MATH 321: Mathematical Statistics

Lecture 1: Random Samples and Common Statistics

Chapter 5: Distributions of Functions of Random Variables (5.5)

MATH 320 vs MATH 321

Relationship between Probability and Statistics

- We studied **Probability** in MATH 320 and are going to study **Statistical Inference** in MATH 321.

  So what is the difference between them?

- In a probability problem, the properties of the population are assumed known, and we use these to infer properties of the sample.

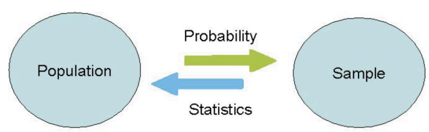


Figure: The reverse actions of Probability and Statistics

- Whereas statistics is concerned with learning (inferring) population properties from sample information (which is the opposite of probability).
- In spite of this difference, statistical inference itself would not be possible without probability because it is based on probability calculations.
- Example:
  - Suppose we know 75% of batteries last longer than 1500 hours. We want to know the chance that in a sample of 30 batteries at least 20 will last more than 1500 hours.
    - What is known (in other words, fixed)? parameters ($p = 0.75$, $n = 30$)

      This is population information

      $\Longrightarrow$ This a probability question.
    - What is unknown (in other words, variable)? sample info e.g) observed $x = ??$
    - We answer probability question using the distribution of X.
  - The most important thing when solving probability questions is the **distribution**. If we know this, we can answer any questions in probability.

    e.g) $P(X \in A)$

    $E(X)$

    $V(X|Y=y)$

- Suppose that in a sample of 30 batteries, only 20 are found to last more than 1500 hours. We want to know if that is enough evidence to conclude that the proportion of all batteries that last more than 1500 hours is less than 75%.
  - What is known (in other words, fixed)? observed $x = 20$, sample proportion $\hat{p} = \frac{20}{30} \approx 0.66$

    This is Sample information

    $\Longrightarrow$ This a statistics question.
  - What is unknown (in other words, variable)? population $p = ??$
  - We answer Statistics question using the distribution of the statistic.
- Whether in a probability or statistics context, we are always looking for the **distribution**.

- What we are going to study in MATH 321:
  - Statistics (and their properties)
  - The distributions of statistics
  - Principles we would like statistics to have
  - Statistical inference (confidence intervals and hypothesis tests)
  - Other applications, such as regression

Example process

1. Collect data, $x = \{1510, 1700, 1400, \ldots\}$ 30 observations
2. Transform data to 0s and 1s (if $x > 1500 \to 1, \text{else } 0$)
3. Summarize with a statistic (which is a random variable)
4. Find distribution
5. Compute stuff (expected value, variance, probabilities, etc.)

Basic concepts of random samples

Random sample

- Motivation: Often, the data collected in an experiment consist of several observations on a variable of interest.

  In this section, we present a model for data collection that is often used to describe this situation, a model referred to as **random sampling**.

- Definition: The random variables $X_1, \ldots, X_n$ are called a **random sample** of size n from the population $f(x)$ if $X_1, \ldots, X_n$ are mutually independent random variables and the marginal pdf or pmf of each X_i is the same function $f(x)$.

  In other words, $X_1, \ldots, X_n$ are **independent and identically distributed (*iid*)** random variables with pdf or pmf $f(x)$.

  random sample $\Longleftrightarrow$ iid

  2 conditions ① mutually $\perp\!\!\!\perp$ ② same distribution

- Lets build up to the model (joint pdf or pmf) that we will be working with from now on starting with what we learned in the multivariate setting:

  0. Joint distribution

     joint distribution $f(x_1, \ldots, x_n)$

  random sample

  1. Mutually independent random variables:

     Split into (different) marginals

     $f_{X_1}(x_1) \cdots f_{X_n}(x_n)$

  2. Identically distributed:

     all same $f(x)$, drop X_i (only observed x_i are different)

     $f(x_1) \cdots f(x_n)$

  3. Parametric family:

     e.g.) Geometric(p), $\exp(\lambda)$

     Same parameter in each term, add θ

     $f(x_1|\theta) \cdots f(x_n|\theta)$

     $\star \Longrightarrow f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$

     final model

- In a statistical setting, if we assume that the population we are observing is a member of a specific parametric family, but the true parameter value is unknown, then a random sample from this population has a joint pdf or pmf of the above form with the value of θ unknown.

