Lecture 1: Sampling distribution

1.1 Background

Definition 1.1 A **random sample** is a set of mutually independent and identically distributed (i.i.d) random variables.

Definition 1.2 Once we model the data generating distribution, the variables associated with the distribution are called the **(population) parameters**, and the set of all possible population parameters is called the **parameter space**.

Definition 1.3 A function of a random sample, called a **statistic**, is used to make inferences about the population distribution and parameters.

Definition 1.4 A statistic is also a random variable, and its distribution is called the **sampling distribution**, which is determined by the population distribution.

Example 1.1 On the defect rate of a factory producing computers, we can model the population distribution by a Bernoulli distribution $\text{Ber}(p)$, where p represents the defect rate.

- p (population parameter)
- $\Omega = \{p : 0 \le p \le 1\}$ (parameter space)
- $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Ber}(p)$ means $X_1, \ldots, X_n$ is a random sample of n units from $\text{Ber}(p)$
- Examples of statistics: the number of defective computers $u_1 = X_1 + \cdots + X_n$, and the sample proportion $u_2 = (X_1 + \cdots + X_n)/n$.

Example 1.2 In many scientific experiments, the distribution of errors is modeled using the normal distribution $\mathcal{N}(\mu, \sigma^2)$.

- $\theta = (\mu, \sigma^2)^{\mathrm{T}}$, corresponding to the mean and variance (population parameters)
- $\Omega = \{(\mu, \sigma^2)^{\mathrm{T}} : -\infty < \mu < +\infty, \sigma^2 > 0\}$ (parameter space)
- $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ (a random sample)
- Examples of statistics: the sample mean and sample variance, respectively:

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n), \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

(statistics for inferring the population mean and the population variance)

Remark 1.1

- Learning the sampling distribution $\equiv$ learning the distribution of a function of random variables (e.g., function of a random sample).
- General idea: if the random variable $\boldsymbol{X} = (X_1, \ldots, X_n)^{\mathrm{T}}$ has a distribution, then to find the distribution of $Y = u(\boldsymbol{X}) = u(X_1, \ldots, X_n)$, we need to use change of variables and calculate the probability corresponding to Y in terms of $\boldsymbol{X}$. In other words, derive the distribution of $Y = u(X)$ if we know the distribution of X and some transformation u.
- Goal: derive popular sampling distributions (χ^2, t, and F).

1.2 Change of variables (transformation)

Theorem 1.1 *Suppose X_1 and X_2 are random variables with supports $\mathbb{X}_1$ and $\mathbb{X}_2$, respectively. Let $f_{X_1,X_2}(x_1, x_2)$ denote their joint pdf. X_1 and X_2 are independent if and only if $f_{X_1,X_2}(x_1, x_2)$ can be written as a product of nonnegative functions of x_1 and x_2, respectively. In other words,*

$$f_{X_1,X_2}(x_1, x_2) = g(x_1) \cdot h(x_2)$$

where $g(x_1) > 0, \forall x_1 \in \mathbb{X}_1$ and $h(x_2) > 0, \forall x_2 \in \mathbb{X}_2$.

Definition 1.5 Transformation of **discrete** random variable:

$$P(Y = y) = P(u(X) = y) = \sum_{x:u(x)=y} P(X = x)$$

$$\Longrightarrow \mathrm{pmf}_Y(y) = \sum_{x:u(x)=y} \mathrm{pmf}_X(x)$$

Example 1.3 If X_1 and X_2 are independent random variables following $\mathrm{Bin}(n_1, p)$ and $\mathrm{Bin}(n_2, p)$, respectively. If $Y = X_1 + X_2$, what is its distribution and pmf?

Solution. We know the pmf for a binomial random variable, say $X \sim \mathrm{Bin}(n, p)$, is given by $\mathrm{pmf}_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$. Next,

$$
\begin{aligned}
\mathrm{pmf}_Y(y) = P(Y = y) &= P(X_1 + X_2 = y) \\
&= \sum_{x_1=0}^{y} P(X_1 = x_1, X_2 = y - x_1) \\
&= \sum_{x_1=0}^{y} P(X_1 = x_1) P(X_2 = y - x_1) \quad \text{(by independence)} \\
&= \sum_{x_1=0}^{y} \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1 - x_1} \binom{n_2}{y - x_1} p^{y - x_1} (1-p)^{n_2 - y + x_1} \\
&= \left\{ \sum_{x_1=0}^{y} \binom{n_1}{x_1} \binom{n_2}{y - x_1} \right\} p^y (1-p)^{n_1 + n_2 - y} \\
&= \binom{n_1 + n_2}{y} p^y (1-p)^{n_1 + n_2 - y}, \quad y = 0, 1, \ldots, n_1 + n_2.
\end{aligned}
$$

