

Chi-Square Distribution

James W. Stoutenborough <jstout@ku.edu> and Paul E. Johnson <pauljohn@ku.edu>

June 10, 2013

1 Introduction

The Chi-Square distribution is a staple of statistical analysis. It is often used to judge “how far away” some number is from some other number.

The simplest place to start is that the Chi-Square distribution is what you get if you take observations from a standard Normal distribution and square them and add them up. If we use Z_1, Z_2 , and so forth to refer to draws from $N(0,1)$, then

$$Z_1^2 + Z_2^2 + \dots + Z_N^2 = \sum_{i=1}^N Z_i^2 \sim \chi_N^2$$

That’s means the sum of Z 's squared has a Chi-Square distribution with N degrees of freedom. The term “degrees of freedom” has some emotional and cognitive implications for psychologists, but it is really just a parameter for us.

Things that are sums of squares have χ^2 distributions.

Now, suppose the numbers being added up are not standardized, but they are centered. That is to say, they have a Normal distribution with a mean of 0 and a standard deviation of sd . That means we would have to divide each observation by sd in order to obtain the Z_i 's which are standardized Normal observations. Obviously,

$$\left(\frac{Y_1}{sd}\right)^2 + \left(\frac{Y_2}{sd}\right)^2 + \dots + \left(\frac{Y_N}{sd}\right)^2 = \frac{1}{sd^2} \sum_{i=1}^N Y_i^2 = \sum_{i=1}^N Z_i^2 \sim \chi_N^2$$

Equivalently, suppose you think of the Y_i as being proportional to the Z_i in this way:

$$Y_i = sd * Z_i$$

The coefficient sd is playing the role of a “scaling coefficient” and without too much effort you find out that if some variable $x_i = \sum Z_i^2$ has a Chi-square distribution, χ_N^2 , then $sd \times x_i$ has a distribution equal to $sd \times \chi_N^2$.

The elementary laws of expected values and variances dictate that

$$E(sd \times x_i) = sd * E(x_i)$$

and

$$Var(sd \times x_i) = sd^2 Var(x_i)$$

In other words, the Chi-square distribution applies not just for a sum of squares of a standardized Normal distribution, but in fact it describes a sum of squares of any Normal distribution that is centered around zero.

2 Mathematical Description

The Chi-Square probability density function for $x_i = \sum Z_i^2$ is defined as:

$$f(x_i) = \frac{x_i^{(N/2-1)} \exp(-x_i/2)}{2^{N/2} \Gamma[N/2]}$$

It is defined on a range of positive numbers, $0 \leq x_i \leq \infty$. Because we are thinking of this value as a sum of squared values, it could not possibly be smaller than zero. It also assumes that $N > 0$, which is obviously true because we are thinking of the variable as a sum of N squared items.

Why does the χ^2 have that functional form? Well, write down the probability model for a standardized Normal distribution, and then realize that the probability of a squared-value of that standardized Normal is EXTREMELY easy to calculate if you know a little bit of mathematical statistics. The only “fancy” bit is that this formula uses our friend the Gamma Function (see my handout on the Gamma distribution), to represent a factorial. But we have it on good authority (Robert V. Hogg and Allen T. Craig, Introduction to Mathematical Statistics, 4ed, New York: Macmillan, 1978, p. 115) that $\Gamma(1/2) = \sqrt{\pi}$.

3 Illustrations

The probability density function of the Chi-Square distribution changes quite a bit when one puts in different values of the parameters. If somebody knows some “interesting” parameter settings, then a clear, beautiful illustration of the Chi-square can be produced. Consider the following code, which can be used to create the illustration of 2 possible Chi-Square density functions in Figure 1.

```
xvals <- seq(0,10,length.out=1000)
chisquare1 <- dchisq(xvals, df=1)
chisquare2 <- dchisq(xvals, df=6)
matplot(xvals, cbind(chisquare1, chisquare2), type="l", xlab="
  possible values of x", ylab="probability of x", ylim=c(0,1),
  main="Chi-Square Probability Densities")
text(.4, .9, "df=1", pos=4, col=1)
text(4, .2, "df=6", pos=4, col=2)
```

The shape of the Chi-Square is primarily dependent upon the degrees of freedom that are witnessed in any particular univariate analysis. The adjustment of the degrees of freedom will have a substantial impact on the shape of the distribution. The following code will produce example density functions for a variety of shapes with a variety of degrees of freedom. Examples of Chi-Square density function with a variety of degrees of freedom are found in Figure 2.