  By considering different possible values of θ, we can study how a random sample would behave for different populations.

- Example: $X_1, \ldots, X_n$ correspond to the times until failure (measured in years) for n identical circuit boards that are put on test and used until they fail. Find the joint pdf of the random sample $X_1, \ldots, X_n$.

  assume $X_i \overset{iid}{\sim} Exp(\lambda) \rightarrow f(x_1, \ldots, x_n|\lambda) = \prod_{i=1}^{n} \underbrace{f(x_i|\lambda)}$ $\hookrightarrow = \lambda e^{-\lambda x_i}$, $x_i > 0$, $\lambda > 0$

  $= \lambda^n e^{-\lambda \sum x_i}$

Scenario

- Suppose we collect information on 30 circuit boards, say $x_1, \ldots, x_{30}$.

  When reporting this information, do we ever report the entire set of observations? Of course Not!

- Instead, we report summary statistics.

  In doing so, we are summarizing the information in a sample by determining a few key features of the sample values. This is usually done by computing statistics (functions of the sample).

- These statistics define a form of data reduction. There are advantages and consequences of this.

  Whether or not statistics are "good enough" is a topic that will be left for grad school, but we will work with the most common ones that have good properties.

Sums of random variables from random samples

Statistics (estimators)

- Definition: Let $X_1, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, \ldots, x_n)$ be a real-valued or vector-valued functions whose domain includes the sample space of $(X_1, \ldots, X_n)$.

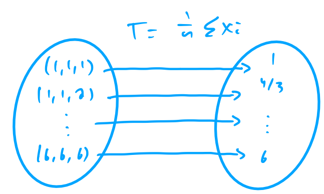
  Then the random variable or random vector $Y = T(X_1, \ldots, X_n)$ is called a **statistic**. The probability distribution of a statistic Y is called the **sampling distribution of** $\boldsymbol{Y}$.

- Breakdown of definition:

  – Statistic $Y = T(X_1, \ldots, X_n)$ is a function. All functions have a domain (set of inputs) and codomain (set of outputs).

    The domain of Y is the sample space of $X_1, \ldots, X_n$ and the codomain of Y depends on the statistic.

  – Example: Roll 3 die (x_1, x_2, x_3)



    Some examples of statistics: Mean, variance, median, min, max, etc.

    Usually we are interested in one of these at a time, but not always (e.g. $(\bar{X}, S^2)$ is a two-dimensional statistic (a vector)).

  – The statistic $Y = T(X_1, \ldots, X_n)$ is a random variable because it is a function of random variables.

- Miscellaneous notes:
  - Statistics and estimators are exactly the same thing.
  - It's called a sampling distribution because it is derived from a random sample.
  - The definition of a statistic is very broad. The only restriction is that a statistic can not be a function of the parameter (because it is unknown).
  - It is not necessary that $X_1, \ldots, X_n$ be *iid* to define a statistic. Even if they are dependent and/or not identically distributed, $Y = T(X_1, \ldots, X_n)$ is still called a statistic, but is is much more difficult to deal with (of course beyond the scope of this class).

- Ultimately, we want to find the distribution, which is usually tractable (able to be found) because of its simple probability structure (*iid*). But we will start with summary measures of two common statistics.

Sample mean and variance

- Definition: The **sample mean** is the arithmetic average of the values in a random sample. It is usually denoted by

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Definition: The **sample variance** is the statistic defined by

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

The **sample standard deviation** is the statistic defined by $S = \sqrt{S^2}$.

- Theorem: Let $X_1, \ldots, X_n$ be a random sample of size n from a population (of any distribution) with mean μ and variance $\sigma^2 < \infty$. Then

(a) $\mu_{\bar{X}} = E(\bar{X}) = \mu$ (mean of sample means)

(b) $\sigma^2_{\bar{X}} = V(\bar{X}) = \dfrac{\sigma^2}{n}$ (variance of sample means)

(c) $E(S^2) = \sigma^2$ (mean of sample variance)

Some proofs: (a) & (b)

$\rightarrow$ linear combinations with $a_i = 1/n$

$\bar{X} = \frac{1}{n} \sum X_i \quad \rightarrow E(\frac{1}{n} \sum X_i) = \frac{1}{n} \sum \underbrace{E(X_i)}_{= \mu} = \frac{1}{n}(n\mu) \overset{\checkmark}{=} \mu$