Therefore, $Y \sim \mathrm{Bin}(n_1 + n_2, p)$. ■

Remark 1.2 Transformation of **continuous** random variable [not precise]:

$$
\begin{aligned}
&P(Y \in [y, y + \Delta y]) = P(u(X) \in [y, y + \Delta y]) \approx \sum_{x: u(x) = y} P(X \in [x, x + \Delta x]) \\
&\implies \mathrm{pdf}_Y(y) |\Delta y| \approx \sum_{x: u(x) = y} \mathrm{pdf}_X(x) |\Delta x| \\
&\implies \mathrm{pdf}_Y(y) \approx \sum_{x: u(x) = y} \mathrm{pdf}_X(x) \left| \frac{\Delta x}{\Delta y} \right|
\end{aligned}
$$

Theorem 1.2 *Let X be a continuous random variable with pdf $f_X(x)$ and support $\mathcal{S}_X$ (i.e., $\forall x \in \mathcal{S}_X, f_X(x) > 0$). Let $Y = g(X)$, where $g(x)$ is a* ***one-to-one differentiable*** *function on S_X. The inverse of g is denoted $x = g^{-1}(y)$ and $dx/dy = d[g^{-1}(y)]/dy$. Then, the pdf of Y is*

$$
f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{dx}{dy} \right|, \qquad \text{for } y \in \mathcal{S}_Y
$$

where $\mathcal{S}_Y = \{y = g(x) \mid x \in \mathcal{S}_X\}$.

Proof. Because $g(x)$ is one-to-one and continuous, it is either monotonically increasing or decreasing. Assume that it is strictly monotonically increasing first. Then, the cdf for Y is

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))$$

Finding the pdf, we differentiate with respect to y,

$$f_Y(y) = \frac{d}{dy}F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \cdot \frac{dg^{-1}(y)}{dy} = f_X(g^{-1}(y)) \cdot \frac{dx}{dy}$$

Since g is assumed to be monotonically increasing, we know $dx/dy = |dx/dy|$.
[Exercise: finish proof assuming g is monotonically decreasing. Reference Hogg & Craig for help.]

□

Example 1.4 Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Find the distribution of $Y = u(X) = (X - \mu)/\sigma$.
Solution. First, (1) u is one-to-one function from $\mathbb{R}$ to $\mathbb{R}$, (2) differentiable in $\mathbb{R}$ with $du(x)/dx = 1/\sigma \neq 0$ for $x \in \mathbb{R}$. Then, according to Theorem 1.2

$$\begin{aligned}
\mathrm{pdf}_Y(y) &= \mathrm{pdf}_X(\sigma y + \mu)\left|du^{-1}(y)/dy\right| \\
&= \sigma \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\sigma y + \mu - \mu)^2} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}.
\end{aligned}$$

Therefore, $Y \sim \mathcal{N}(0, 1)$. In other words, we can say $X \overset{\mathrm{d}}{\equiv} \sigma Y + \mu$. (Note: μ and σ are often referred to as "location" and "scale" parameters, respectively.) ■

Theorem 1.3 **(Multivariate transformation)** *Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T \in \mathcal{S}$ be a continuous n-dimensional random variable and $\boldsymbol{u} = (u_1, \ldots, u_n)^T : \mathcal{S} \to \mathcal{T}$ be an n-dimensional function that maps $\mathcal{S}$ onto $\mathcal{T}$. Next, define $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^T = (u_1(\boldsymbol{X}), \ldots, u_n(\boldsymbol{X}))^T \in \mathcal{T}$. Assuming*

- ***$\boldsymbol{u}$ is one-to-one***
- *the first partial derivatives of the inverse functions (e.g., $\partial x_i/\partial y_j$ $i, j \in \{1, \ldots, n\}$) are continuous (let the inverse functions be denoted as $x_j = w_j(y_1, \ldots, y_n)$ for $j \in \{1, \ldots, n\}$)*
- *the Jacobian (J) is nonzero in $\mathcal{T}$*

then the joint pdf of $\boldsymbol{Y}$ is given by

$$f_{\boldsymbol{Y}}(y_1, \ldots, y_n) = f_{\boldsymbol{X}}(w_1(y_1, \ldots, y_n), \ldots, w_n(y_1, \ldots, y_n)) \cdot |J|$$

for $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{T}$, and 0 elsewhere ("the joint pdf of n functions of n random variables" (Hogg & Craig 2019)).

