CLARK UNIVERSITY

# The sample variance and the $\chi^2$ distribution

Math 218, Mathematical Statistics
D Joyce, Spring 2016

The *sample variance* for a sample $X_1, X_2, \ldots, X_n$ is sometimes defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

but sometimes as

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

We'll use the first, since that's what our text uses.

In the same way that the normal distribution is used in the approximation of means, a distribution called the $\chi^2$ distribution is used in the approximation of variances.

Let $Z_1, Z_2, \ldots, Z_\nu$ be $\nu$ independent standard normal variables. Then the sum of their squares

$$X = Z_1^2 + Z_2^2 + \cdots + Z_\nu^2$$

has what is called a $\chi^2$ *distribution with* $\nu$ *degrees of freedom*, written $\chi^2_\nu$, or more simply $\chi^2$ when $\nu$ is understood. The letter $\nu$ is the Greek letter nu, and it is often used for the number of degrees of freedom.

The main purpose of a $\chi^2$ distribution is its relation to the sample variance for a normal sample.

Suppose the sample $X_1, X_2, \ldots, X_n$ is from a normal distribution with mean $\mu$ and variance $\sigma^2$, then the sample variance $S^2$ is a scaled version of a $\chi^2$ distribution with $n-1$ degrees of freedom

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

The details of the proof are given at the end of section 5.2 of the text.

**The relation between $\chi^2$ distributions and Gamma distributions, and $\Gamma$ functions.** Recall that the Gamma distribution is one of the distributions that comes up in the Poisson process, the others being the exponential distribution and the Poisson distribution. A Poisson process is when events occur uniformly at random over time at a constant rate of $\lambda$ events per unit time. The time $T$ it takes for the $r^{\text{th}}$ event to occur has what is called a Gamma distribution with parameters $\lambda$ and $r$.

The gamma distribution also has applications when $r$ is not an integer. For that generality the factorial function is replaced by the gamma function, where

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt.$$

The gamma function and a related function called the beta function were invented by Euler in 1729. One of the main purposes of the gamma function is to generalize the factorial function to nonintegers. When $r$ is a nonnegative integer,

$$r! = \Gamma(r+1)$$

so that $1 = 0! = \Gamma(1)$, $1 = 1! = \Gamma(2)$, $2 = 2! = \Gamma(3)$, $6 = 3! = \Gamma(4)$, etc. In fact, $\Gamma(x)$ is defined except when $x = 0, -1, -2, -3, \ldots$. It's also defined for complex numbers. The gamma function has several properties including these

$$\begin{aligned}
\Gamma(x+1) &= x\Gamma(x) \\
\Gamma(x)\Gamma(1-x) &= \frac{\pi}{\sin \pi x} \\
\Gamma(x)\Gamma(x+\tfrac{1}{2}) &= 2^{1-2x}\sqrt{\pi}\,\Gamma(2x) \\
\Gamma(\tfrac{1}{2}) &= \sqrt{\pi}
\end{aligned}$$

It turns out that the $\chi^2$ with $\nu$ degrees of freedom is gamma distribution with a fractional value for $r$, namely $r = \nu/2$, and $\lambda = 1/2$. That implies that the density function for a $\chi^2_\nu$ distribution is

$$f(x) = \frac{x^{n/2-1}e^{x/2}}{2^{n/2}\Gamma(n/2)}, \text{ for } x \in [0, \infty),$$

its mean is $\mu = n$, and its variance is $\sigma^2 = 2n$.

Our text has a table of values for the $\chi^2$ distribution, Table A.5 on page 676.

We'll work through example 5.5 on page 178.