- Recall: the likelihood test statistic $\Lambda(\theta)$ follows a $\chi^2(1)$ distribution
  - Wait what's a Chi-squared distribution?
- The **Chi-squared distribution with** $k$ **degrees of freedom** is denoted $\chi^2_k$ or $\chi^2(k)$.
- The probability density function is a bit messy:

$$\frac{1}{2^{k/2}\Gamma(k/2)} y^{k/2-1} e^{-y/2}$$

↳ Gamma function

## Visualizing the Chi-squared Distribution

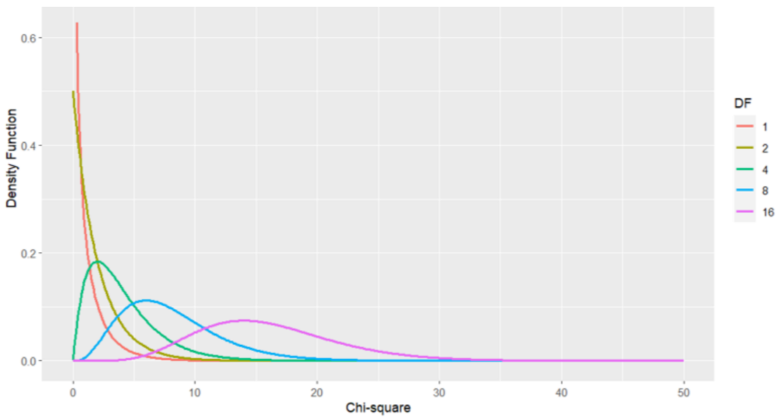



Figure 1: Chi-squared distribution probability density functions for different degrees of freedom.

heavy right skew

Useful R commands:

- pchisq(w, k) — P(W ≤ w; k) where k = degrees of freedom
- qhisq(p, k) — value of w such that P(W ≤ w; k) = p

- If $W \sim \chi^2(k)$, then:

$$E(W) = k, \; Var(W) = 2k$$

- **Proof**: pages 160-162 of the course notes
- If $W_1, ..., W_n$ are independent random variables with $W_i \sim \chi^2(k_i)$, then

$$\sum_{i=1}^n W_i \sim \chi^2_{\sum k_i}$$

- **Proof**: problem 21 in Chapter 4

Relating to the normal distribution:

- **Theorem**: if $Z \sim G(0,1)$ then $Z^2 \sim \chi^2(1)$
- **Corollary**: if $Z_1, ... Z_n$ are mutually independent $G(0,1)$ random variables, then

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

↳ n degrees of freedom

Ex. $W \sim \chi^2(1)$

$$\Rightarrow \text{Then } P(W \geq w) = 1 - P(W \leq w) = 1 - P(Z^2 \leq w) = 1 - P(|Z| \leq \sqrt{w})$$

## Student's t Distribution

- The **Student's t distribution (t distribution)** has the following probability density function:

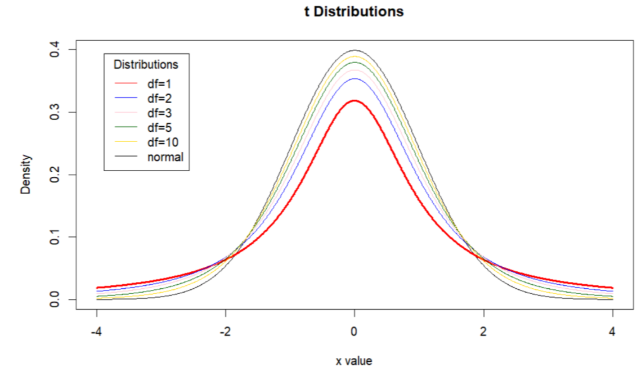
$$f(t;k) = c_k\left(1 + \frac{t^2}{k}\right)^{-(k+1)/2}$$

- The constant $c_k$ is given by

$$c_k = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi}\,\Gamma(\frac{k}{2})}$$

k - degrees of freedom

### Visualizing the t Distribution




As k -> ∞:

- Distribution looks closer to Gaussian
- Kurtosis gets smaller

Notation: T ~ t(k)

- Let $Z \sim G(0,1)$ and $U \sim \chi^2(k)$, independently
- Then

$$T = \frac{Z}{\sqrt{U/k}}$$

has a Student's t distribution with $k$ degrees of freedom

- Recall: if $\sigma$ is known, and $Y_1, \ldots, Y_n$ are independently and identically distributed from a $G(\mu,\sigma)$ distribution, then

$$Q(Y;\theta) = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim G(0,1)$$

is a pivotal quantity
- What if $\sigma$ is also unknown?

