Chi-Squared Distribution

Fall 2016

A New Distribution

A distribution that is used frequently in methods of estimation and hypothesis testing is a distribution called the Chi-squared distribution.

This distribution is not typically used for modeling data.

The Chi-squared distribution is a special case of a distribution called the Gamma distribution which is used extensively for modeling lifetime data.

Before introducing the Chi-squared distribution we recall a special function from STAT 230 - the Gamma Function.

The Gamma Function

Definition

The Gamma function is defined as

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \qquad \alpha > 0$$

Properties of the Gamma function:

1) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$

2) $\Gamma(\alpha) = (\alpha - 1)!$ for $\alpha = 1, 2, \ldots$

3) $\Gamma(1/2) = \sqrt{\pi}$

The Chi-squared Distribution

Suppose X is a random variable with probability density function

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \quad \text{for } x > 0 \quad \text{and} \quad k = 1, 2, \ldots$$

then X is said to have a Chi-squared distribution with parameter k which is usually called the "degrees of freedom".

We write $X \backsim \chi^2(k)$ and read this as X has a Chi-squared distribution on k degrees of freedom (df).

Chi-squared Probability Density Function Shapes

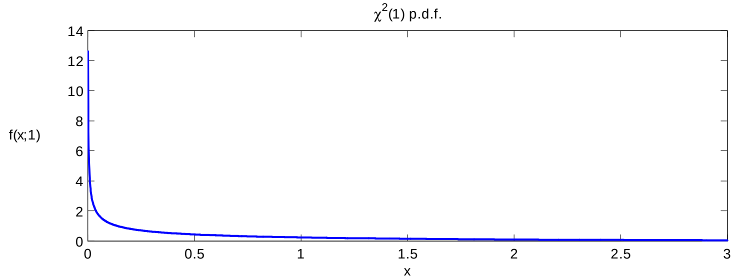
For $k > 2$ the probability density function

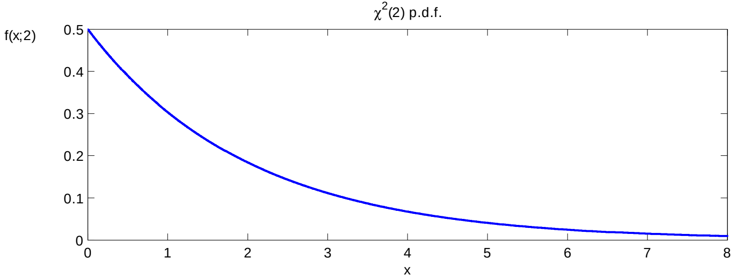
$$f(x; k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} \quad \text{for } x > 0 \quad \text{and} \quad k = 1, 2, \ldots$$

is unimodal with a global maximum at $x = k - 2$.

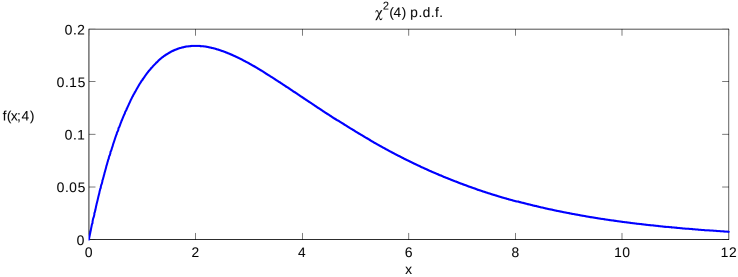
(Can you show this?)

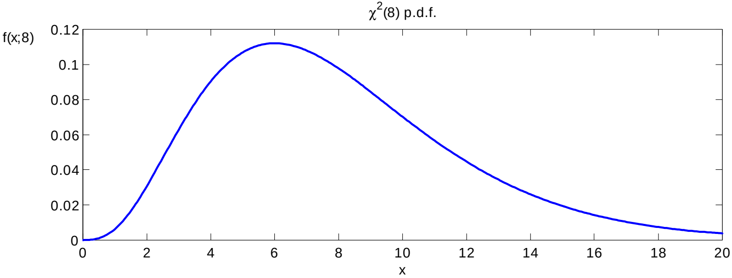
Chi-squared probability density functions for k=1 and k=2



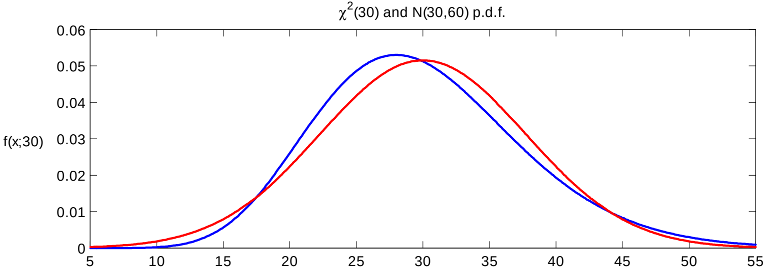


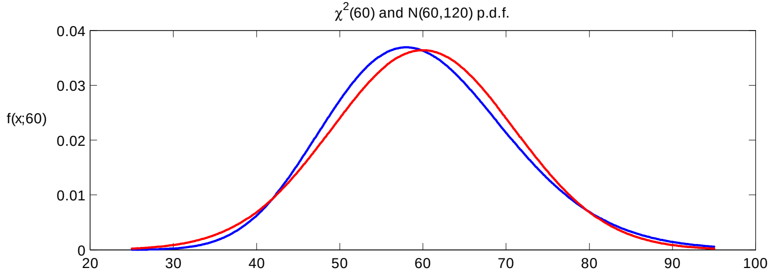
Chi-squared probability density functions for k=4 and k=8





Chi-squared p.d.f.'s for k=30 and k=60 (Normal in red)





The case k=2 and the Exponential Distribution

For $k = 2$ the probability density function is given by

$$f(x;2) = \frac{1}{2}e^{-x/2} \quad \text{for } x > 0$$

which we recognize as an Exponential(2) probability density function.