Example 1.5 Let X_1 and X_2 be independent random variables with $X_1 \sim \text{Gamma}(\alpha_1, \beta)$ and $X_2 \sim \text{Gamma}(\alpha_2, \beta)$. Suppose $Y_1 = X_1/(X_1 + X_2)$ and $Y_2 = X_1 + X_2$. Find the joint probability density function of Y_1 and Y_2 and the marginal probability density functions of each.
Solution. The joint pdf of X_1 and X_2 (by independence) is

$$\text{pdf}_{X_1,X_2}(x_1, x_2) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} x_1^{\alpha_1-1} x_2^{\alpha_2-1} e^{-(x_1+x_2)/\beta} \mathbf{1}_{(x_1,x_2>0)}$$

Let $u(x_1, x_2) = \left(\frac{x_1}{x_1+x_2}, x_1 + x_2\right)^{\mathrm{T}}$ denote the transformation function. The domain of u is $\mathbb{X} = \{(x_1, x_2)^{\mathrm{T}} : \text{pdf}_{X_1,X_2}(x_1, x_2) > 0\} = \{(x_1, x_2)^{\mathrm{T}} : x_1 > 0, x_2 > 0\}$. Next, we need to verify that the function satisfies the conditions of Theorem 1.3. Given $u(x_1, x_2) = (y_1, y_2)^{\mathrm{T}}$, the following holds:

$$u : \begin{cases} y_1 = \frac{x_1}{x_1+x_2} \\ y_2 = x_1 + x_2 \end{cases} \qquad \text{implies} \quad u^{-1} : \begin{cases} x_1 = y_1 y_2 \\ x_2 = y_2(1 - y_1) \end{cases}$$

The range of the function is $\mathbb{Y} = \{(y_1, y_2)^{\mathrm{T}} : y_1 y_2 > 0, y_2(1 - y_1) > 0\} = \{(y_1, y_2)^{\mathrm{T}} : 0 < y_1 < 1, y_2 > 0\}$. The Jacobian determinant of u^{-1} is

$$J_{u^{-1}} = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \det \begin{pmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{pmatrix} = y_2$$

Thus, the joint probability density function of Y_1 and Y_2 is

$$\begin{aligned} \text{pdf}_{Y_1,Y_2}(y_1, y_2) &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} (y_1 y_2)^{\alpha_1-1} (y_2(1 - y_1))^{\alpha_2-1} e^{-y_2/\beta} y_2 \mathbf{1}_{(0,1)}(y_1) \mathbf{1}_{(0,\infty)}(y_2) \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1-1} (1 - y_1)^{\alpha_2-1} y_2^{\alpha_1+\alpha_2-1} e^{-y_2/\beta} \mathbf{1}_{(0,1)}(y_1) \mathbf{1}_{(0,\infty)}(y_2). \end{aligned}$$

Since the density is a product of functions of y_1 and y_2, Y_1 and Y_2 are independent, and their marginal densities are:

$$\text{pdf}_{Y_1}(y_1) = \int_0^\infty \text{pdf}_{Y_1,Y_2}(y_1, y_2) dy_2 = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1} (1 - y_1)^{\alpha_2-1} \mathbf{1}_{(0,1)}(y_1),$$

$$\text{pdf}_{Y_2}(y_2) = \int_0^1 \text{pdf}_{Y_1,Y_2}(y_1, y_2) dy_1 = \frac{1}{\Gamma(\alpha_1 + \alpha_2)\beta^{\alpha_1+\alpha_2}} y_2^{\alpha_1+\alpha_2-1} e^{-y_2/\beta} \mathbf{1}_{(0,\infty)}(y_2).$$

Thus we see $Y_1 \sim \text{Beta}(\alpha_1, \alpha_2)$ and $Y_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$. ■

Definition 1.6 If $X \sim \text{Beta}(\alpha_1, \alpha_2)$, then X can be represented as $X \equiv Z_1/(Z_1 + Z_2)$, where $Z_i \sim \text{Gamma}(\alpha_i, \beta)$ $(i = 1, 2)$, independent. Then, the pdf of X is

$$\text{pdf}_X(x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1-1} (1 - x)^{\alpha_2-1} \mathbf{1}_{(0,1)}(x).$$

Remark 1.3 What happens when the transformation, u, is no longer one-to-one, but instead "many-to-one" (or m-to-one for $m > 1$)? It can best be answered through Example 1.6.