- First, recall that the maximum likelihood *estimator* for $\sigma^2$ is
$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$
- The sample estimator for $\sigma^2$ is
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$
- Using the sample estimator $S^2$, it can be shown that
$$V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

- It can then be shown that if $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim G(0,1)$ and $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$,
$$\frac{\bar{Y}-\mu}{S/\sqrt{n}} = \frac{Z}{\sqrt{V/(n-1)}} \sim t(n-1)$$
- Note that this is also a pivotal quantity (why?)

← does not depend on σ

- We can apply the same process using $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$ and quantiles $a$ from the t-distribution such that $P(T \le a) = (1+p)/2$:
$$P(-a \le \frac{\bar{Y}-\mu}{S/\sqrt{n}} \le a) = p$$
$$P(-aS/\sqrt{n} \le \bar{Y} - \mu \le aS/\sqrt{n}) = p$$
$$P(\bar{Y} - aS/\sqrt{n} \le \mu \le \bar{Y} + aS/\sqrt{n}) = p$$
- Substituting our observed data,
$$[\bar{y} - \underbrace{as}_{}/\sqrt{n}, \bar{y} + as/\sqrt{n}]$$

calculated value of σ using estimator

## Two Different Confidence Intervals for $\mu$

- We now have two different confidence interval forms for $\mu$
- If $\sigma$ is known:
$$[\bar{y} - a\sigma/\sqrt{n}, \bar{y} + a\sigma/\sqrt{n}]$$
- Here, $P(Z \le a) = (1+p)/2$, $Z \sim G(0,1)$
- If $\sigma$ is not known:
$$[\bar{y} - as/\sqrt{n}, \bar{y} + as/\sqrt{n}]$$
- Here, $P(T \le a) = (1+p)/2$, $T \sim t(n-1)$

## Sample Size Calculations: Gaussian Data

- With Gaussian data, we now have two $100p\%$ confidence intervals:
$$[\bar{y} - a\sigma/\sqrt{n}, \bar{y} + a\sigma/\sqrt{n}]$$
and
$$[\bar{y} - as/\sqrt{n}, \bar{y} + as/\sqrt{n}]$$
- The widths are therefore
$$2a\sigma/\sqrt{n}$$
or
$$2as/\sqrt{n}$$
- **Question**: do you notice a problem with the last line?

s is the variance, which is calculated in terms of n. This means that we cannot isolate for n.

There is also no "worst case" value for s (unlike in the binomial, where the worst case MLE is 0.5), so we cannot cheat

## Sample Size Calculation with Gaussian Data

- If we want the width of our confidence interval to be of length $2\ell$, then we want
$$\frac{2a\sigma}{\sqrt{n}} = 2\ell$$
and, subsequently,
$$n \approx (\frac{a\sigma}{\ell})^2$$
- Since we usually have some uncertainty about $\sigma$ when assuming a particular value for it, we'd usually choose $n$ larger than $(\frac{a\sigma}{\ell})^2$

## Constructing Confidence Intervals for $\sigma^2$

- Suppose we now want to construct a $100p\%$ confidence interval for $\sigma^2$
- So far, the process consists of three general steps:
  1. Obtain a pivotal quantity $Q(Y;\theta)$ with a $G(0,1)$ or $t(n-1)$ distribution
  2. Obtain quantiles such that $P(Z \leq a) = (1+p)/2$ or $P(T \leq a) = (1+p)/2$
  3. Using $P(-a \leq Q(Y;\theta) \leq a) = p$, isolate for $\theta$ in the middle of the inequality
  4. Obtain
$$[L(y), U(y)]$$

Ex. Example: Suppose we collected data on STAT 231 exam scores:

$\bar{y} = 72.5$, $s = 9.4$, $n = 200$

Let $Y_i \sim G(\mu, \sigma)$, $i = 1, 2, \dots, 200$ represent an exam score.

To find a 98% C.I. for $\mu$:

$$\boxed{\bar{y} \pm a s/\sqrt{n}}$$

Since sigma is not known, we will replace it with an estimator for S:

$$T = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t(\underbrace{n-1}_{\text{degrees of freedom}})$$

Note that $Z = \dfrac{\bar{Y} - \mu}{\sigma/\sqrt{n}}$ and $T = \dfrac{Z}{\sqrt{U}/\sqrt{k}}$

Substituting $k = n - 1$: