

4. The Chi-Square Distribution

In this section we will study a distribution that has special importance in statistics. In particular, this distribution will arise in the study of the [sample variance](#) when the underlying distribution is [normal](#) and in [goodness of fit tests](#).

The Density Function

For $n > 0$, the [gamma distribution](#) with shape parameter $k = \frac{n}{2}$ and scale parameter 2 is called the **chi-square distribution** with n **degrees of freedom**.

1. Show that the chi-square distribution with n degrees of freedom has probability density function

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

2. In the [random variable experiment](#), select the chi-square distribution. Vary n with the scroll bar and note the shape of the density function. For selected values of n , run the simulation 1000 times with an update frequency of 10 and note the apparent convergence of the empirical density function to the true density function.

3. Show that the chi-square distribution with 2 degrees of freedom is the exponential distribution with scale parameter 2.

4. Draw a careful sketch of the chi-square probability density function in each of the following cases:
- $0 < n < 2$.
 - $n = 2$. This is the probability density function of the exponential distribution.
 - $n > 2$. In this case, show that the mode occurs at $x = n - 2$

The [distribution function](#) and the [quantile function](#) do not have simple, closed-form representations. Approximate values of these functions can be obtained from the [table of the chi-square distribution](#), from the [quantile applet](#), and from most mathematical and statistical software packages.

5. In the [quantile applet](#), select the chi-square distribution. Vary the parameter and note the shape of the density function and the distribution function. In each of the following cases, find the median, the first and third quartiles, and the interquartile range.
- $n = 1$
 - $n = 2$
 - $n = 5$
 - $n = 10$



Moments

The [mean](#), [variance](#), moments, and [moment generating function](#) of the chi-square distribution can be obtained easily from general results for the gamma distribution. In the following exercises, suppose that X has the chi-square distribution with n degrees of freedom.

6. Show that

- a. $\mathbb{E}(X) = n$
- b. $\text{var}(X) = 2n$

7. Show that the moment of order $k > 0$ is

$$\mathbb{E}(X^k) = 2^k \frac{\Gamma(n/2 + k)}{\Gamma(n/2)}$$

8. Show that the moment generating function is given by

$$\mathbb{E}(e^{tX}) = \frac{1}{(1 - 2t)^{n/2}}, \quad t < \frac{1}{2}$$

9. In the simulation of the [random variable experiment](#), select the chi-square distribution. Vary n with the scroll bar and note the size and location of the mean/standard deviation bar. For selected values of n , run the simulation 1000 times with an update frequency of 10 and note the apparent convergence of the empirical moments to the distribution moments.

Transformations

10. Suppose that Z has the [standard normal distribution](#). Use [change of variable techniques](#) to show that $U = Z^2$ has the chi-square distribution with 1 degree of freedom.

11. Use moment generating functions or properties of the gamma distribution to show that if X has the chi-square distribution with m degrees of freedom, Y has the chi-square distribution with n degrees of freedom, and X and Y are independent, then $X + Y$ has the chi-square distribution with $m + n$ degrees of freedom.

12. Suppose that (Z_1, Z_1, \dots, Z_n) is a sequence of [independent](#) standard normal variables (that is, a [random sample](#) of size n from the standard normal distribution). Use the results of the previous two exercises to show that

$$V = \sum_{i=1}^n Z_i^2$$

has the chi-square distribution with n degrees of freedom.

The result of the last exercise is the reason that the chi-square distribution deserves a name of its own. Sums of squares of independent normal variables occur frequently in statistics. On the other hand, the following exercise shows that any gamma distributed variable can be re-scaled into a variable with a chi-square distribution.

13. Suppose that X has the gamma distribution with shape parameter k and scale parameter b . Show that $Y = \frac{2X}{b}$ has the chi-square distribution with $2k$ degrees of freedom.

14. Suppose that a missile is fired at a target at the origin of a plane coordinate system, with units in meters. The

missile lands at (X, Y) where X and Y are independent and each has the normal distribution with mean 0 and variance 100. The missile will destroy the target if it lands within 20 meters of the target. Find the probability of this event.



Normal Approximation

From the [central limit theorem](#), and previous results for the gamma distribution, it follows that if n is large, the chi-square distribution with n degrees of freedom can be approximated by the normal distribution with mean n and variance $2n$. More precisely, if X_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 $n \rightarrow \infty$

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

15. In the simulation of the [random variable experiment](#), select the chi-square distribution. Start with $n = 1$ and increase n . Note the shape of the density function. For selected values of n , 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.

16. Suppose that X has the chi-square distribution with $n = 18$ degrees of freedom. For each of the following, compute the true value using the [quantile applet](#) and then compute the normal approximation. Compare the results.

- $\mathbb{P}(15 < X < 20)$
- The 75th percentile of X .