Example 1.6 Let $X \sim \text{Unif}(-1, 1)$. Find the probability density function of $Y = X^2$ where $\mathbb{X} = (-1, 0) \cup (0, 1)$ to $\mathbb{Y} = (0, 1)$.

Solution. The function $u(x) = x^2$ is a many-to-one function from $\mathbb{X} = (-1, 0) \cup (0, 1)$ to $\mathbb{Y} = (0, 1)$ where $m = 2$. In this case, we need to represent the domain $\mathbb{X}$ as a disjoint union of sets such that u is one-to-one for each distinct disjoint set of $\mathbb{X}$.

Let $A_1 = \{x \mid -1 < x < 0\}$ and $A_2 = \{x \mid 0 < x < 1\}$. In this case, $A_1 \cup A_2 = \mathbb{X}$ and u maps A_1 to $\{y \mid 0 < y < 1\} = \mathbb{Y}$, and the same for A_2. Given A_1 and A_2 meet this condition where u is a one-to-one from each domain (A_1, A_2) to $\mathbb{Y}$, we need the m-distinct inverse mappings from $\mathbb{Y}$ to A_1 and A_2, respectively. In other words

$$\begin{cases} w_1(y) = -\sqrt{y} & \text{maps } \mathbb{Y} \to A_1 \\ w_2(y) = \sqrt{y} & \text{maps } \mathbb{Y} \to A_2. \end{cases}$$

Let us recall that the goal is to learn the pdf of Y. A more intuitive question to answer is, what is $P(Y \in B)$ for $B \subset \mathbb{Y}$? We know $Y \in B$ if and only if $X \in \mathcal{B}_1 = \{x \mid x = -\sqrt{y}, y \in B\}$ or $X \in \mathcal{B}_2 = \{x \mid x = \sqrt{y}, y \in B\}$. Therefore $P(Y \in B) = P(X \in \mathcal{B}_1) + P(X \in \mathcal{B}_2)$, or in other words

$$P(Y \in B) = P(X \in \mathcal{B}_1) + P(X \in \mathcal{B}_2) = \int_{\mathcal{B}_1} f_X(x)dx + \int_{\mathcal{B}_2} f_X(x)dx$$

and for each of those integrals, we can use our approaches from before the transform them in terms of Y! Hence, the pdf for Y can be written as

$$\begin{aligned} f_Y(y) &= f_X(-\sqrt{y}) \cdot \left|\frac{-1}{2\sqrt{y}}\right| + f_X(\sqrt{y}) \cdot \left|\frac{1}{2\sqrt{y}}\right|, \quad y \in \mathbb{Y} \\ &= \frac{1}{2} \cdot \frac{1}{2\sqrt{y}} + \frac{1}{2} \cdot \frac{1}{2\sqrt{y}} \cdot \mathbf{1}_{y \in (0,1)} \\ &= \frac{1}{2\sqrt{y}} \cdot \mathbf{1}_{y \in (0,1)} \end{aligned}$$

■

Note: this process can be abstracted to any $m > 1$ and follows the same steps of breaking the domain into m disjoint sets and finding each inverse over each domain, then summing the distinct densities. [See Example 2.7.3 in Hogg & Craig for $m = 4$.]

Remark 1.4 The most important example for the case of multivariate transformations of m-to-one functions is **order statistics**, which we study in the next section. If the population distribution is continuous and its probability density function is $f(x)$, the order statistics of a random sample $X_1, X_2, \ldots, X_n$ arranged in increasing order are denoted as $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$. If the function representing the order statistics is $u(X_1, \ldots, X_n) = (X_{(1)}, \ldots, X_{(n)})^{\mathrm{T}}$, then the function u from $\mathbb{X} = \{(x_1, \ldots, x_n)^{\mathrm{T}} : f(x_i) > 0\,(i = 1, \ldots, n),\, x_1, \ldots, x_n$ are distinct real numbers$\}$ to $\mathbb{Y} = \{(y_1, \ldots, y_n)^{\mathrm{T}} : f(y_i) > 0\,(i = 1, \ldots, n),\, y_1 < \cdots < y_n\}$, is a $n!$-to-1 mapping that satisfies the conditions for a multivariate transformation.

