The Chi-Squared Distribution

There is a special gamma distribution:

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

The Chi-Squared Distribution

There is a special gamma distribution:

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

$$f_X(x) = \frac{1}{\Gamma(n/2)}\left(\frac{1}{2}\right)^{n/2} x^{n/2-1}\, e^{-\frac{1}{2}x}\, I_{(0,\infty)}(x)$$

We write $X \sim \chi^2(n)$.

The Chi-Squared Distribution

There is a special gamma distribution:

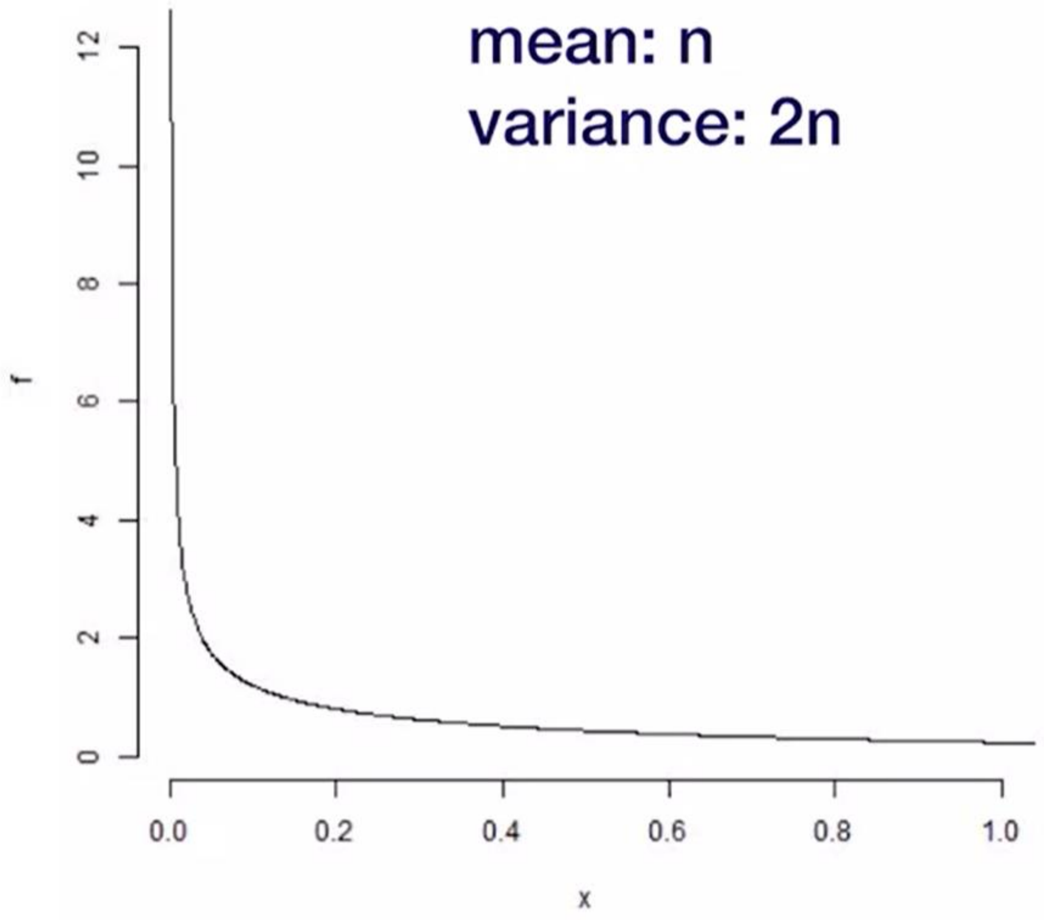
$$X \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

$$f_X(x) = \frac{1}{\Gamma(n/2)}\left(\frac{1}{2}\right)^{n/2} x^{n/2-1}\, e^{-\frac{1}{2}x}\, I_{(0,\infty)}(x)$$

We write $X \sim \chi^2(n)$.

"A chi-squared distribution with n degrees of freedom."

