This ***sinh-arcsinh normal*** distribution was first introduced in 2009 as a generalization of the normal distribution. While the normal distribution is symmetric and has light to moderate tails and can be defined by just two parameters (\mu for location and \sigmafor scale), the sinh-arcsinh distribution has two more parameters which control asymmetry and tail weight.

Given the 4 parameters, the sinh-arcsinh normal distribution is defined as

\begin{aligned} X = \mu + \sigma \cdot \text{sinh}\left[ \frac{\text{sinh}^{-1} (Z) + \nu}{\tau} \right], \end{aligned}

where \text{sinh}(x) = \dfrac{e^x - e^{-x}}{2}and \text{sinh}^{-1}(x) = \log \left( x + \sqrt{1 + x^2} \right)are the hyperbolic sine function and its inverse.

* \mucontrols the location of the distribution (where it is “centered” at),
* \sigmacontrols the scale (the larger it is, the more spread out the distribution is),
* \nucontrols the asymmetry of the distribution (can be any real value, more positive means more right skew, more negative means more left skew), and
* \taucontrols tail weight (any positive real value, \tau > 1means lighter than normal distribution, \tau < 1means heavier).

From the expression, we can also see that when \nu = 0and \tau = 1, the distribution reduces to the normal distribution with mean \muand standard deviation \sigma.

In R, the gamlss.dist package provides functions for plotting this distribution. The package provides functions for 3 different parametrizations of this distribution; the parametrization above corresponds to the SHASHo set of functions. As is usually the case in R, dSHASHo, pSHASHo, qSHASHo and rSHASHo are for the density, distribution function, quantile function and random generation for the distribution.

First, we demonstrate the effect of skewness (i.e. varying \nu).

library(gamlss.dist)

library(dplyr)

library(ggplot2)

x <- seq(-6, 6, length.out = 301)

nu\_list <- -3:3

df <- data.frame()

for (nu in nu\_list) {

temp\_df <- data.frame(x = x,

y = dSHASHo2(x, mu = 0, sigma = 1, nu = nu, tau = 1))

temp\_df$nu <- nu

df <- rbind(df, temp\_df)

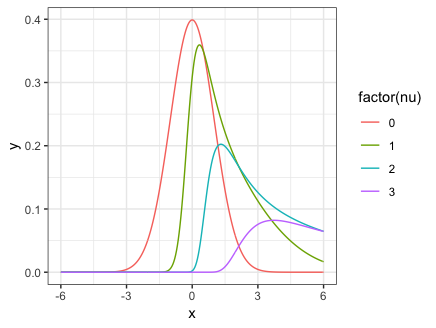
}

As \nubecomes more positive, the distribution becomes more right-skewed:

df %>% filter(nu >= 0) %>%

ggplot(aes(x = x, y = y, col = factor(nu))) +

geom\_line() + theme\_bw()

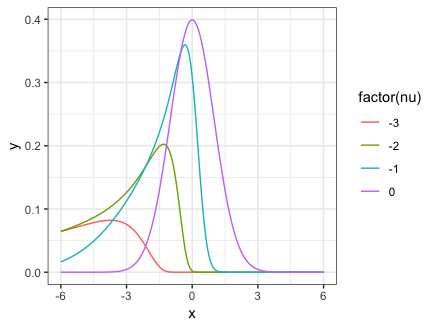


As \nubecomes more negative, the distribution becomes more left-skewed:

df %>% filter(nu <= 0) %>%

ggplot(aes(x = x, y = y, col = factor(nu))) +

geom\_line() + theme\_bw()



Next, we demonstrate the effect varying \tauhas on the weight of the tails. The code and picture below is for when there is no skewness in the distribution:

tau\_list <- c(0.25, 0.75, 1, 1.5)

df <- data.frame()

for (tau in tau\_list) {

temp\_df <- data.frame(x = x,

y = dSHASHo(x, mu = 0, sigma = 1, nu = 0, tau = tau))

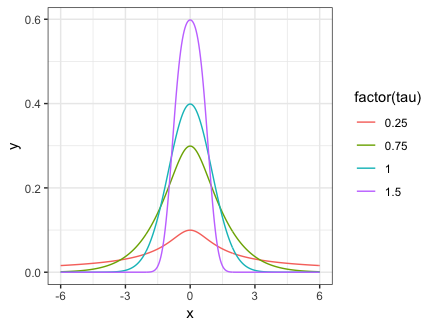
temp\_df$tau <- tau

df <- rbind(df, temp\_df)

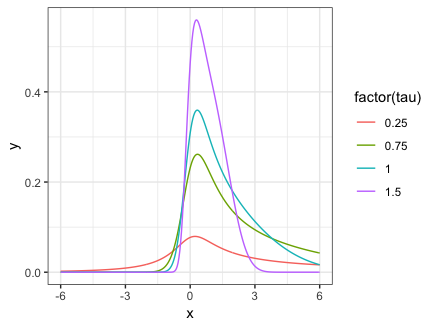
}

ggplot(data = df, aes(x = x, y = y, col = factor(tau))) +

geom\_line() + theme\_bw()



By changing nu = 0 to nu = 1 in the code above, we see the effect of tail weight when there is skewness:



(**Note:** For reasons unclear to me, the *Significance* article uses different symbols for the 4 parameters: \xiinstead of \mu, \etainstead of \sigma, \epsiloninstead of \nuand \deltainstead of \tau.)