. By definition of convergence in distribution, the central limit theorem states that  $F_n(z) \rightarrow \Phi(z)$  as  $n \rightarrow \infty$  for each  $z \in \mathbb{R}$ , where  $F_n$  is the distribution function of  $Z_n$  and  $\Phi$  is the standard normal distribution function:

$$\Phi(z) = \int_{-\infty}^z \phi(x) dx = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx, \quad z \in \mathbb{R} \quad (6.4.8)$$

An equivalent statement of the central limit theorem involves convergence of the corresponding characteristic functions. This is the version that we will give and prove, but first we need a generalization of a famous limit from calculus.

Suppose that  $(a_1, a_2, \dots)$  is a sequence of real numbers and that  $a_n \rightarrow a \in \mathbb{R}$  as  $n \rightarrow \infty$ . Then

$$\left(1 + \frac{a_n}{n}\right)^n \rightarrow e^a \text{ as } n \rightarrow \infty \quad (6.4.9)$$

Now let  $\chi$  denote the characteristic function of the standard score of the sample variable  $X$ , and let  $\chi_n$  denote the characteristic function of the standard score  $Z_n$ :

$$\chi(t) = \mathbb{E} \left[ \exp \left( it \frac{X - \mu}{\sigma} \right) \right], \quad \chi_n(t) = \mathbb{E}[\exp(itZ_n)]; \quad t \in \mathbb{R} \quad (6.4.10)$$

Recall that  $t \mapsto e^{-\frac{1}{2}t^2}$  is the characteristic function of the standard normal distribution. We can now give a proof.

*The central limit theorem.* The distribution of  $Z_n$  converges to the standard normal distribution as  $n \rightarrow \infty$ . That is,  $\chi_n(t) \rightarrow e^{-\frac{1}{2}t^2}$  as  $n \rightarrow \infty$  for each  $t \in \mathbb{R}$ .

Proof

Note that  $\chi(0) = 1$ ,  $\chi'(0) = 0$ ,  $\chi''(0) = -1$ . Next

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} \quad (6.4.11)$$

From properties of characteristic functions,  $\chi_n(t) = \chi^n(t/\sqrt{n})$  for  $t \in \mathbb{R}$ . By Taylor's theorem (named after Brook Taylor),

$$\chi\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{1}{2}\chi''(s_n)\frac{t^2}{n} \text{ where } |s_n| \leq \frac{|t|}{n} \quad (6.4.12)$$

But  $s_n \rightarrow 0$  and hence  $\chi''(s_n) \rightarrow -1$  as  $n \rightarrow \infty$ . Finally,

$$\chi_n(t) = \left[1 + \frac{1}{2}\chi''(s_n)\frac{t^2}{n}\right]^n \rightarrow e^{-\frac{1}{2}t^2} \text{ as } n \rightarrow \infty \quad (6.4.13)$$

## Normal Approximations

The central limit theorem implies that if the sample size  $n$  is “large” then the distribution of the partial sum  $Y_n$  is approximately normal with mean  $n\mu$  and variance  $n\sigma^2$ . Equivalently the sample mean  $M_n$  is approximately normal with mean  $\mu$  and variance  $\sigma^2/n$ . The central limit theorem is of fundamental importance, because it means that we can approximate the distribution of certain statistics, even if we know very little about the underlying sampling distribution.

Of course, the term “large” is relative. Roughly, the more “abnormal” the basic distribution, the larger  $n$  must be for normal approximations to work well. The rule of thumb is that a sample size  $n$  of at least 30 will usually suffice if the basic distribution is not too weird; although for many distributions smaller  $n$  will do.

Let  $Y$  denote the sum of the variables in a random sample of size 30 from the uniform distribution on  $[0, 1]$ . Find normal approximations to each of the following:

1.  $\mathbb{P}(13 < Y < 18)$
2. The 90th percentile of  $Y$

Answer

1. 0.8682
2. 17.03

Random variable  $Y$  in the previous exercise has the *Irwin-Hall distribution* of order 30. The Irwin-Hall distributions are studied in more detail in the chapter on Special Distributions and are named for Joseph Irwin and Phillip Hall.

In the special distribution simulator, select the Irwin-Hall distribution. Vary  $n$  from 1 to 10 and note the shape of the probability density function. With  $n = 10$  run the experiment 1000 times and compare the empirical density function to the true probability density function.

