

4. Partial Sums and 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 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 $X = (X_1, X_2, \dots)$ is a sequence of [independent](#), identically distributed, real-valued [random variables](#) with common [probability density function](#) f , [mean](#) μ , and [variance](#) σ^2 . Let

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

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

$$M_n = \frac{Y_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

By the [law of large numbers](#), $M_n \rightarrow \mu$ as $n \rightarrow \infty$ with probability 1.

Stationary, Independent Increments

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

2. Show that 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 Y has **independent increments**.

3. Conversely, suppose that $V = (V_0, V_1, \dots)$ is a random process with stationary, independent increments, in the sense of [Exercise 1](#) and [Exercise 2](#). Define $U_i = V_i - V_{i-1}$ for $i \in \{1, 2, \dots\}$. Show that U is a sequence of

independent, identically distributed variables and that 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

4. Suppose that $n \in \mathbb{N}$. Use basic properties of [expected value](#) and [variance](#) to show that

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

5. Suppose that $n \in \mathbb{N}_+$ and $m \in \mathbb{N}$ with $m \leq n$. Use basic properties of [covariance](#) and the [stationary](#) and [independence](#) properties to verify the following results. *Hint:* Note that $Y_n = Y_m + (Y_n - Y_m)$.

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

6. Suppose that X has [moment generating function](#) G . Show that Y_n has moment generating function G^n

Distributions

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

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

8. Suppose that $n_1 < n_2 < \dots < n_k$. Show that $(Y_{n_1}, Y_{n_2}, \dots, Y_{n_k})$ has joint probability density function

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

The Central Limit Theorem

We will now make the central limit theorem precise. From [Exercise 4](#), we cannot expect Y_n itself to have a limiting distribution. Note that

- $\text{var}(Y_n) \rightarrow \infty$ as $n \rightarrow \infty$ if $\sigma > 0$
- $\mathbb{E}(Y_n) \rightarrow \infty$ as $n \rightarrow \infty$ if $\mu > 0$ and $\mathbb{E}(Y_n) \rightarrow -\infty$ as $n \rightarrow \infty$ if $\mu < 0$

Thus, to obtain a limiting distribution that is not degenerate, we need to consider, not Y_n itself, but the [standard score](#) of Y_n . Thus, let

$$Z_n = \frac{Y_n - n\mu}{\sqrt{n}\sigma}$$

9. Show that

- a. $\mathbb{E}(Z_n) = 0$
- b. $\text{var}(Z_n) = 1$

10. Show that Z_n is also the standard score of the sample mean M_n :

$$Z_n = \frac{M_n - \mu}{\sigma / \sqrt{n}}$$

The **central limit theorem** states that the distribution of the standard score Z_n converges to the **standard normal distribution** as $n \rightarrow \infty$. 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.

Proof of the Central Limit Theorem

We need to show 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 Φ the distribution function of the standard normal distribution. Equivalently we will show that

$$\chi_n(t) \rightarrow e^{-\frac{1}{2}t^2} \text{ as } n \rightarrow \infty \text{ for each } t \in \mathbb{R}$$

where χ_n is the **characteristic function** of Z_n , and the expression on the right is the characteristic function of the standard normal distribution.

The following exercises sketch the proof of the central limit theorem. Ultimately, the proof hinges on a generalization of a famous limit from calculus.

11. Suppose that $a_n \rightarrow a$ as $n \rightarrow \infty$. Show that

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

Now let χ denote the characteristic function of the standard score of a sample variable X , and let χ_n denote the characteristic function of the standard score Z_n :

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

12. Show that

- a. $\chi(0) = 1$
- b. $\chi'(0) = 0$
- c. $\chi''(0) = -1$

13. Show that

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

14. Use properties of characteristic functions to show that

$$\chi_n(t) = \chi\left(\frac{t}{\sqrt{n}}\right)^n, \quad t \in \mathbb{R}$$

15. Use Taylor's theorem (named after **Brook Taylor**) to show that

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

16. In the context of previous exercise, show that $s_n \rightarrow 0$ and hence $\chi''(s_n) \rightarrow -1$ as $n \rightarrow \infty$.

17. Finally, show that

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

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 μ and variance $\frac{\sigma^2}{n}$. 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; although for many distributions smaller n will do.

18. 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:

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



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

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



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.

20. Suppose that X takes integer values and hence so does the partial sum Y_n . Show that for any $h \in (0, 1]$ and $k \in \mathbb{Z}$, the event $\{k - h < Y_n < k + h\}$ is equivalent to the event $\{Y_n = k\}$.

In the context of the previous exercise, 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 $h = 0.5$ for the normal approximation. This is sometimes called the **continuity correction**. The continuity correction is extended to other events in the natural way, using the additivity of probability.

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

22. 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, updating every 10 runs. Compute the following and compare with the result in the previous exercise:

- $\mathbb{P}(60 \leq Y \leq 75)$
- The relative frequency of the event $\{60 \leq Y \leq 75\}$

Normal Approximation to the Gamma Distribution

If Y_k has the **gamma distribution** with shape parameter $k \in \mathbb{N}_+$ and scale parameter $b \in (0, \infty)$ then

$$Y_k = \sum_{i=1}^k X_i$$

where (X_1, X_2, \dots, X_k) is a sequence of independent variables, each having the exponential distribution with scale parameter b . Since $\mathbb{E}(X_i) = b$ and $\text{var}(X_i) = b^2$, it follows that if k is large, the gamma distribution can be approximated by the normal distribution with mean $k b$ and variance $k b^2$. The same statement actually holds when k is not an integer; more precisely, the distribution of the standardized variable below converges to the standard normal distribution as $k \rightarrow \infty$.

$$Z_k = \frac{Y_k - k b}{\sqrt{k} b}$$

23. In the **gamma experiment**, vary k and b and note the shape of the probability density function. With $k = 10$ and $b = 2$, run the experiment 1000 times with an update frequency of 10 and note the apparent convergence of the empirical density function to the true density function.

