3. Distribution of the studentized sample mean

a. Again suppose $X_j \sim \mathcal{N}(\mu, \sigma^2)$, independent.

b. Then $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim \mathcal{N}(0, 1)$, because:
   
   i. Note that $\bar{X}_n - \mu \sim \mathcal{N}(0, \sigma^2/n)$.

   c. When $\sigma$ is unknown, it is often approximated by $\hat{\sigma} = \sqrt{Q_n/(n-1)}$.

   d. Unfortunately, $T = \sqrt{n}(\bar{X}_n - \mu)/\hat{\sigma}$ is not $\sim \mathcal{N}(0, 1)$, because:
      
      i. because of the variability in $\hat{\sigma}$.
      
      ii. Difference is noticeable for moderate $n$ (smaller than 40).
      
      iii. A quantity divided by its standard deviation is called studentized.

   e. Distribution $T$ of a standard normal divided by the square root of an independent $\chi^2$ divided by degrees of freedom is called $t$ distribution on $n - 1$ degrees of freedom.

   f. $T$ probability density function $\propto (1 + t^2/k)^{-k/2-1/2}$, for $k$ degrees of freedom of denominator, because:
      
      i. Standardize:
      
      • Divide numerator and denominator by $\sigma$
      
      • Subtract $\mu$ from $X$. 
ii. \( T = XW^{-1/2}\sqrt{k} \), for \( X \perp W \), \( X \sim \mathcal{N}(0, 1) \), \( W \sim \chi^2_k \).

iii. \( Z = W \)

   - \( X = \sqrt{ZT}/\sqrt{k} \), \( W = Z \)
   - \( dX/dT = \sqrt{Z}/\sqrt{k} \), \( dW/dZ = 1 \), \( dW/dT = 0 \), \( dX/dZ \) irrelevant,
   - Jacobian is \( \sqrt{Z}/\sqrt{k} \).

iv. Joint probability density function of \( X \) and \( W \) is

\[
\frac{\exp(-x^2/2)}{\sqrt{2\pi}} \times w^{k/2-1} \exp(-w/2) 2^{-k/2} \frac{\Gamma(k/2)}{\sqrt{2\pi}}.
\]

v. Joint probability density function of \( T \) and \( Z \) is

\[
\frac{\exp(-z t^2/(2k))}{\sqrt{2\pi}} z^{k/2-1} \exp(-z/2) 2^{k/2} \Gamma(k/2) \sqrt{k}.
\]

vi. \( Z = K \exp(-z(t^2/k + 1)/2) z^{k/2-1/2} \) for \( K = 2^{-k/2}/(\sqrt{2\pi}\Gamma(k/2)) \).

vii. Marginal probability density function of \( T \) is

\[
K'(t^2/k + 1)^{-k/2-1/2},
\]

because:

\[
f_T(t) = \int_0^\infty K \exp(-z(t^2/k + 1)/2) z^{k/2-1/2} \, dz = \int_0^\infty K \exp(-v) v^{k/2-1/2}(t^2/k + 1)^{-k/2-1/2} \, dv.
\]

viii. Distribution depends on degrees of freedom.

ix. Moment of order \( r \) for a \( T_k \) distribution exists only if \( k > r \).
x. \( k = \infty \) is equivalent to standard normal.

xi. Distribution with df 1 coincides with Cauchy. See Fig. 44.

\[ e^{-x^2 / 2} \]

\[ T \sim T_k \]

\[ \Pr(T \leq t) \]

Fig. 44: T Densities

\begin{align*}
\text{DF} & \quad 1 \\
\text{DF 5} & \quad 30 \\
\text{DF \infty} & \quad \infty
\end{align*}

\[ x \]

\[ 0 \]

\[ 0.1 \]

\[ 0.2 \]

\[ 0.3 \]

\[ 0.4 \]

\[ -4 \]

\[ -2 \]

\[ 0 \]

\[ 2 \]

\[ 4 \]

\[ g. \] R calculates distributional quantities for \( T \sim T_k \).

\[ i. \] Calculate probabilities \( \Pr(T \leq t) \) using \( pt(t, k) \).
ii. Calculate quantiles using \( qt(q, k) \) for \( q \in (0, 1) \).

4. Distribution of the ratio of sums of squares.
   
a. Suppose that
   
   i. \( Q_a \) and \( Q_b \) are \( \chi^2_a \) and \( \chi^2_b \) variables respectively.
   
   ii. \( Q_a \perp Q_b \)

b. Then the distribution of variable \( F = (Q_a/a)/(Q_b/b) \) is called the \( F \)-distribution with \( a \) and \( b \) degrees of freedom.

c. Distribution depends on numerator, denominator df.
   
   i. When \( a = 1 \), then \( F \) has the distribution of a \( t_b \) variable, squared.
   
   ii. When \( b \to \infty \), then \( F \) has the \( \chi^2_a \) distribution. See Fig. 45.

d. R can calculate distributional quantities for \( F \sim F_{a,b} \).
   
   i. Calculate probabilities \( P(F \leq f) \) using \( pf(f, a, b) \).
   
   ii. Calculate quantiles using \( qf(q, a, b) \) for \( q \in (0, 1) \).
Fig. 45: F Densities

- Densities for different degrees of freedom (DF):
  - DF 2,2
  - DF 2,20
  - DF 20,2
  - DF 20,20