Let  $M$  denote the sample mean of a random sample of size 50 from the distribution with probability density function  $f(x) = \frac{3}{x^4}$  for  $1 \leq x < \infty$ . This is a *Pareto distribution*, named for Vilfredo Pareto. Find normal approximations to each of the following:

1.  $\mathbb{P}(M > 1.6)$
2. The 60th percentile of  $M$

Answer

1. 0.2071
2. 1.531

## The Continuity Correction

A slight technical problem arises when the sampling distribution is discrete. In this case, the partial sum also has a discrete distribution, and hence we are approximating a discrete distribution with a continuous one. Suppose that  $X$  takes integer values (the most common case) and hence so does the partial sum  $Y_n$ . For any  $k \in \mathbb{Z}$  and  $h \in [0, 1]$ , note that the event  $\{k - h \leq Y_n \leq k + h\}$  is equivalent to the event  $\{Y = k\}$ . Different values of  $h$  lead to different normal approximations, even though the events are equivalent. The smallest approximation would be 0 when  $h = 0$ , and the approximations increase as  $h$  increases. It is customary to split the difference by using  $h = \frac{1}{2}$  for the normal approximation. This is sometimes called the *half-*

*unit continuity correction* or the *histogram correction*. The continuity correction is extended to other events in the natural way, using the additivity of probability.

Suppose that  $j, k \in \mathbb{Z}$  with  $j \leq k$ .

1. For the event  $\{j \leq Y_n \leq k\} = \{j-1 < Y_n < k+1\}$ , use  $\{j - \frac{1}{2} \leq Y_n \leq k + \frac{1}{2}\}$  in the normal approximation.
2. For the event  $\{j \leq Y_n\} = \{j-1 < Y_n\}$ , use  $\{j - \frac{1}{2} \leq Y_n\}$  in the normal approximation.
3. For the event  $\{Y_n \leq k\} = \{Y_n < k+1\}$ , use  $\{Y_n \leq k + \frac{1}{2}\}$  in the normal approximation.

Let  $Y$  denote the sum of the scores of 20 fair dice. Compute the normal approximation to  $\mathbb{P}(60 \leq Y \leq 75)$ .

Answer

0.6741

In the dice experiment, set the die distribution to fair, select the sum random variable  $Y$ , and set  $n = 20$ . Run the simulation 1000 times and find each of the following. Compare with the result in the previous exercise:

1.  $\mathbb{P}(60 \leq Y \leq 75)$
2. The relative frequency of the event  $\{60 \leq Y \leq 75\}$  (from the simulation)

### Normal Approximation to the Gamma Distribution

Recall that the *gamma distribution* with shape parameter  $k \in (0, \infty)$  and scale parameter  $b \in (0, \infty)$  is a continuous distribution on  $(0, \infty)$  with probability density function  $f$  given by

$$f(x) = \frac{1}{\Gamma(k)b^k} x^{k-1} e^{-x/b}, \quad x \in (0, \infty) \quad (6.4.14)$$

The mean is  $kb$  and the variance is  $kb^2$ . The gamma distribution is widely used to model random times (particularly in the context of the Poisson model) and other positive random variables. The general gamma distribution is studied in more detail in the chapter on Special Distributions. In the context of the Poisson model (where  $k \in \mathbb{N}_+$ ), the gamma distribution is also known as the *Erlang distribution*, named for Agner Erlang; it is studied in more detail in the chapter on the Poisson Process. Suppose now that  $Y_k$  has the gamma (Erlang) distribution with shape parameter  $k \in \mathbb{N}_+$  and scale parameter  $b > 0$  then

$$Y_k = \sum_{i=1}^k X_i \quad (6.4.15)$$

where  $(X_1, X_2, \dots)$  is a sequence of independent variables, each having the exponential distribution with scale parameter  $b$ . (The exponential distribution is a special case of the gamma distribution with shape parameter 1.) It follows that if  $k$  is large, the gamma distribution can be approximated by the normal distribution with mean  $kb$  and variance  $kb^2$ . The same statement actually holds when  $k$  is not an integer. Here is the precise statement:

Suppose that  $Y_k$  has the gamma distribution with scale parameter  $b \in (0, \infty)$  and shape parameter  $k \in (0, \infty)$ . Then the distribution of the standardized variable  $Z_k$  below converges to the standard normal distribution as  $k \rightarrow \infty$ :

$$Z_k = \frac{Y_k - kb}{\sqrt{kb}} \quad (6.4.16)$$

In the special distribution simulator, select the gamma distribution. Vary  $k$  and  $b$  and note the shape of the probability density function. With  $k = 10$  and various values of  $b$ , run the experiment 1000 times and compare the empirical density function to the true probability density function.