Chi-Square Probability Densities

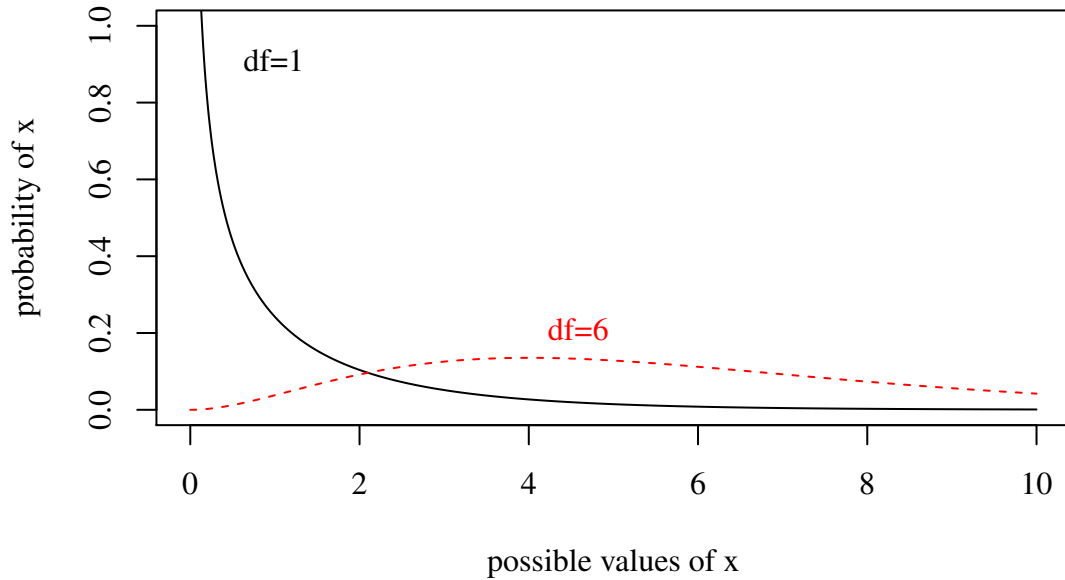


Figure 1: χ^2 Density Functions

4 Expected Value, Variance, and the role of the parameters

The Chi-Square distribution is a form of the Gamma distribution, and most treatments of the Chi-Square rely on the general results about the Gamma to state the characteristics of the special-case Chi-square. The Gamma distribution $G(\alpha, \beta)$ is a two parameter distribution, with parameters shape (α) and scale (β).

$$\text{Gamma probability density} = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

Note that if the shape parameter of a Gamma distribution is $\frac{N}{2}$ and the scale parameter is equal to 2, then this probability density is identical to the Chi-square distribution with degrees of freedom equal to N .

Since it is known that the expected value of a Gamma distribution is $\alpha\beta$ and the variance is $\alpha\beta^2$, that means that the expected value of a Chi-square for N observations is

$$E(x) = N$$

and the variance of a Chi-square variable is

$$\text{Var}(x) = 2N$$

Chi-Square Probability Densities

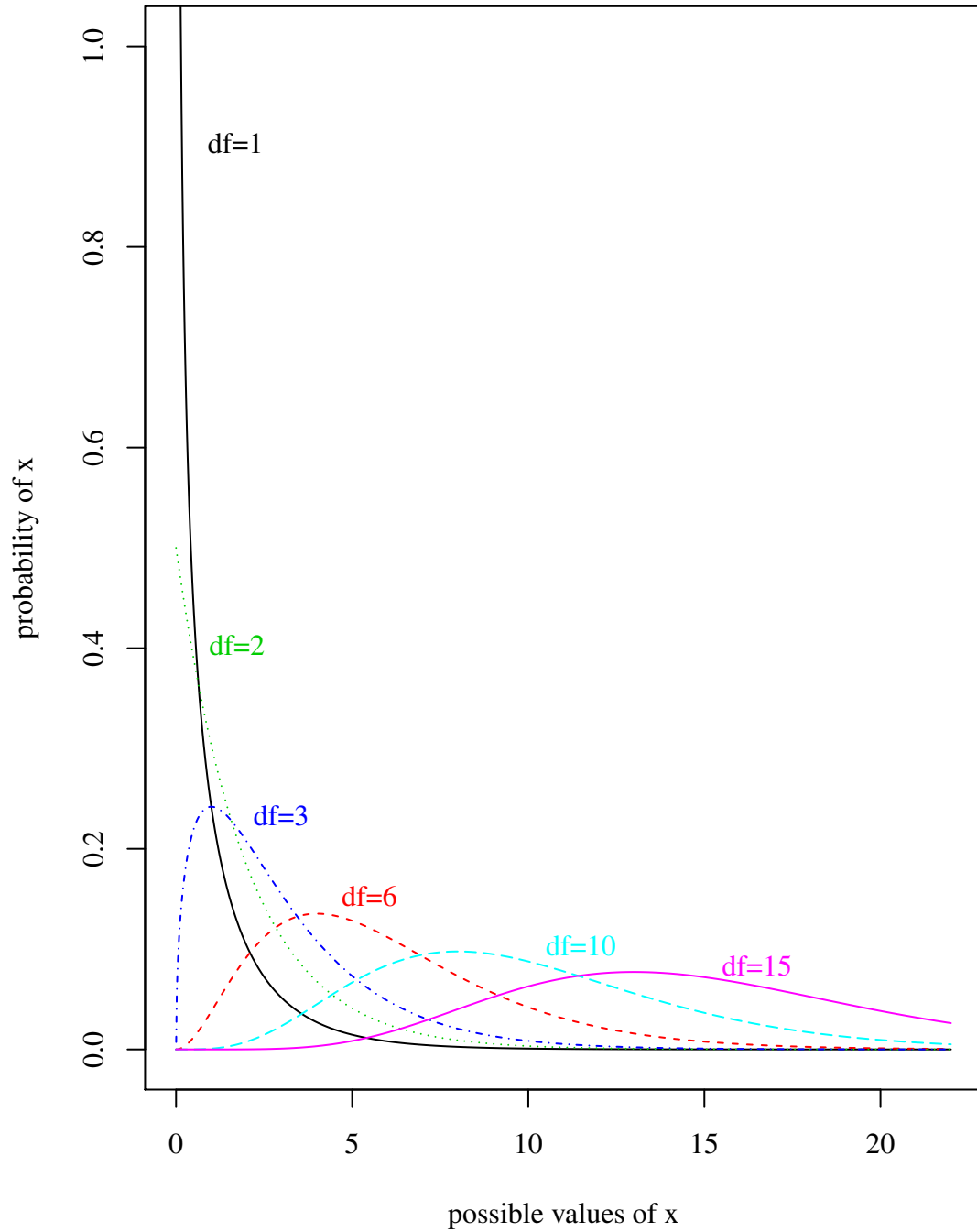


Figure 2: Chi-Square Densities various df

Now, if a variable is proportional to a Chi-Square x_i , $y_i = \sigma x_i$, we know that y_i has a distribution

$$y_i \sim \sigma \chi_N^2$$

and the probability density is (via a “change of variables”)