Therefore if $X \backsim \chi^2(2)$ we can evaluate probabilities for X using

$$P(X \leq x) = \int_0^x \frac{1}{2}e^{-u/2}du = 1 - e^{-x/2} \quad \text{for } x > 0$$

and

$$P(X > x) = 1 - P(X \leq x) = e^{-x/2} \quad \text{for } x > 0$$

Mean and Variance of a Chi-squared Random Variable

If $X \backsim \chi^2(k)$ then by using the method of substitution (change of variable) and the Gamma function it can be shown that

$$E(X) = k$$

and

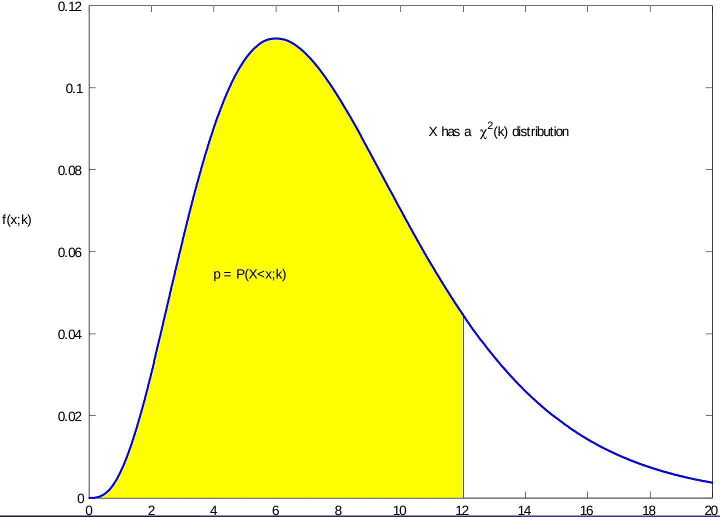
$$Var(X) = 2k.$$

Normal Approximation to the Chi-squared Distribution:

If $k \geq 30$ and $X \backsim \chi^2(k)$ then $X \backsim N(k, 2k)$ approximately.

Chi-squared Table - Learn How to Read

Chi-squared table gives the value p for $p = P(X \leq x; k)$ where $X \backsim \chi^2(k)$.



Determining Chi-squared Values

Chapter 4, Problem 17

Important problem to do before Tutorial Test 2!

Distribution of a Sum of Independent Chi-squared Random Variables

Theorem

Suppose $W_1, W_2, \ldots, W_n$ *are independent random variables and* $W_i \backsim \chi^2(k_i)$, $i = 1, 2, \ldots, n$.

Then

$$\sum_{i=1}^{n} W_i \backsim \chi^2\left(\sum_{i=1}^{n} k_i\right)$$

Relationship between the Normal and Chi-squared Distributions

Theorem

If $Z \backsim N(0,1)$ *then* $W = Z^2 \backsim \chi^2(1)$.

Relationship between the Normal and Chi-squared Distributions

Theorem

If $Z \backsim N(0,1)$ *then* $W = Z^2 \backsim \chi^2(1)$.

Proof.

The cumulative distribution function of W is

$$P(W \leq w) = P(Z^2 \leq w)$$

$$\begin{aligned} &= P(-\sqrt{w} \leq Z \leq \sqrt{w}) \\ &= \Phi(-\sqrt{w}) - \Phi(\sqrt{w}) \\ &= 2\Phi(\sqrt{w}) - 1 \end{aligned}$$

where Φ is the c.d.f. of a $N(0,1)$ random variable. $\square$

Proof Cont'd

Proof.

Let $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ which is the p.d.f. of a $N(0,1)$ random variable. Then the p.d.f. of W is

$$\begin{aligned} & \frac{d}{dw}[2\Phi(\sqrt{w}) - 1] \\ = \; & 2\phi(\sqrt{w})\left(\frac{1}{2}\right) w^{-1/2} \\ = \; & \frac{1}{\sqrt{2\pi}} w^{-1/2} e^{-1/2} \quad \text{for } w > 0 \end{aligned}$$

which is the p.d.f. of a $\chi^2(1)$ random variable as required. □

Useful Results

This theorem also gives us a way to calculate probabilities for the random variable $W \backsim \chi^2(1)$ using Normal tables since

$$P(W \leq w) = P(|Z| \leq \sqrt{w}) = 2P(Z \leq \sqrt{w}) - 1$$

where $Z \backsim N(0, 1)$.

Also

$$P(W > w) = P(|Z| > \sqrt{w}) = 2P(Z > \sqrt{w}) = 2[1 - P(Z \leq \sqrt{w})].$$

Distribution of a Sum of Independent N(0,1) Random Variables Squared

Using the previous two theorems we have the following result:

Theorem

If $Z_1, Z_2, \ldots, Z_n$ are independent and identically distributed N(0,1) random variables then

$$S = \sum_{i=1}^{n} Z_i^2 \backsim \chi^2(n)$$

Distribution of a Sum of Independent Standardized Normal Random Variables Squared

Corollary

If $X_1, X_2, \ldots, X_n$ are independent and identically distributed $N(\mu, \sigma^2)$ random variables then

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \backsim \chi^2(n)$$