Suppose that  $Y$  has the gamma distribution with shape parameter  $k = 10$  and scale parameter  $b = 2$ . Find normal approximations to each of the following:

1.  $\mathbb{P}(18 \leq Y \leq 23)$
2. The 80th percentile of  $Y$

Answer

1. 0.3063
2. 25.32

### Normal Approximation to the Chi-Square Distribution

Recall that the *chi-square distribution* with  $n \in (0, \infty)$  degrees of freedom is a special case of the gamma distribution, with shape parameter  $k = n/2$  and scale parameter  $b = 2$ . Thus, the chi-square distribution with  $n$  degrees of freedom has probability density function

$$f(x) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2}, \quad 0 < x < \infty \quad (6.4.17)$$

When  $n$  is a positive, integer, the chi-square distribution governs the sum of  $n$  independent, standard normal variables. For this reason, it is one of the most important distributions in statistics. The chi-square distribution is studied in more detail in the chapter on Special Distributions. From the [previous discussion](#), it follows that if  $n$  is large, the chi-square distribution can be approximated by the normal distribution with mean  $n$  and variance  $2n$ . Here is the precise statement:

Suppose that  $Y_n$  has the chi-square distribution with  $n \in (0, \infty)$  degrees of freedom. Then the distribution of the standardized variable  $Z_n$  below converges to the standard normal distribution as  $n \rightarrow \infty$ :

$$Z_n = \frac{Y_n - n}{\sqrt{2n}} \quad (6.4.18)$$

In the special distribution simulator, select the chi-square distribution. Vary  $n$  and note the shape of the probability density function. With  $n = 20$ , run the experiment 1000 times and compare the empirical density function to the probability density function.

Suppose that  $Y$  has the chi-square distribution with  $n = 20$  degrees of freedom. Find normal approximations to each of the following:

1.  $\mathbb{P}(18 < Y < 25)$
2. The 75th percentile of  $Y$

Answer

1. 0.4107
2. 24.3

### Normal Approximation to the Binomial Distribution

Recall that a *Bernoulli trials sequence*, named for Jacob Bernoulli, is a sequence  $(X_1, X_2, \dots)$  of independent, identically distributed indicator variables with  $\mathbb{P}(X_i = 1) = p$  for each  $i$ , where  $p \in (0, 1)$  is the parameter. In the usual language of reliability,  $X_i$  is the outcome of trial  $i$ , where 1 means success and 0 means failure. The common mean is  $p$  and the common variance is  $p(1 - p)$ .

Let  $Y_n = \sum_{i=1}^n X_i$ , so that  $Y_n$  is the number of successes in the first  $n$  trials. Recall that  $Y_n$  has the *binomial distribution* with parameters  $n$  and  $p$ , and has probability density function

$$f(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k \in \{0, 1, \dots, n\} \quad (6.4.19)$$

The binomial distribution is studied in more detail in the chapter on Bernoulli trials.

It follows from the central limit theorem that if  $n$  is large, the binomial distribution with parameters  $n$  and  $p$  can be approximated by the normal distribution with mean  $np$  and variance  $np(1 - p)$ . The rule of thumb is that  $n$  should be large enough for  $np \geq 5$  and  $n(1 - p) \geq 5$ . (The first condition is the important one when  $p < \frac{1}{2}$  and the second condition is the important one when  $p > \frac{1}{2}$ .) Here is the precise statement:

Suppose that  $Y_n$  has the binomial distribution with trial parameter  $n \in \mathbb{N}_+$  and success parameter  $p \in (0, 1)$ . Then the distribution of the standardized variable  $Z_n$  given below converges to the standard normal distribution as  $n \rightarrow \infty$ :

$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}} \quad (6.4.20)$$

In the binomial timeline experiment, vary  $n$  and  $p$  and note the shape of the probability density function. With  $n = 50$  and  $p = 0.3$ , run the simulation 1000 times and compute the following:

1.  $\mathbb{P}(12 \leq Y \leq 16)$
2. The relative frequency of the event  $\{12 \leq Y \leq 16\}$  (from the simulation)

Answer

1. 0.5448

Suppose that  $Y$  has the binomial distribution with parameters  $n = 50$  and  $p = 0.3$ . Compute the normal approximation to  $\mathbb{P}(12 \leq Y \leq 16)$  (don't forget the continuity correction) and compare with the results of the previous exercise.