1.3 Popular sampling distributions

We will focus on the following sampling distributions: χ^2 distribution, t ditsribution, F distribution. These three distributions are important as they relate to the sample mean and sample variance when the population distribution is assumed to be normally distributed.

1.3.1 χ^2 (chi-square) distribution

Definition 1.7 Suppose $X_1, \ldots, X_r \overset{iid}{\sim} N(0, 1)$ (i.e., $X_1, \ldots, X_r$ are independent and identically distributed as Normal with mean 0 and variance 1). Then, $Y = X_1^2 + \cdots + X_r^2$ follows a $\boldsymbol{\chi^2}$ **distribution** with degrees of freedom r, denoted by $Y \sim \chi^2(r)$.

Remark 1.5 It can be shown that $\chi^2(r) \overset{d}{\equiv} \text{Gamma}(r/2, 2)$ where the pdf of $Y \sim \chi^2(r)$ is

$$\text{pdf}_Y(y) = \frac{1}{\Gamma(r/2)2^{r/2}} y^{(r/2)-1} e^{-y/2} \cdot \mathbf{1}_{(0,\infty)}(y).$$

Proof. Homework. $\square$

1.3.2 t distribution

Definition 1.8 We say X follows a $\boldsymbol{t}$ **distribution** with degree of freedom $r > 0$, denoted by $X \sim t_r$ if and only if

1. $X \overset{d}{\equiv} \frac{Z}{\sqrt{V/r}}$, $\quad Z \sim N(0, 1)$, $\quad V \sim \chi^2(r)$, $\quad Z$ and V are independent, or
2. $\text{pdf}_X(x) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{r}{2}\right)\sqrt{r}} \left(1 + \frac{x^2}{r}\right)^{-\frac{r+1}{2}}$.

Remark 1.6 Items (1) and (2) from Definition 1.8 are equivalent.

Proof. Given (1), let $X = \frac{Z}{\sqrt{V/r}}$ and $Y = V$. Our goal is to get the pdf of X, thus we need a transformation from the distributions we know (Z, V) to the distributions we care about (X, Y). The inverse relations are given by

$$Z = X\sqrt{\frac{Y}{r}}, \quad V = Y,$$

and because this is a one-to-one transformation, the joint pdf of X and Y, by Theorem 1.3, is:

$$\begin{aligned}
\mathrm{pdf}_{X,Y}(x,y) &= \mathrm{pdf}_{Z,V}(z,v)\left|\det\left(\frac{\partial(z,v)}{\partial(x,y)}\right)\right| = \mathrm{pdf}_Z(z)\mathrm{pdf}_V(v)\left|\det\begin{pmatrix}\frac{\partial z}{\partial x} & \frac{\partial z}{\partial y}\\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\end{pmatrix}\right| = \\
&= \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\frac{1}{\Gamma(r/2)2^{r/2}}v^{r/2-1}e^{-v/2}\mathbf{1}_{(v>0)}\left|\det\begin{pmatrix}\frac{\partial z}{\partial x} & \frac{\partial z}{\partial y}\\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\end{pmatrix}\right| \\
&= \frac{1}{\sqrt{2\pi}}e^{-x^2y/2r}\frac{1}{\Gamma(r/2)2^{r/2}}y^{r/2-1}e^{-y/2}\mathbf{1}_{(y>0)}\left|\det\begin{pmatrix}\sqrt{\frac{y}{r}} & \frac{x}{2\sqrt{ry}}\\ 0 & 1\end{pmatrix}\right| \\
&= \frac{1}{\Gamma(1/2)\Gamma(r/2)2^{(r+1)/2}\sqrt{r}}y^{(r+1)/2-1}\exp\left(-\frac{y}{2}\cdot\left(1+\frac{x^2}{r}\right)\right)\mathbf{1}_{(y>0)}.
\end{aligned}$$

From this, the marginal probability density function of X is given by

$$\begin{aligned}
\mathrm{pdf}_X(x) &= \int_0^\infty \mathrm{pdf}_{X,Y}(x,y)\,dy \\
&= \frac{1}{\Gamma(1/2)\Gamma(r/2)2^{(r+1)/2}\sqrt{r}}\int_0^\infty y^{(r+1)/2-1}e^{-(1+x^2/r)y/2}\,dy
\end{aligned}$$