$\rightarrow V(\frac{1}{n} \sum X_i) \underset{\text{b/c } \perp}{=} \frac{1}{n^2} \sum \underbrace{V(X_i)}_{= \sigma^2} = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n} \Rightarrow$ goes to zero as $n \to \infty$
(smaller variance with more data)

- Notes:
  - It can be shown why we use $n-1$ in the definition of S^2 rather than just n (i.e. it is necessary to be unbiased, which means the statistic's expected value is equal to the parameter it is estimating).

    $$E\left[\frac{1}{n}\sum(x_i-\bar{x})^2\right] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

  - The statistic $\bar{X}$ is an unbiased estimator of μ, and S^2 is an unbiased estimator of σ^2. We discuss this more later.

Sampling distribution of $\bar{X}$

- Now we would like to study the sampling distribution of $\bar{X}$.

  Recall whenever finding the distribution of a sum of random variables, we want to use the mgf technique.

- Theorem: Let $X_1, \ldots, X_n$ be a random sample from a population with mgf $M_X(t)$. Then the mgf of the sample mean is

  $$M_{\bar{X}}(t) = [M_X(t/n)]^n$$

  Note this theorem is a combo of previous theorems where we found the mgf of *iid* $X_1 + \cdots + X_n$ and also now with all coefficients $a_i = 1/n$.

- Examples:
  1. Continuing circuit board failure time example: Find the distribution of $\bar{X}$ the mean time until failure (measured in years) for n identical circuit boards that are put on test and used until they fail. (recall $M_X(t) = \lambda/(\lambda - t)$)

     $$M_{\bar{X}}(t) = [M_X(t/n)]^n = \left[\frac{\lambda}{\left(\frac{n}{n}\right)\lambda - \frac{t}{n}}\right]^n = \left[\frac{\lambda}{\left(\frac{n\lambda - t}{n}\right)}\right]^n = \left[\frac{n\lambda}{n\lambda - t}\right]^n \sim \text{Gamma}(\alpha = n, \beta = n\lambda)$$

  2. Suppose $X_i \overset{iid}{\sim} \text{Binomial}\,(p = 0.6, n = 10)$, for $i = 1, \ldots, 5$.

     Find $E(\bar{X})$, $V(\bar{X})$, and the distribution of $\bar{X}$. (recall $M_X(t) = (q + pe^t)^n$)

Binomial

$\mu = np = 10(0.6) = 6$

$\sigma^2 = npq = 10(0.6)(0.4) = 2.4$

$\Longrightarrow$

$\mu_{\bar{x}} = \mu = 6$ ✓

$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{2.4}{5} = 0.48$ ✓

$\hookrightarrow$ sample size

(much less variable than X distribution)

$$M_{\bar{X}}(t) = [M_X(t/5)]^5 = \left[\left[0.4 + 0.6e^{t/5}\right]^{10}\right]^5 = \left(0.4 + 0.6e^{t/5}\right)^{50} \sim ??$$

doesn't follow a familiar mgf

Sampling from the normal distribution

Introduction

- This section deals with the properties of sample quantities drawn from a normal population – still one of the most widely used statistical models.
- In practice, often
  1. We assume population has a normal distribution (e.g. heights, test scores, errors in measurements, etc.).
  2. Then sample from population and form statistics in order to estimate the mean and variance of the population.
- Thus we wish to know the distribution of these statistics which will allow us to determine accuracy, form confidence intervals, hypothesis testing, etc.

  Sampling from a normal population leads to many useful properties of sample statistics and also to many well-known sampling distribution.

Theorem

- Let $X_1, \ldots, X_n$ be a random sample of size n from a Normal (μ, σ^2) distribution. Then
  (a) $\bar{X}$ and S^2 are independent random variables.
  (b) $\bar{X}$ has a Normal $(\mu, \dfrac{\sigma^2}{n})$ distribution.
  (c) $\dfrac{(n-1)}{\sigma^2} S^2$ has a chi squared distribution with $n-1$ degrees of freedom (df).

- Notes / proofs:
  (a) $\bar{X} \perp\!\!\!\perp S^2 \rightarrow$ This is just a necessary fact when deriving (c), we will not prove this.