Answer

- 0.5383

### Normal Approximation to the Poisson Distribution

Recall that the *Poisson distribution*, named for Simeon Poisson, is a discrete distribution on  $\mathbb{N}$  with probability density function  $f$  given by

$$f(x) = e^{-\theta} \frac{\theta^x}{x!}, \quad x \in \mathbb{N} \quad (6.4.21)$$

where  $\theta > 0$  is a parameter. The parameter is both the mean and the variance of the distribution. The Poisson distribution is widely used to model the number of “random points” in a region of time or space, and is studied in more detail in the chapter on the Poisson Process. In this context, the parameter is proportional to the size of the region.

Suppose now that  $Y_n$  has the Poisson distribution with parameter  $n \in \mathbb{N}_+$ . Then

$$Y_n = \sum_{i=1}^n X_i \quad (6.4.22)$$

where  $(X_1, X_2, \dots, X_n)$  is a sequence of independent variables, each with the Poisson distribution with parameter 1. It follows from the central limit theorem that if  $n$  is large, the Poisson distribution with parameter  $n$  can be approximated by the normal distribution with mean  $n$  and variance  $n$ . The same statement holds when the parameter  $n$  is not an integer. Here is the precise statement:

. Suppose that  $Y_\theta$  has the Poisson distribution with parameter  $\theta \in (0, \infty)$ . Then the distribution of the standardized variable  $Z_\theta$  below converges to the standard normal distribution as  $\theta \rightarrow \infty$ :

$$Z_\theta = \frac{Y_\theta - \theta}{\sqrt{\theta}} \quad (6.4.23)$$

Suppose that  $Y$  has the Poisson distribution with mean 20.

1. Compute the true value of  $\mathbb{P}(16 \leq Y \leq 23)$ .
2. Compute the normal approximation to  $\mathbb{P}(16 \leq Y \leq 23)$ .

Answer

1. 0.6310
2. 0.6259

In the Poisson experiment, vary the time and rate parameters  $t$  and  $r$  (the parameter of the Poisson distribution in the experiment is the product  $rt$ ). Note the shape of the probability density function. With  $r = 5$  and  $t = 4$ , run the experiment 1000 times and compare the empirical density function to the true probability density function.

### Normal Approximation to the Negative Binomial Distribution

The general version of the *negative binomial distribution* is a discrete distribution on  $\mathbb{N}$ , with shape parameter  $k \in (0, \infty)$  and success parameter  $p \in (0, 1)$ . The probability density function  $f$  is given by

$$f(n) = \binom{n+k-1}{n} p^k (1-p)^n, \quad n \in \mathbb{N}_+ \quad (6.4.24)$$

The mean is  $k(1-p)/p$  and the variance is  $k(1-p)/p^2$ . The negative binomial distribution is studied in more detail in the chapter on Bernoulli trials. If  $k \in \mathbb{N}_+$ , the distribution governs the number of failures  $Y_k$  before success number  $k$  in a sequence of Bernoulli trials with success parameter  $p$ . Thus in this case,

$$Y_k = \sum_{i=1}^k X_i \quad (6.4.25)$$

where  $(X_1, X_2, \dots, X_k)$  is a sequence of independent variables, each having the geometric distribution on  $\mathbb{N}$  with parameter  $p$ . (The geometric distribution is a special case of the negative binomial, with parameters 1 and  $p$ .) In the context of the Bernoulli trials,  $X_1$  is the number of failures before the first success, and for  $i \in \{2, 3, \dots\}$ ,  $X_i$  is the number of failures between success number  $i-1$  and success number  $i$ . It follows that if  $k$  is large, the negative binomial distribution can be approximated by the normal distribution. The same statement holds if  $k$  is not an integer. Here is the precise statement:

Suppose that  $Y_k$  has the negative binomial distribution with shape parameter  $k \in (0, 1)$  and scale parameter  $p \in (0, 1)$ . Then the distribution of the standardized variable  $Z_k$  below converges to the standard normal distribution as  $k \rightarrow \infty$ :

$$Z_k = \frac{pY_k - k(1-p)}{\sqrt{k(1-p)}} \quad (6.4.26)$$

Another version of the negative binomial distribution is the distribution of the trial number  $V_k$  of success number  $k \in \mathbb{N}_+$ . So  $V_k = k + Y_k$  and  $V_k$  has mean  $k/p$  and variance  $k(1-p)/p^2$ . The normal approximation applies to the distribution of  $V_k$  as well, if  $k$  is large, and since the distributions are related by a location transformation, the standard scores are the same. That is

$$\frac{pV_k - k}{\sqrt{k(1-p)}} = \frac{pY_k - k(1-p)}{\sqrt{k(1-p)}} \quad (6.4.27)$$

In the negative binomial experiment, vary  $k$  and  $p$  and note the shape of the probability density function. With  $k = 5$  and  $p = 0.4$ , run the experiment 1000 times and compare the empirical density function to the true probability density function.

