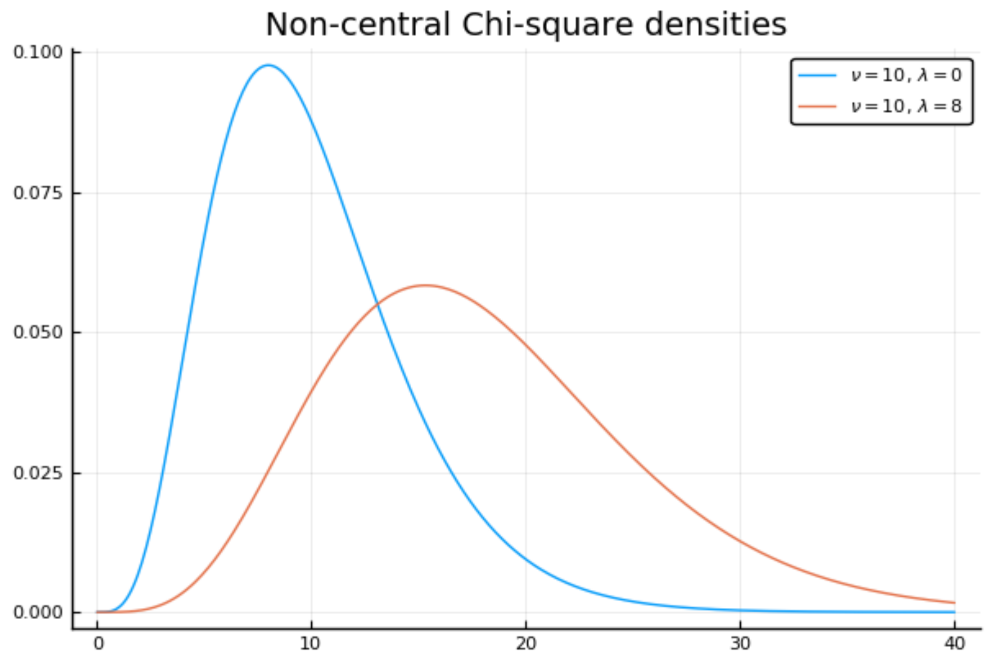
Distribución Ji-cuadrada no central

In [23]:
```
using WebIO, Interact, Distributions, Plots, LaTeXStrings
pyplot();
```

In [2]:
```
xrange = collect(0:0.1:40)
plot( xrange, pdf.( NoncentralChisq(10, 0.0), xrange), label=L"\nu=10,\, \lambda=
plot!(xrange, pdf.( NoncentralChisq(10, 8.0), xrange), label=L"\nu=10,\, \lambda=
title!("Non-central Chi-square densities")
```

Out[2]:



In [3]:
```
function plotNoncentralChisqPdf(xrange,ν,λ)
    yrange = pdf.( NoncentralChisq(ν, λ), xrange)
    plot(xrange,yrange,label=latexstring("\$ \\nu = \$ $(ν), \$ \\lambda = \$ $(λ
end
```

Out[3]: plotNoncentralChisqPdf (generic function with 1 method)

In [4]:
```
function interactiveNoncentralChisqPdf(xrange,νrange,λrange)
    νslider = slider( νrange, label="ν")
    λslider = slider( λrange, label="λ")
    vslider = [νslider,λslider]
    display(νslider)
    display(λslider)
    map((x...)->plotNoncentralChisqPdf( xrange, x[1], x[2]), vslider...)
end
```

Out[4]: interactiveNoncentralChisqPdf (generic function with 1 method)

In [8]:
```
interactiveNoncentralChisqPdf(-0.1:0.001:40 , 1:1:10, 0.0:0.5:10.0 )
```

Out[8]:

Ejemplos relacionados al modelo de regresión lineal múltiple con errores normales