Then, rather than computing the integral, it is often easier to do "kernel matching" to a distribution where we know the integral resolves to 1. In particular, the pdf of $Y \sim \mathrm{Gamma}(\alpha,\theta)$ is

$$f_Y(y) = \frac{1}{\Gamma(\alpha)\theta^\alpha}y^{\alpha-1}\exp(-y/\theta).$$

Hence, by letting $(1+x^2/r)y/2 = t$,

$$\begin{aligned}
\mathrm{pdf}_X(x) &= \frac{1}{\Gamma(1/2)\Gamma(r/2)2^{(r+1)/2}\sqrt{r}}\int_0^\infty t^{(r+1)/2-1}e^{-t}\,dt\left(\frac{2}{1+x^2/r}\right)^{(r+1)/2} \\
&= \frac{1}{\Gamma(1/2)\Gamma(r/2)2^{(r+1)/2}\sqrt{r}}\left(\frac{2}{1+x^2/r}\right)^{(r+1)/2}\Gamma((r+1)/2)\int_0^\infty \frac{1}{\Gamma((r+1)/2)}t^{(r+1)/2-1}e^{-t}\,dt \\
&= \frac{\Gamma((r+1)/2)}{\Gamma(1/2)\Gamma(r/2)\sqrt{r}}\left(1+\frac{x^2}{r}\right)^{-(r+1)/2}.
\end{aligned}$$

□

Remark 1.7 Given a random sample $X_1, X_2, \ldots, X_n$ from a normal distribution $\mathcal{N}(\mu,\sigma^2)$, suppose that we do not know the values of both μ and σ^2. Recall from introductory statistics that

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1},$$

where S is the sample standard deviation (see Theorem 1.4 for proof). We can use this centered and scaled statistic to make inference for μ by using the confidence interval

$$P\left(\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right) = 1-\alpha.$$

Remark 1.8 The t-distribution has no mgf because it does not have moments of all orders. In fact, if there are p degrees of freedom, then there are only $p-1$ moments. Hence, a t_1 has no mean, a t_2 has no variance, etc. It is easy to check that if T_p is a random variable with a t_p distribution, then

$$\mathbb{E}(T_p) = 0, \quad \text{if} \;\; p > 1$$

and

$$\mathrm{Var}(T_p) = \frac{p}{p-2}, \text{ if } p > 2$$

(Tip: use the item (1) from Definition 1.8).

Theorem 1.4 *Given $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, the following holds:*

(a) The sample mean $\bar{X} = (X_1 + \cdots + X_n)/n$ follows $\mathcal{N}(\mu, \sigma^2/n)$.

(b) The centered samples $(X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_n - \bar{X})$ is independent of the sample mean $\bar{X}$. The sample variance $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ is independent of the sample mean $\bar{X}$.

(c) $(n-1)S^2/\sigma^2$ follows a χ^2 distribution with $(n-1)$ degrees of freedom. That is,

$$(n-1)S^2/\sigma^2 = \sum_{i=1}^n (X_i - \bar{X})^2/\sigma^2 \sim \chi^2(n-1)$$

(d) $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$.

Proof.

(a) Using the properties of mgfs

$$\begin{aligned}
\mathrm{mgf}_{\bar{X}}(s) &= \mathbb{E}e^{s\bar{X}} = \mathbb{E}e^{s(X_1+\cdots+X_n)/n} \\
&= \mathrm{mgf}_{X_1+\cdots+X_n}\left(\frac{s}{n}\right) = \mathrm{mgf}_{X_1}\left(\frac{s}{n}\right)\cdots \mathrm{mgf}_{X_n}\left(\frac{s}{n}\right) \\
&= \prod_{i=1}^n \exp\left\{\mu\left(\frac{s}{n}\right) + \sigma^2\left(\frac{s}{n}\right)^2/2\right\} = \exp\left\{\mu s + \left(\frac{\sigma^2}{n}\right)\frac{s^2}{2}\right\},
\end{aligned}$$

which is the mgf for a normal distribution with mean μ and variance σ^2/n.