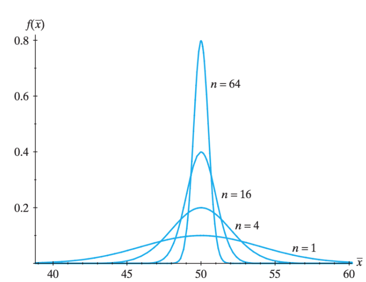
  (b) $\bar{X} \sim \text{Normal}\,(\mu, \dfrac{\sigma^2}{n})$

  We know this from prior theorems (parameters: mean and variance of sample means), dsitribution (sum of normals = normal).

  Example: Let $X_1, \ldots, X_n$ be a random sample from $N(\mu = 50, \sigma^2 = 16)$.

  See the effect of n in the distributions.

  $P(49 < \bar{X}_{64} < 51) \approx 0.9545$
  $P(49 < X < 51) \approx 0.1974$.



pdfs of means of samples from $N(50, 16)$

(c) $\dfrac{(n-1)}{\sigma^2}S^2 \sim \chi^2(n-1) \rightarrow$ This is a new distribution, so lets learn about this before motivating the idea behind this part of the theorem.

Chi-square distribution

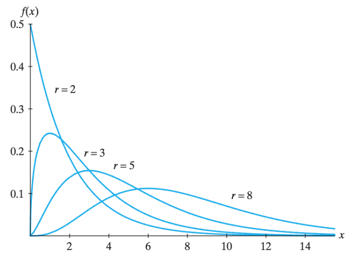
Definition

- The chi-square distribution is a special case of the gamma distribution that plays an important role in statistics.
- If $X \sim \chi^2$ with r degrees of freedom (often written $\chi^2{}_r$ or $\chi^2(r)$), then
  (i) Pdf:
  $$f(x) = \frac{1}{\Gamma(\frac{r}{2})2^{r/2}}\, x^{\frac{r}{2}-1}\, \mathrm{e}^{-\frac{x}{2}}, \qquad x \geq 0$$
  (ii) Mean and variance: (scale gamma)
  $$E(X) = \alpha\theta = \frac{r}{2}(2) = r \qquad V(X) = \alpha\theta^2 = \frac{r}{2}(2^2) = 2r$$
  (iii) Mgf:
  $$M_X(t) = \left(\frac{1}{1-\theta t}\right)^{\alpha} = \left(\frac{1}{1-2t}\right)^{r/2}, \quad t < 1/2$$
- Notes about the chi-square distribution.
  - Special case of gamma: Chi-square is more commonly represented as the **scale parameterization** of the gamma (not the version we used).

    Scale $\theta = \dfrac{1}{\text{rate}} = \dfrac{1}{\beta}$

    This changes the pdf, expected value, variance and mgf slightly.
  - Pdf: Just the (scale) gamma density function with $\alpha = r/2, \;\; \theta = 2.$
  - Mean and variance: Just the mean and variance of (scale) gamma with the specific parameter values above.
  - Characteristics: Right-skewed density function. Unbounded support. range $0 \leq x < \infty$
  - Probabilities: Just like the gamma distribution, probabilities need to be found using software. There is also a table of probabilities (just like a Z-table), because it is used in statistical tests.
  - Parameter: The degrees of freedom r must be a positive integer when used in the gamma distribution. $r = \{1, 2, 3, \ldots\} \longrightarrow \chi^2_r \Rightarrow \alpha : \{0.5, 1, 1.5, \ldots\}$

    Here is how it affects the shape (recall α is the shape parameter) of the density curve: Larger values of r shifts the probability to the right.



Important facts about the chi-square random variables

1. If $Z \sim \text{Normal}\,(0,1)$, then $Z^2 \sim \chi^2\,(1)$.
   - So the square of a standard normal random variable follows a chi squared distribution with 1 degree of freedom.
   - We will not prove this, but this means we can start with a normal random variable X with mean μ and standard deviation σ and end up with a chi-square random variable:

     standardize and square $\left(\frac{\overset{\text{Normal}}{X-\mu}}{\sigma}\right)^2 = Z^2 \sim \chi^2_1$