En las notas sobre el modelo de regresión lineal múltiple con errores normales, en la sección refrente al coeficiente de determinación R^2, vimos que

$$SSR = \hat{\boldsymbol{\beta}}_1' \boldsymbol{X}_c' \boldsymbol{X}_c \hat{\boldsymbol{\beta}}_1 = \boldsymbol{y}' \boldsymbol{X}_c \left(\boldsymbol{X}_c' \boldsymbol{X}_c\right)^{-1} \boldsymbol{X}_c' \boldsymbol{X}_c \left(\boldsymbol{X}_c' \boldsymbol{X}_c\right)^{-1} \boldsymbol{X}_c' \boldsymbol{y} = \boldsymbol{y}' \boldsymbol{X}_c \left(\boldsymbol{X}_c' \boldsymbol{X}_c\right)^{-1} \boldsymbol{X}_c' \boldsymbol{y}$$

y en clase hemos visto que

$$SST = \boldsymbol{y}' \boldsymbol{y} - n\bar{y}^2 = \boldsymbol{y}' \boldsymbol{y} - \frac{1}{n} \boldsymbol{y}' \boldsymbol{j} \boldsymbol{j}' \boldsymbol{y} = \boldsymbol{y}' \left(\boldsymbol{I} - \frac{1}{n} \boldsymbol{J} \right) \boldsymbol{y}$$

Por lo que de $SST = SSR + SSE$ se sigue que

$$SSE = \boldsymbol{y}' \left(\boldsymbol{I} - \frac{1}{n} \boldsymbol{J} - \boldsymbol{X}_c \left(\boldsymbol{X}_c' \boldsymbol{X}_c\right)^{-1} \boldsymbol{X}_c' \right) \boldsymbol{y}$$

En clase vimos el siguiente resultado

Teorema Si $\boldsymbol{y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ entonces $\frac{1}{\sigma^2} \boldsymbol{y}' \boldsymbol{A} \boldsymbol{y} \sim \chi^2(r, \frac{1}{2} \boldsymbol{\mu}' \boldsymbol{A} \boldsymbol{\mu})$ si y sólo si $\boldsymbol{A\Sigma}$ es idempotente de rango r.

Observemos que

$$\left(\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\right)^2 = \boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c' = \boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'$$

por lo que $\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'$ es idempotente y su rango, que coincide con la traza, es k. Recordemos que las columnas de $\boldsymbol{X}_c$ suman cero por lo que $\boldsymbol{X}_c'\boldsymbol{j} = \boldsymbol{0}$ y observemos que en el modelo centrado $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{y}] = (\boldsymbol{j}, \boldsymbol{X}_c)\begin{pmatrix}\alpha\\ \boldsymbol{\beta}_1\end{pmatrix}$ por lo que

$$\begin{aligned}
\lambda_{SSR} &= \frac{1}{2\sigma^2}\left(\alpha, \boldsymbol{\beta}_1'\right)\begin{pmatrix}\boldsymbol{j}'\\ \boldsymbol{X}_c'\end{pmatrix}\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\left(\boldsymbol{j}, \boldsymbol{X}_c\right)\begin{pmatrix}\alpha\\ \boldsymbol{\beta}_1\end{pmatrix}\\
&= \frac{1}{2\sigma^2}\left(\alpha, \boldsymbol{\beta}_1'\right)\begin{pmatrix}\left(\boldsymbol{X}_c'\boldsymbol{j}\right)'\\ \boldsymbol{X}_c'\boldsymbol{X}_c\end{pmatrix}\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\left(\boldsymbol{X}_c'\boldsymbol{j}, \boldsymbol{X}_c'\boldsymbol{X}_c\right)\begin{pmatrix}\alpha\\ \boldsymbol{\beta}_1\end{pmatrix}\\
&= \frac{1}{2\sigma^2}\left(\alpha, \boldsymbol{\beta}_1'\right)\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{X}_c'\boldsymbol{X}_c\end{pmatrix}\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\left(\boldsymbol{0}, \boldsymbol{X}_c'\boldsymbol{X}_c\right)\begin{pmatrix}\alpha\\ \boldsymbol{\beta}_1\end{pmatrix}\\
&= \frac{1}{2\sigma^2}\boldsymbol{\beta}_1'\boldsymbol{X}_c'\boldsymbol{X}_c\boldsymbol{\beta}_1\\
&= \lambda_1
\end{aligned}$$

Por lo que concluimos que

$$\frac{\boldsymbol{y}'\left(\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\right)\boldsymbol{y}}{\sigma^2} = \frac{\mathrm{SSR}}{\sigma^2} \sim \chi^2(k, \lambda_1)$$

Por otro lado se tiene que

$$\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\boldsymbol{J} = \boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\boldsymbol{j}\boldsymbol{j}' = \boldsymbol{0}$$