(b) The sample variance

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

is a function of $(X_1 - \bar{X}, \ldots, X_n - \bar{X})^{\mathrm{T}}$. If we show that $\bar{X}$ and $(X_1 - \bar{X}, \ldots, X_n - \bar{X})^{\mathrm{T}}$ are independent, we know that $\bar{X}$ and S^2 are independent since S^2 is a function of $(X_1 -$

$\bar{X}, \ldots, X_n - \bar{X})^{\mathrm{T}}$. In other words, independence is preserved under shifting and scaling transformations. We can use the properties of mgfs to show independence between $\bar{X}$ and $\boldsymbol{Y} = (X_1 - \bar{X}, \ldots, X_n - \bar{X})^{\mathrm{T}}$ by showing

$$\mathrm{mgf}_{\bar{X},\boldsymbol{Y}}(s, \boldsymbol{t}) = \mathrm{mgf}_{\bar{X}}(s) \cdot \mathrm{mgf}_{\boldsymbol{Y}}(\boldsymbol{t})$$

$$\begin{aligned}
\mathrm{mgf}_{\bar{X},\boldsymbol{Y}}(s, \boldsymbol{t}) &= \mathbb{E}\left[\exp\left\{s\bar{X} + \boldsymbol{t}^T\boldsymbol{Y}\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{s\bar{X} + t_1(X_1 - \bar{X}) + \ldots + t_n(X_n - \bar{X})\right\}\right] \\
&= \mathbb{E}\left[\exp\left\{\left(\frac{s}{n} + (t_1 - \bar{t})\right) X_1 + \ldots + \left(\frac{s}{n} + (t_n - \bar{t})\right) X_n\right\}\right] \\
&= \prod_{i=1}^{n} \mathrm{mgf}_{X_i}\left(\frac{s}{n} + (t_i - \bar{t})\right) \quad \text{(by independence)} \\
&= \exp\left\{\sum_{i=1}^{n}\left[\mu\left(\frac{s}{n} + (t_i - \bar{t})\right)\right] + \frac{1}{2}\sigma^2\left(\frac{s}{n} + (t_i - \bar{t})\right)^2\right\} \\
&= \exp\left\{\mu s + \frac{1}{2}\sigma^2\frac{s^2}{n}\right\} \exp\left\{\frac{1}{2}\sigma^2\sum_{i=1}^{n}(t_i - \bar{t})^2\right\}
\end{aligned}$$

The joint moment generating function of $\bar{X}$ and $\boldsymbol{Y}$ can be expressed as the product of the marginal moment generating functions, so they are independent.

(c) Understanding that $\sum_{i=1}^{n} X_i - \bar{X} = 0$, we know the following relationship

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}\left\{(X_i - \bar{X}) + (\bar{X} - \mu)\right\}^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2
\end{aligned}$$

Next, define

$$U = \sum_{i=1}^{n}\frac{(X_i - \mu)^2}{\sigma^2}, \quad V = \sum_{i=1}^{n}\frac{(X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2}, \quad W = \frac{n(\bar{X} - \mu)^2}{\sigma^2}.$$

The distribution of V is what we need to find. First, note that V and W are independent from (b) and $U = V + W$. Therefore, the moment-generating function (MGF) of U is given by:

$$\mathrm{mgf}_U(t) = \mathrm{mgf}_V(t) \times \mathrm{mgf}_W(t)$$

Given $(X_i - \mu)/\sigma \overset{iid}{\sim} N(0, 1)$ for $i = 1, \ldots, n$, we can use Definition 1.7 to get

$$U = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n), \quad \Longrightarrow \quad \mathrm{mgf}_U(t) = (1 - 2t)^{-n/2} I_{(t<1/2)}$$

Then, from (a), $\sqrt{n}(\bar{X}-\mu)/\sigma$ follows the standard normal distribution $N(0,1)$. Therefore:

$$W = \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1), \quad \Longrightarrow \quad \text{mgf}_W(t) = (1-2t)^{-1/2} I_{(t<1/2)}$$

Thus, the moment-generating function for V is

$$\text{mgf}_V(t) = (1-2t)^{-(n-1)/2} I_{(t<1/2)}, \quad \Longrightarrow \quad V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

(d)

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \frac{1}{\sqrt{(n-1)S^2/\sigma^2/(n-1)}} \overset{d}{\equiv} \frac{N(0,1)}{\sqrt{\chi^2(n-1)/(n-1)}} \sim t_{n-1}$$

□

1.3.3 F-distribution

Definition 1.9 We say X follows an ***F*-distribution** with degrees of freedom r_1, r_2, denoted by $X \sim F(r_1, r_2)$ $(r_i > 0, i = 1, 2)$, if and only if

1. $X \overset{d}{\equiv} \frac{V_1/r_1}{V_2/r_2}, \quad V_i \sim \chi^2(r_i) \quad (i = 1, 2), \quad V_1$ and V_2 are independent, or
2. $\text{pdf}_X(x) = \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)}{\Gamma(\frac{r_1}{2})\Gamma(\frac{r_2}{2})} \left(\frac{r_1}{r_2}\right)^{r_1/2} x^{r_1/2-1} \left(1 + \frac{r_1 x}{r_2}\right)^{-(r_1+r_2)/2} \mathbf{1}_{(x>0)}$.