2. If $X_1, \ldots, X_n$ are mutually independent and $X_i \sim \chi^2\,(r_i)$ for $i = 1, \ldots, n$, then $Y = X_1 + \cdots + X_n \sim \chi^2\,(r_1 + \cdots + r_n)$.
   - The degrees of freedom are additive.
   - Proof:

     $$M_Y(t) = M_{X_1+\cdots+X_n}(t) = M_{X_1}(t) \cdot \ldots \cdot M_{X_n}(t)$$
     $$= \left(\frac{1}{1-2t}\right)^{r_1/2} \cdot \ldots \cdot \left(\frac{1}{1-2t}\right)^{r_n/2}$$
     $$= \left(\frac{1}{1-2t}\right)^{\frac{r_1+\cdots+r_n}{2}}$$
     $$\sim \chi^2_{r_1+\cdots+r_n}$$

   - Result / extension of this: If $X_1, \ldots, X_n$ are mutually independent random variables with $X_i \sim \text{Normal}\,(\mu_i, \sigma_i)$ for $i = 1, \ldots, n$, then

     $$\sum_{i=1}^{n} \left(\frac{X_i-\mu_i}{\sigma_i}\right)^2 = \sum_{i=1}^{n} \underbrace{Z_i^2}_{\sim \chi^2_1}$$
     $$\sim \chi^2_n$$

     This is one way to think about what the parameter r represents:

     $r =$ the number of independent standard normals that we are adding together.

Return to theorem for $\bar{X}$ and S^2

(c) $\dfrac{(n-1)}{\sigma^2}S^2 = \displaystyle\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2$ has a chi squared distribution with $n-1$ degrees of freedom.

- We will not prove this, but we can understand the pieces of the theorem (the coefficients in front of S^2 and the chi-square result). Recall $X_i \overset{iid}{\sim} \mathrm{N}(\mu, \sigma^2)$:

Don't know this distribution (pdf)

start

want to get ≈ here

know this distribution

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \longrightarrow \sum_{i=1}^{n}\Big(\underbrace{\frac{X_i-\mu}{\sigma}}_{=Z}\Big)^2 \sim \chi^2(n)$$

$$(n-1) * S^2 = \cancel{\frac{1}{n-1}} \sum (X_i - \bar{X})^2 * \cancel{(n-1)}$$

$$\frac{1}{\sigma^2} * (n-1)\, S^2 = \sum (X_i - \bar{X})^2 * \frac{1}{\sigma^2}$$

$$\frac{n-1}{\sigma^2} S^2 = \sum \left(\frac{X_i - \bar{X}}{\sigma}\right)^2$$

Only difference is μ vs

→ estimate μ with because it is unknown

⇒ 1 d f is lost

⇒ $\chi^2_{(n-1)}$ now

Example

- Let $X_1, \ldots, X_4$ be a random sample from of size 4 from a normal distribution $N(\mu = 76, \sigma = 383)$ and

$$U = \sum_{i=1}^{4}\Big(\underbrace{\frac{X_i - 76}{383}}_{=Z_i}\Big)^2 \sim \chi^2_4 \quad \text{and} \quad W = \sum_{i=1}^{4}\left(\frac{X_i - \bar{X}}{383}\right)^2 \sim \chi^2_{4-1} = \chi^2_3$$

μ

Compute $P(0.7 < U < 7.8)$ and $P(0.7 < W < 7.8)$.

χ^2 cdf (lower = 0.7, upper = 7.8, df = 4) = < now with df =3 >

$\approx 0.8521 \quad > \quad \approx 0.8229$

Derived distributions: Student's t and Snedecor's F

Derivation of the t distribution

- We are studying the properties of $\bar{X}$ in order to make inferences about μ.
- If $X_1, \ldots, X_n$ are a random sample for a $N(\mu, \sigma^2)$, we know that the quantity

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1) = Z$$

- If we knew the value of σ and we measured $\bar{X}$, then we could use this as a basis for inference about μ, since μ would then be the only unknown quantity.

  However, most of the time σ is unknown. So, Student (W. S. Gosset) looked at the distribution of

$$\frac{\bar{X} - \mu}{s/\sqrt{n}}$$

  as a quantity that could be used as a basis for inference about μ when σ was unknown.

- When deriving this, we have the form of the statistic, but don't know its pdf (distribution). So here is the logic to get this statistic into a form that we are then able to find the actual equation for the pdf.