Por lo cual se obtiene que $\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J} - \boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\right)$ es idempotente y su rango, que coincide con la traza, es $n - 1 - k = n - (k+1)$. Trabajando con el modelo no centrado y recordando que $\boldsymbol{X}_c = \left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\boldsymbol{X}_1$ se tiene

$$\begin{aligned}
\lambda_{SSE} &= \frac{1}{2\sigma^2}\left(\beta_0, \boldsymbol{\beta}_1'\right)\begin{pmatrix}\boldsymbol{j}'\\ \boldsymbol{X}_1'\end{pmatrix}\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\left(\boldsymbol{j}, \boldsymbol{X}_1\right)\begin{pmatrix}\beta_0\\ \boldsymbol{\beta}_1\end{pmatrix} - \lambda_{SSR}\\
&= \frac{1}{2\sigma^2}\left(\beta_0, \boldsymbol{\beta}_1'\right)\begin{pmatrix}\boldsymbol{j}'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\boldsymbol{j}\\ \boldsymbol{X}_1'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\boldsymbol{X}_1\end{pmatrix}\begin{pmatrix}\beta_0\\ \boldsymbol{\beta}_1\end{pmatrix} - \lambda_1\\
&= \frac{1}{2\sigma^2}\left(\beta_0, \boldsymbol{\beta}_1'\right)\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{X}_1'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\boldsymbol{X}_1\end{pmatrix}\begin{pmatrix}\beta_0\\ \boldsymbol{\beta}_1\end{pmatrix} - \lambda_1\\
&= \frac{1}{2\sigma^2}\left(\beta_0, \boldsymbol{\beta}_1'\right)\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{X}_c'\boldsymbol{X}_c\end{pmatrix}\begin{pmatrix}\beta_0\\ \boldsymbol{\beta}_1\end{pmatrix} - \lambda_1\\
&= \frac{1}{2\sigma^2}\boldsymbol{\beta}_1'\boldsymbol{X}_c'\boldsymbol{X}_c\boldsymbol{\beta}_1 - \lambda_1\\
&= \lambda_1 - \lambda_1\\
&= 0
\end{aligned}$$

Concluimos que

$$\frac{\boldsymbol{y}'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J} - \boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\right)\boldsymbol{y}}{\sigma^2} = \frac{\mathrm{SSE}}{\sigma^2} \sim \chi^2(n - k - 1)$$

En clase vimos el siguiente resultado

Teorema Si $\boldsymbol{y} \sim N_p(\boldsymbol{\mu}, \sigma^2\boldsymbol{I})$ entonces $\boldsymbol{y}'\boldsymbol{B}\boldsymbol{y}$ y $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y}$ son independientes si y sólo si $\boldsymbol{B}\boldsymbol{A} = \boldsymbol{0}$.

Vimos que $SSR = \boldsymbol{y}'\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\boldsymbol{y}$ y $SSE = \boldsymbol{y}'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J} - \boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\right)\boldsymbol{y}$. Para ver si SSR es idenpendiente de SSE necesitamos ver si $\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J} - \boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\right) = \boldsymbol{0}$. Previamente vimos que $\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\boldsymbol{J} = \boldsymbol{0}$ y que $\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'$ es idempotente por lo que

$$\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J} - \boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\right) = \boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c' - \left(\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right)^{-1}\boldsymbol{X}_c'\boldsymbol{X}_c\left(\boldsymbol{X}_c'\boldsymbol{X}_c\right.\right.$$

Concluimos que SSR es independiente de SSE.

Distribución F-cuadrada no central

Si $U \sim \chi^2(n)$ y $V \sim \chi^2(m)$ con $U \perp V$ entonces

$$W = \frac{U/n}{V/m} \sim F(n, m).$$

Se tiene que

$$\mathbb{E}[W] = \frac{m}{m-2}, \qquad \mathrm{Var}\,(W) = \frac{2m^2(n+m-2)}{n(m-1)^2(m-4)}$$