Remark 1.9 Suppose $X_{11}, X_{12}, \ldots, X_{1n_1} \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ and $X_{21}, X_{22}, \ldots, X_{2n_2} \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$, and assume the two random samples are independent. We can test whether or not the variances σ_1^2, σ_2^2 are the same (equality test) by using the following statistic:

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1),$$

where $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i, \quad S_i^2 = \sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2/(n_i - 1) \quad (i = 1, 2)$ [Homework: proving this distributional relationship]. We can do inference by computing the confidence interval of the ratio by

$$\text{P}\left\{\frac{S_1^2/S_2^2}{F_{\alpha/2}(n_1-1, n_2-1)} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2/S_2^2}{F_{\alpha/2}(n_2-1, n_1-1)}\right\} = 1 - \alpha.$$

Remark 1.10 The probability density function of the F-distribution is derived as follows:

Proof. From Definition 1.9, let $Y = V_2$ and X is provided in (1). We are interested in the pdf of

X, thus we transform (V_1, V_2) to (X, Y). The inverses are given by:

$$V_1 = \frac{r_1 XY}{r_2}, \quad V_2 = Y$$

Since this is a one-to-one transformation, the joint pdf of (X, Y) (according to Theorem 1.3) is

$$\begin{aligned}
\mathrm{pdf}_{X,Y}(x, y) &= \mathrm{pdf}_{V_1,V_2}(v_1, v_2) \left| \det \left(\frac{\partial(v_1, v_2)}{\partial(x, y)} \right) \right| \\
&= \prod_{i=1}^{2} \left[\frac{1}{\Gamma(r_i/2) 2^{r_i/2}} v_i^{r_i/2-1} e^{-v_i/2} \mathbf{1}_{(v_i>0)} \right] \left| \det \begin{pmatrix} r_1 y/r_2 & r_1 x/r_2 \\ 0 & 1 \end{pmatrix} \right| \\
&= \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2) 2^{(r_1+r_2)/2}} \left(\frac{r_1}{r_2} xy \right)^{r_1/2-1} y^{r_2/2-1} e^{-(1+r_1 x/r_2)y/2} \left(\frac{r_1}{r_2} y \right) \mathbf{1}_{(x>0, y>0)}.
\end{aligned}$$

From this, the marginal probability density function of X is given by

$$\begin{aligned}
\mathrm{pdf}_X(x) &= \int_0^\infty \mathrm{pdf}_{X,Y}(x, y)\, dy \\
&= \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2) 2^{(r_1+r_2)/2}} \left(\frac{r_1}{r_2} \right)^{r_1/2} x^{r_1/2-1} \mathbf{1}_{(x>0)} \int_0^\infty y^{(r_1+r_2)/2-1} e^{-(1+r_1 x/r_2)y/2}\, dy,
\end{aligned}$$

and letting $(1 + r_1 x/r_2)y/2 = t$,

$$\begin{aligned}
\mathrm{pdf}_X(x) &= \frac{1}{\Gamma(r_1/2)\Gamma(r_2/2) 2^{(r_1+r_2)/2}} \left(\frac{r_1}{r_2} \right)^{r_1/2} x^{r_1/2-1} \mathbf{1}_{(x>0)} \int_0^\infty t^{(r_1+r_2)/2-1} e^{-t} \left(\frac{2}{1 + r_1 x/r_2} \right)^{(r_1+r_2)/2} dt \\
&= \frac{\Gamma((r_1+r_2)/2)}{\Gamma(r_1/2)\Gamma(r_2/2)} \left(\frac{r_1}{r_2} \right)^{r_1/2} x^{r_1/2-1} \left(1 + \frac{r_1 x}{r_2} \right)^{-(r_1+r_2)/2} \mathbf{1}_{(x>0)}.
\end{aligned}$$

$\square$