The Chi-Squared Distribution



mean: n
variance: 2n

Things About $\chi^2(n)$

$$X \sim \Gamma(\alpha, \beta)$$

$$\Rightarrow \quad M_X(t) = \left(\frac{\beta}{\beta - t}\right)^{\alpha}$$

Things About $\chi^2(n)$

$$X \sim \Gamma(\alpha, \beta)$$

$$\Rightarrow \quad M_X(t) = \left(\frac{\beta}{\beta - t}\right)^{\alpha}$$

So,

$$X \sim \chi^2(n) = \Gamma(n/2, 1/2)$$

$$\Rightarrow \quad M_X(t) = \left(\frac{1/2}{1/2 - t}\right)^{n/2}$$

Things About $\chi^2(n)$

Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables with $X_i \sim \chi^2(n_i)$.

$$\text{Let } Y = \sum_{i=1}^{k} X_i$$

$$M_Y(t) = E[e^{tY}] = E\left[e^{t\sum_{i=1}^{k} X_i}\right]$$

$$= E\left[\prod_{i=1}^{k} e^{tX_i}\right] \overset{\text{indep}}{=} \prod_{i=1}^{k} E\left[e^{tX_i}\right]$$

Things About $\chi^2(n)$

Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables with $X_i \sim \chi^2(n_i)$. Let $Y = \sum_{i=1}^{k} X_i$

$$M_Y(t) = \prod_{i=1}^{k} E\left[e^{tX_i}\right]$$

Things About $\chi^2(n)$

Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables with $X_i \sim \chi^2(n_i)$. Let $Y = \sum_{i=1}^{k} X_i$

$$M_Y(t) = \prod_{i=1}^{k} E\left[e^{tX_i}\right] = \prod_{i=1}^{k} M_{X_i}(t)$$

$$= \prod_{i=1}^{k} \left(\frac{1/2}{1/2 - t}\right)^{n_i/2} = \left(\frac{1/2}{1/2 - t}\right)^{\sum_{i=1}^{k} n_i/2}$$

A Chi-Squared Property

Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables with $X_i \sim \chi^2(n_i)$.

Let $Y = \sum_{i=1}^{k} X_i$

$$M_Y(t) = \prod_{i=1}^{k} \left(\frac{1/2}{1/2 - t} \right)^{\sum_{i=1}^{n} n_i/2}$$

A Chi-Squared Property

Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables with $X_i \sim \chi^2(n_i)$. Let $Y = \sum_{i=1}^{k} X_i$

$$M_Y(t) = \prod_{i=1}^{k} \left(\frac{1/2}{1/2 - t} \right)^{\sum_{i=1}^{n} n_i/2}$$

$$\Rightarrow \quad Y \sim \chi^2(n_1 + n_2 + \cdots + n_k)$$

The sum of chi-squareds is chi-squared!

The Chi-Squared Normal Relationship

Let $X \sim N(0, 1)$.

Let $Y = X^2$.

The Chi-Squared Normal Relationship

Let $X \sim N(0, 1)$.

Let $Y = X^2$.

Then $Y \sim \chi^2(1)$.

We could show this using moment generating functions, as described in Module 1,Lesson 8.

Recall from Module 1, Lesson 4, the pdf of the transformation $Y = g(x)$ is

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

Recall from Module 1, Lesson 4, the pdf of the transformation $Y = g(x)$ is

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

There is a bivariate version of this to go from X_1 and X_2 to $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$.

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)) \cdot |J|$$

where

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

The t-Distribution

Let $Z \sim N(0,1)$ and $W \sim \chi^2(n)$ be independent random variables..

Define

$$T = \frac{Z}{\sqrt{W/n}}$$

The t-Distribution

Let $Z \sim N(0,1)$ and $W \sim \chi^2(n)$ be independent random variables..

Define
$$T = \frac{Z}{\sqrt{W/n}}$$

Do the "Jacobian transformation" with

$$X_1 = Z \qquad Y_1 = \frac{X_1}{\sqrt{X_2/n}}$$
$$X_2 = W$$
$$Y_2 = g_2(X_1, X_2)$$

The t-Distribution

Can show that T has pdf

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

The t-Distribution

Can show that T has pdf

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

mean: 0

variance: $\dfrac{n}{n-2}$

We write $T \sim t(n)$.

The t-Distribution

Can show that T has pdf

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

mean: 0

variance: $\frac{n}{n-2}$

We write $T \sim t(n)$.

"A t distribution with n degrees of freedom."