Por otro lado, si $U \sim \chi^2(n, \lambda)$ y $V \sim \chi^2(m)$ con $U \perp V$ entonces

$$Z = \frac{U/n}{V/m} \sim F(n, m, \lambda).$$

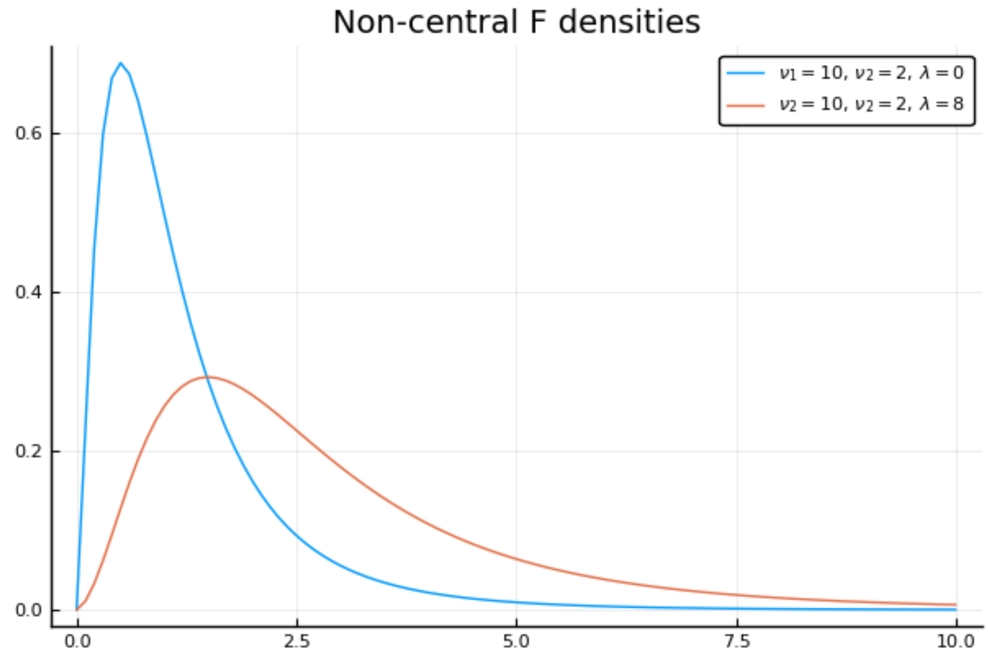
Se tiene que

$$\mathbb{E}[Z] = \frac{m}{m-2}\left(1 + \frac{2\lambda}{m}\right),$$

cuyo valor es mayor que $\mathbb{E}[W]$

In [22]:
```
xrange = collect(0:0.1:10)
plot( xrange, pdf.( NoncentralF(5, 10, 0.0), xrange), label=L"\nu_1=10,\, \nu_2=2
plot!(xrange, pdf.( NoncentralF(5, 10, 7.0), xrange), label=L"\nu_2=10,\, \nu_2=2
title!("Non-central F densities")
```

Out[22]:



Ejemplos relacionados al modelo de regresión lineal múltiple con errores normales

Sabemos que

$$\frac{\mathrm{SSR}}{\sigma^2} \sim \chi^2(k, \lambda_1)$$

y

$$\frac{\mathrm{SSE}}{\sigma^2} \sim \chi^2(n-k-1)$$

Por lo que

$$\boldsymbol{F} = \frac{SSR/\left(\sigma^2 k\right)}{SSE/\left(\sigma^2(n-k-1)\right)} = \frac{SSR/k}{SSE/(n-k-1)} \sim F(k, n-k-1, \lambda_1)$$

Distribución t no central

Recordamos que si $X \sim N(0,1)$ y $Z \sim \chi^2(p)$ con $X \perp Z$ entonces

$$t = \frac{X}{\sqrt{\frac{Z}{p}}} \sim t(p)$$

Si $Y \sim N(\mu, 1)$ y $Z \sim \chi^2(p)$ con $Y \perp Z$ entonces definimos la distribución t no central como

$$t = \frac{Y}{\sqrt{\frac{Z}{p}}} \sim t(p, \mu)$$

con p grados de libertad y parámetro de no centralidad μ.

Observamos que si $Y \sim N(\mu, \sigma^2)$ con $\perp Z$ entonces

$$\frac{\frac{Y}{\sigma}}{\sqrt{\frac{Z}{p}}} \sim t\left(p, \frac{\mu}{\sigma}\right)$$

```
In [ ]:
```