Suppose that  $Y$  has the negative binomial distribution with trial parameter  $k = 10$  and success parameter  $p = 0.4$ . Find normal approximations to each of the following:

1.  $\mathbb{P}(20 < Y < 30)$
2. The 80th percentile of  $Y$

Answer

1. 0.6318
2. 30.1

### Partial Sums with a Random Number of Terms

Our last topic is a bit more esoteric, but still fits with the general setting of this section. Recall that  $\mathbf{X} = (X_1, X_2, \dots)$  is a sequence of independent, identically distributed real-valued random variables with common mean  $\mu$  and variance  $\sigma^2$ . Suppose now that  $N$  is a random variable (on the same probability space) taking values in  $\mathbb{N}$ , also with finite mean and variance. Then

$$Y_N = \sum_{i=1}^N X_i \quad (6.4.28)$$

is a random sum of the independent, identically distributed variables. That is, the terms are random of course, but so also is the number of terms  $N$ . We are primarily interested in the moments of  $Y_N$ .

### Independent Number of Terms

Suppose first that  $N$ , the number of terms, is independent of  $\mathbf{X}$ , the sequence of terms. Computing the moments of  $Y_N$  is a good exercise in conditional expectation.

The conditional expected value of  $Y_N$  given  $N$ , and the expected value of  $Y_N$  are

1.  $\mathbb{E}(Y_N | N) = N\mu$
2.  $\mathbb{E}(Y_N) = \mathbb{E}(N)\mu$

The conditional variance of  $Y_N$  given  $N$  and the variance of  $Y_N$  are

1.  $\text{var}(Y_N | N) = N\sigma^2$
2.  $\text{var}(Y_N) = \mathbb{E}(N)\sigma^2 + \text{var}(N)\mu^2$

Let  $H$  denote the probability generating function of  $N$ . Show that the moment generating function of  $Y_N$  is  $H \circ G$ .

1.  $\mathbb{E}(e^{tY_N} | N) = [G(t)]^N$
2.  $\mathbb{E}(e^{tY_N}) = H(G(t))$

### Wald's Equation

The result in [Exercise 29 \(b\)](#) generalizes to the case where the random number of terms  $N$  is a *stopping time* for the sequence  $\mathbf{X}$ . This means that the event  $\{N = n\}$  depends only on (technically, is measurable with respect to)  $(X_1, X_2, \dots, X_n)$  for each  $n \in \mathbb{N}$ . The generalization is known as *Wald's equation*, and is named for Abraham Wald. Stopping times are studied in much more technical detail in the section on Filtrations and Stopping Times.

If  $N$  is a stopping time for  $\mathbf{X}$  then  $\mathbb{E}(Y_N) = \mathbb{E}(N)\mu$ .

Proof

First note that  $Y_N = \sum_{i=1}^{\infty} X_i \mathbf{1}(i \leq N)$ . But  $\{i \leq N\} = \{N \geq i\}^c$  depends only on  $\{X_1, \dots, X_{i-1}\}$  and hence is independent of  $X_i$ . Thus  $\mathbb{E}[X_i \mathbf{1}(i \leq N)] = \mu \mathbb{P}(N \geq i)$ . Suppose that  $X_i \geq 0$  for each  $i$ . Taking expected values term by term gives Wald's equation in this special case. The interchange of sum and expected value is justified by the monotone convergence theorem. Now Wald's equation can be established in general by using the dominated convergence theorem.

An elegant proof of Wald's equation is given in the chapter on Martingales.

Suppose that the number of customers arriving at a store during a given day has the Poisson distribution with parameter 50. Each customer, independently of the others (and independently of the number of customers), spends an amount of money that is uniformly distributed on the interval  $[0, 20]$ . Find the mean and standard deviation of the amount of money that the store takes in during a day.

Answer

500, 81.65

When a certain critical component in a system fails, it is immediately replaced by a new, statistically identical component. The components are independent, and the lifetime of each (in hours) is exponentially distributed with scale parameter  $b$ . During the life of the system, the number of critical components used has a geometric distribution on  $\mathbb{N}_+$  with parameter  $p$ . For the total life of the critical component,

1. Find the mean.
2. Find the standard deviation.
3. Find the moment generating function.

4. Identify the distribution by name.

Answer

1.  $b/p$

2.  $b/p$

3.  $t \mapsto \frac{1}{1-(b/p)t}$

4. Exponential distribution with scale parameter  $b/p$

---

This page titled [6.4: The Central Limit Theorem](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.