The t-Distribution

Can show that T has pdf

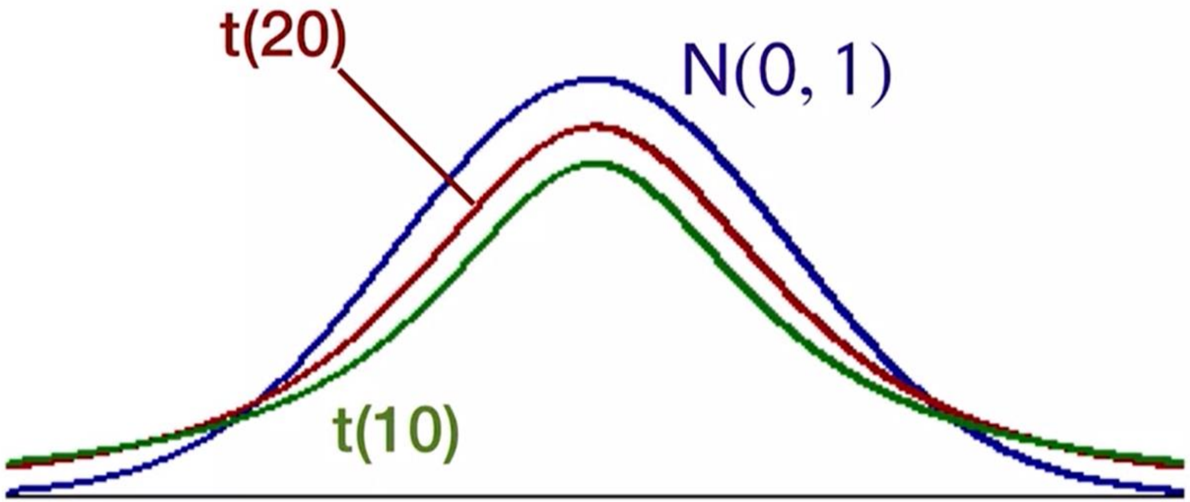
$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

mean: 0

variance: $\dfrac{n}{n-2}$

We write $T \sim t(n)$.

"A t distribution with n degrees of freedom."


t(20)
N(0, 1)
t(10)

Why now? Suppose that

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

Why now? Suppose that

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

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

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

$$+ n(\overline{X} - \mu)^2$$

Why now? Suppose that

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$$

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

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

$$+ n(\overline{X} - \mu)^2$$

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

$$= \sum_{i=1}^{n} X_i - n\frac{1}{n}\sum_{i=1}^{n} X_i = 0$$

Why now?

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

Divide through by σ^2.

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}+\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

Why now?

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

Divide through by σ^2.

$$\underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{Y_1}=\underbrace{\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}}_{Y_2}+\underbrace{\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2}_{Y_3}$$

$$X_i \sim N(\mu, \sigma^2)$$

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

$$X_i \sim N(\mu, \sigma^2)$$

$$\Downarrow$$

$$\frac{X_i - \mu}{\sigma} \sim N(0, 1)$$

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

$$X_i \sim N(\mu, \sigma^2)$$

$$\Downarrow$$

$$\frac{X_i - \mu}{\sigma} \sim N(0, 1)$$

$$\Downarrow$$

$$\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(1)$$

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

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

$$\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(1)$$

$$\Downarrow$$

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

Why now?

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

Divide through by σ^2.

$$\underbrace{\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2}_{Y_1} = \underbrace{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{\sigma^2}}_{Y_2} + \underbrace{\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \right)^2}_{Y_3}$$

$Y_1 = Y_2 + Y_3$

- $Y_1 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$
- $Y_2 = \frac{(n-1)S^2}{\sigma^2} \sim ?$
- $Y_3 = \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \sim \chi^2(1)$

$$Y_1 = Y_2 + Y_3$$

- $Y_1 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$
- $Y_2 = \frac{(n-1)S^2}{\sigma^2} \sim ?$
- $Y_3 = \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \sim \chi^2(1)$

independent

???

$$Y_1 = Y_2 + Y_3$$

$$M_{Y_1}(t) = M_{Y_2+Y_3}(t)$$

$$\overset{\text{indep}}{=} M_{Y_2}(t) \cdot M_{Y_3}(t)$$

$$\Rightarrow \; M_{Y_2}(t) = \frac{M_{Y_1}(t)}{M_{Y_3}(t)}$$

$$M_{Y_2}(t) = \left(\frac{1/2}{1/2 - t} \right)^{n-1}$$

$$M_{Y_2}(t) = \left(\frac{1/2}{1/2 - t} \right)^{n-1}$$

$$\Rightarrow \quad Y_2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

for $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$.