24. 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:

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



Normal Approximation to the Chi-Square Distribution

The **chi-square distribution** with degrees of freedom $n \in \mathbb{N}_+$ is the gamma distribution with shape parameter $k = \frac{n}{2}$ and scale parameter $b = 2$. From the previous subsection, it follows that if n is large the chi-square distribution can be approximated by the normal distribution with mean n and variance $2n$. More precisely, if Y_n has the chi-square distribution with n degrees of freedom, then the distribution of the standardized variable below converges to the standard normal distribution as $n \rightarrow \infty$

$$Z_n = \frac{Y_n - n}{\sqrt{2n}}$$

25. In the **chi-square experiment**, vary n and note the shape of the density function. With $n = 20$, run the experiment 1000 times with an update frequency of 10 and note the apparent convergence of the empirical density function to the probability density function.

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

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



Normal Approximation to the Binomial Distribution

If Y_n has the **binomial distribution** with trial parameter $n \in \mathbb{N}_+$ and success parameter $p \in (0, 1)$, then

$$Y_n = \sum_{i=1}^n X_i$$

where (X_1, X_2, \dots, X_n) is a Bernoulli trials sequence with success parameter p , that is, a sequence of independent indicator variables with $\mathbb{P}(X_i = 1) = p$ for each i . It follows 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 $n(1-p) \geq 5$. More precisely, the distribution of the standardized variable Z_n given below converges to the standard normal distribution as $n \rightarrow \infty$:

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

27. 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, updating every 10 runs. Compute the following:

- $\mathbb{P}(12 \leq Y \leq 16)$
- The relative frequency of the event $\{12 \leq Y \leq 16\}$

28. Suppose that X 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.



Normal Approximation to the Poisson Distribution

If Y_n has the **Poisson distribution** with parameter $n \in \mathbb{N}_+$, then

$$Y_n = \sum_{i=1}^n X_i$$

where (X_1, X_2, \dots, X_n) is a sequence of independent variables, each with the Poisson distribution with parameter 1. Since $\mathbb{E}(X_i) = \text{var}(X_i) = 1$, 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 n is not an integer; more precisely, the distribution of the standardized variable below converges to the standard normal distribution as $n \rightarrow \infty$.

$$Z_n = \frac{Y_n - n}{\sqrt{n}}$$

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

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



30. 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 $r t$). Note the shape of the density function. With $r = 5$ and $t = 4$, run the experiment 1000 times with an update frequency of 10 and note the apparent convergence of the empirical density function to the probability density function.

Normal Approximation to the Negative Binomial Distribution

If Y_k has the **negative binomial distribution** with trial parameter $k \in \mathbb{N}_+$ and success parameter $p \in (0, 1)$ then

$$Y_k = \sum_{i=1}^k X_i$$

where (X_1, X_2, \dots, X_k) is a sequence of independent variables, each having the **geometric distribution** on \mathbb{N}_+ with success parameter p . Since $\mathbb{E}(X_i) = \frac{1}{p}$ and $\text{var}(X_i) = \frac{1-p}{p^2}$, it follows that if k is large, the negative binomial distribution can be approximated by the normal distribution with mean $\frac{k}{p}$ and variance $k \frac{1-p}{p^2}$. More precisely, the distribution of the standardized variable below converges to the standard normal distribution as $k \rightarrow \infty$.

$$Z_k = \frac{p Y_k - k}{\sqrt{k(1-p)}}$$

31. 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 with an update frequency of 10 and note the apparent convergence of the empirical density function to the true density function.

32. 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:

- $\mathbb{P}(20 \leq Y \leq 30)$
- The 80th percentile of Y .



Partial Sums with a Random Number of Terms

Suppose now that N is a random variable taking values in \mathbb{N} , with finite mean and variance. Then

$$Y_N = \sum_{i=1}^N X_i$$

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 X , the sequence of terms. Computing the moments of Y_N is a good exercise in [conditional expectation](#).

33. Show that

- $\mathbb{E}(Y_N|N) = N \mu$
- $\mathbb{E}(Y_N) = \mathbb{E}(N) \mu$

34. Show that

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

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

- $\mathbb{E}\left(e^{t Y_N} \middle| N\right) = G(t)^N$
- $\mathbb{E}\left(e^{t Y_N}\right) = H(G(t))$

Wald's Equation

Some of these results generalize to the case where the random number of terms N is a **stopping time** for the sequence 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 $n \in \mathbb{N}$.

36. Prove **Wald's equation** named after **Abraham Wald**: $\mathbb{E}(Y_N) = \mathbb{E}(N) \mu$.

- a. Show that $Y_N = \sum_{i=1}^{\infty} X_i \mathbf{1}(i \leq N)$
- b. Show that X_i and $\{i \leq N\}$ are independent for each i .
- c. Conclude that $\mathbb{E}(X_i \mathbf{1}(i \leq N)) = \mu \mathbb{P}(N \geq i)$
- d. Suppose that $X_i \geq 0$ for each i . Take expected values term by term in (a) to establish Wald's equation in this special case. The interchange of sum and expected value is justified by the monotone convergence theorem.
- e. Now establish Wald's equation in general by using the dominated convergence theorem