$$f(y_i) = \frac{y_i^{\left(\frac{N}{2}-1\right)} \exp\left(-\frac{y_i}{2\sigma}\right)}{\sigma^{N/2} 2^{N/2} \Gamma[N/2]}$$

and

$$E(y_i) = \sigma N$$

$$Var(y_i) = \sigma^2 N$$

The mode (for $N > 2$) is

$$mode(y_i) = \sigma(N - 2)$$

The Chi-Square is related to the Poisson distributions with parameter and expected value equal to $\frac{x_i}{2}$ by:

$$P[Chi - Square(n) \geq x_i] = P\left[Poisson\left(\frac{x_i}{2}\right) \leq \frac{n}{2} - 1\right]$$

5 How is this useful in Bayesian analysis?

In statistical problems, we often confront 2 kinds of parameters. The “slope coefficients” of a regression model are one type, and we usually have priors that are single-peaked and symmetric. The prior for such a coefficient might be Uniform, Normal, or any other mathematically workable distribution.

Sometimes other coefficients are not supposed to be symmetrical. For example, the variance of a distribution cannot be negative, so we need a distribution that is shaped to have a minimum at zero. The Gamma, or its special case the Chi-square, is an obvious candidate.

The most important aspect of the Chi-square, however, is that it is very mathematically workable! If one is discussing a Normal distribution, for example, $N(\mu, \sigma^2)$ one must specify prior beliefs about the distributions of μ and σ^2 . Recall that in Bayesian updating, we calculate the posterior probability as the product of the likelihood times the prior, so some formula that makes that result as simple as possible would be great.

$$p(\sigma^2|y) = p(y|\sigma^2)p(\sigma^2)$$

From the story that we told about where Chi-square variables come from, it should be very obvious that if y is normal, we can calculate $p(y|\sigma^2)$ (assuming μ is taken as “given” for the moment). So all we need is a prior that makes $p(\sigma^2|y)$ as simple as possible. If you choose $p(\sigma^2)$ to be Chi-squared, then it turns out to be very workable.

Suppose you look at the numerator from the Chi-Square, and “guess” that you want to put $1/\sigma^2$ in place of x_i . You describe your prior opinion about σ^2

$$\text{prior} : p(\sigma^2) \propto (\sigma^2)^{-N/2-1} \exp\left(-\frac{1}{2}S_0/\sigma^2\right)$$

We use N and S_0 as a “scaling factors” to describe how our beliefs vary from one situation to another. N is the “degrees of freedom”.

Note that is very convenient if your Normal theory for y says:

$$p(y_i|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(y_i - \mu)^2}{\sigma^2}\right)$$

Suppose the sample size of the dataset is n . If you let $S = \sum_{i=1}^n (y_i - \mu)^2$ represent the sum of squares, then we rearrange to find a posterior:

$$p(\sigma^2|y) \propto (\sigma^2)^{-(N+n)/2-1} \exp\left(-\frac{1}{2}(S_0 + S)/\sigma^2\right)$$

Look how similar the prior is to the posterior.

It gets confusing discussing σ^2 and $1/\sigma^2$. Bayesians don't usually talk about estimating the variance of σ^2 , but rather the precision, which is defined as

$$\tau = \frac{1}{\sigma^2}$$

Hence, the distribution of the “precision” is given as a Chi-Square variable, and if your prior is

$$\text{prior} : p(\tau) \propto \tau^{N/2-1} \exp\left(-\frac{1}{2}S_0\tau\right)$$

then the posterior is a Chi-Square variable

$$(S_0 + S)\tau \sim \chi_{N+n}^2$$

If you really do want to talk about the variance, rather than the precision, then you are using a prior that is an INVERSE Chi-Square. Your prior is the inverse chi-square

$$\sigma^2 \sim S_0 X_N^{-2}$$

which I've seen referred to as

$$\sigma^2 \sim (S_0 + S)\chi_{N+n}^{-2}$$

As a result, a prior for a variance parameter is often given as an inverse Chi-square, while the prior for a precision parameter is given as a Chi-square.