$\to$ Derivation $\to$ $\dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{\sigma}{\sigma}\left(\dfrac{\bar{x} - \mu}{s/\sqrt{n}}\right) = \dfrac{\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}}{\sqrt{s^2/\sigma^2}}$ rearrange

Independent because $\bar{X} \perp\!\!\!\perp s^2$

$\to$ Numerator $\to$ $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim Z$

$\to$ Denominator $\to$ $\sqrt{\dfrac{s^2}{\sigma^2}} = \sqrt{\dfrac{n-1}{\sigma^2} s^2 / (n-1)} = \sqrt{\chi^2_{n-1}/(n-1)}$

$\to$ Combined $\to$ $\dfrac{Z}{\sqrt{\chi^2_{n-1}/(n-1)}}$

- Thus, the distribution of interest can be found by solving the simplified problem of finding the distribution of $\dfrac{Z}{\sqrt{U/r}}$, where $Z \sim \text{Normal}\,(0,1)$, $U \sim \chi^2\,(r)$ and $Z \perp\!\!\!\perp U$.

  There are two ways to go from here, but we will not show this.

Definition of the t distribution

- Let $X_1, \ldots, X_n$ be a random sample form a $N(\mu, \sigma^2)$ distribution. If

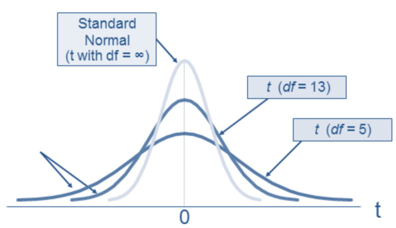
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \qquad \text{then} \qquad T \sim t_{\,n-1}$$

  Equivalently, a random variable T has **Student's *t* distribution** with r degrees of freedom, and we write $T \sim t_{\,r}$ if it has pdf

$$f_T(t) = \frac{\Gamma(\frac{r+1}{2})}{\frac{1}{\sqrt{r\pi}}\Gamma(\frac{r}{2})} \left(\frac{1}{(1 + t^2/r)^{(r+1)/2}} \right), \qquad -\infty < t < \infty$$

Notes about the t distribution

- Density curve, relationship to the standard normal distribution, and probabilities.



  - The t-distribution has the "shape" of a normal distribution (bell-shaped and symmetric about zero), but it has heavier tails. This means there is more probability in the further from the center (and less around the center, zero).
  - As the degrees of freedom r increases, more probability shifts towards the center. Theoretically, as $r \to \infty$, t-distribution tends to the standard normal.
  - Example: Let $T \sim \text{t}_{\,10}$ and $Z \sim \text{N}\,(0,1)$. Compare the following probabilities (t probabilities must be found using software or a t-table):

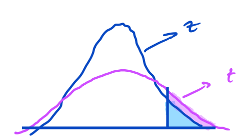
    graphing calc

    (a) Central interval probability: $\quad P(-1 < Z < 1) \quad$ and $\quad P(-1 < T < 1)$ $\quad$ tcdf $\begin{pmatrix} \text{lower} = -1 \\ \text{upper} = 1 \\ df = 10 \end{pmatrix}$

    $\approx 0.682 \quad > \quad \approx 0.659$

    (b) Tail probability: $\quad P(Z > 2) \quad$ and $\quad P(T > 2)$.

    $\approx 0.0228 \quad < \quad \approx 0.0367$



Makes sense $E(\overline{X}-\mu) = 0$

↳ numerator of t

- Mean and variance.
  - If $T \sim t_r$, then

$$E(T) = 0 \quad \text{if } r > 1$$ only exists if 2 or more df 3

$$V(T) = \frac{r}{r-2} \quad \text{if } r > 2$$ only exists if or more df

- Moments and mgf.
  - In general, if there are r degrees of freedom, then there are only $r-1$ moments (recall first moment is $E(X)$ and $V(X)$ is the second central moment).
  - Student's t distribution does not have an mgf.
  - Informal theorem for existence of mgf:

    If the mgf exists, then all moments exist. But the opposite is not true i.e. all moments existing doesn't always mean that the mgf exists).

    MGF exists $\Longrightarrow$ All moments exist
    $\not\Longleftarrow$

    If any moment doesn't exist, then the mgf doesn't exist. "missing" moment $\Longrightarrow$ No MGF

- Use in statistics.
  - The t-distribution is very important in inferential statistics and is used in one sample tests and confidence intervals of populations means, as well as simple linear regression for testing a population slope.

Derivation of the F distribution

- Another important derived distribution is Snedecor's F, whose derivation is quite similar to that of Student's t.

