8.3.3 Confidence Intervals for Normal Samples

In the above discussion, we assumed n to be large so that we could use the CLT. An interesting aspect of the confidence intervals that we obtained was that they often did not depend on the details of the distribution from which we obtained the random sample. That is, the confidence intervals only depended on statistics such as $\overline{X}$ and S^2. What if n is not large? In this case, we cannot use the CLT, so we need to use the probability distribution from which the random sample is obtained. A very important case is when we have a sample X_1, X_2, X_3, $\ldots$, X_n from a normal distribution. Here, we would like to discuss how to find interval estimators for the mean and the variance of a normal distribution. Before doing so, we need to introduce two probability distributions that are related to the normal distribution. These distributions are useful when finding interval estimators for the mean and the variance of a normal distribution.

Chi-Squared Distribution

Let us remember the gamma distribution. A continuous random variable X is said to have a *gamma* distribution with parameters $\alpha > 0$ and $\lambda > 0$, shown as $X \sim Gamma(\alpha, \lambda)$, if its PDF is given by

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Now, we would like to define a closely related distribution, called the chi-squared distribution. We know that if Z_1, Z_2, $\cdots$, Z_n are independent standard normal random variables, then the random variable

$$X = Z_1 + Z_2 + \cdots + Z_n$$

is also normal. More specifically, $X \sim N(0, n)$. Now, if we define a random variable Y as

$$Y = Z_1^2 + Z_2^2 + \cdots + Z_n^2,$$

then Y is said to have a **chi-squared** distribution with n **degrees of freedom** shown by

$$Y \sim \chi^2(n).$$

It can be shown that the random variable Y has, in fact, a gamma distribution with parameters $\alpha = \frac{n}{2}$ and $\lambda = \frac{1}{2}$,

$$Y \sim Gamma\left(\frac{n}{2}, \frac{1}{2}\right).$$

Figure 8.5 shows the PDF of $\chi^2(n)$ distribution for some values of n.

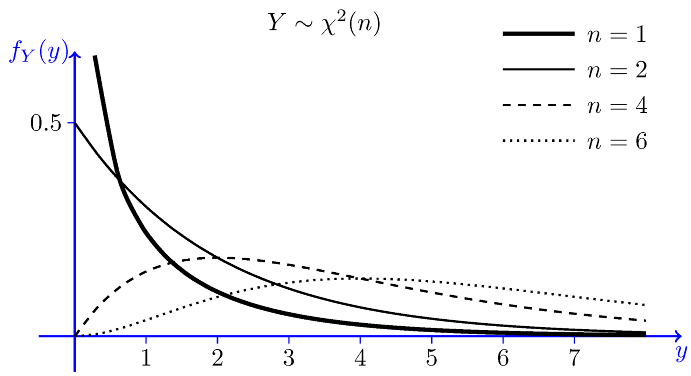

Figure 8.5 - The PDF of $\chi^2(n)$ distribution for some values of n.

So, let us summarize the definition and some properties of the chi-squared distribution.

The Chi-Squared Distribution

Definition 8.1. If Z_1, Z_2, $\cdots$, Z_n are independent standard normal random variables, the random variable Y defined as

$$Y = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

is said to have a *chi-squared* distribution with n *degrees of freedom* shown by

$$Y \sim \chi^2(n).$$

Properties:

1. The chi-squared distribution is a special case of the gamma distribution. More specifically,

$$Y \sim Gamma\left(\frac{n}{2}, \frac{1}{2}\right).$$

   Thus,

$$f_Y(y) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}, \quad \text{for } y > 0.$$

2. $EY = n$, $\mathrm{Var}(Y) = 2n$.

3. For any $p \in [0,1]$ and $n \in \mathbb{N}$, we define $\chi^2_{p,n}$ as the real value for which

$$P(Y > \chi^2_{p,n}) = p,$$

   where $Y \sim \chi^2(n)$. Figure 8.6 shows $\chi^2_{p,n}$. In MATLAB, to compute $\chi^2_{p,n}$ you can use the following command: $\mathtt{chi2inv(1-p,n)}$