- Setup: Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu_X, \sigma_X^2)$ population, and let $Y_1, \ldots, Y_m$ be a random sample from an independent $N(\mu_Y, \sigma_Y^2)$ population.
- Goal: If we were interested in comparing the variability of the populations, one quantity of interest would be the ratio of population variances $\dfrac{\sigma_X^2}{\sigma_Y^2}$. ⟶ If similar ≈ 1

  We can estimate this with ratio of sample variances $\dfrac{s_X^2}{s_Y^2}$

- The F distribution allows us to compare these quantities by giving us a distribution of

$$\frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2}$$

- Again we have the form of the statistic and need to rearrange until we have the form of some familiar distributions (and then the pdf can be found).

$$\frac{\text{ratio of sample variances}}{\text{ratio of population variances}} = \frac{s_X^2/s_Y^2}{\sigma_X^2/\sigma_Y^2}$$

$$= \frac{\left(\frac{n-1}{\sigma_X^2} s_X^2\right)\frac{1}{n-1}}{\left(\frac{m-1}{\sigma_Y^2} s_Y^2\right)\left(\frac{1}{m-1}\right)}$$

$$= \frac{\chi^2_{n-1}/(n-1)}{\chi^2_{m-1}/(m-1)}$$

= ratio of chi - squares ÷ their respective df

- The distribution of the above can be founded by solving the simplified problem of finding the distribution of $\dfrac{X_1/r_1}{X_2/r_2}$, where $X_1 \sim \chi^2(r_1)$, $X_2 \sim \chi^2(r_2)$ and $X_1 \perp\!\!\!\perp X_2$. Again we will not show this.

Definition of the F distribution

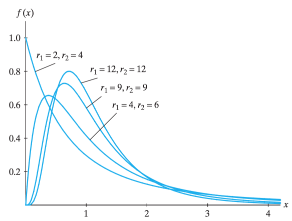
- Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu_X, \sigma_X^2)$ population, and let $Y_1, \ldots, Y_m$ be a random sample from an independent $N(\mu_Y, \sigma_Y^2)$ population. If

$$W = \frac{S_X^2/S_Y^2}{\sigma_X^2/\sigma_Y^2} \qquad \text{then} \qquad W \sim F(n-1, m-1)$$

In general, if W has an F distribution with r_1 numerator degrees of freedom and r_2 denominator degrees of freedom, then we write $W \sim F(r_1, r_2)$.

Notes about the F distribution

- Density curve (more probability gets centered around 1 as *df* increase; range $0 < F < \infty$).



- Pdf, mean, variance and mgf (which doesn't exist).
  - We aren't going to worry about the pdf (it is more ugly than the t pdf) or ever calculate means and variances for this distribution.
- Probabilities.
  - Again, we need software such as R to calculate these.
  - Example: Let $X_1 \sim F(8, 4)$ and $X_2 \sim F(16, 8)$. Find the following probabilities.

    $P(X_1 > 5) =$ Fcdf (lower = 5, upper = 10000, dfNumer = 8, dfDenom = 4) ≈ 0.0686

    $P(X_2 > 5) =$ Fcdf (5, 10000, 16, 8) ≈ 0.0134

- Relationship to other distributions.
  - Theorem:

    (a) If $X \sim F(r_1, r_2)$ then $1/X \sim F(r_2, r_1)$.

    The reciprocal of an F variable is again an F variable.

    $$F_{(r_1, r_2)} = \frac{\chi^2_{r_1}/r_1}{\chi^2_{r_2}/r_2} \xrightarrow{\text{Flip}} \frac{\chi^2_{r_2}/r_2}{\chi^2_{r_1}/r_1} \overset{\checkmark}{=} F_{(r_2, r_1)}$$

    (b) If $X \sim t_r$ then $X^2 \sim F(1, r)$.

    The square of a t random variable is an F variable with 1 and r *df*.

    $$T^2 = \left(\frac{Z}{\sqrt{\chi^2_r/r}}\right)^2 = \frac{Z^2}{\chi^2_r/r} = \frac{\chi^2_1/1}{\chi^2_r/r} \overset{\checkmark}{=} F_{(1,r)}$$

- Use in statistics.
  - The F-distribution is also very important in inferential statistics and is used in ANOVA when comparing means of two populations and also has lots of applications in regression.