Multivariate Distribution, Continuous Distribution, Random Sample

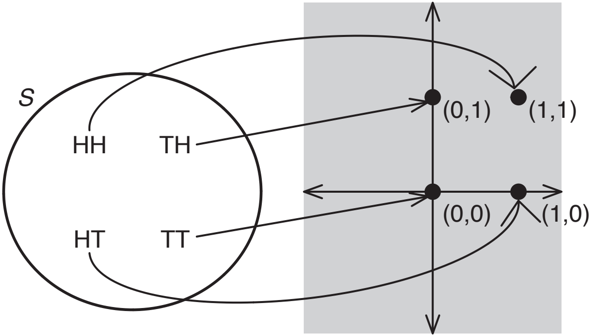
Keng-Chi Chang

Department of Political Science
University of California San Diego

2023-09-21

Multivariate Random Variables

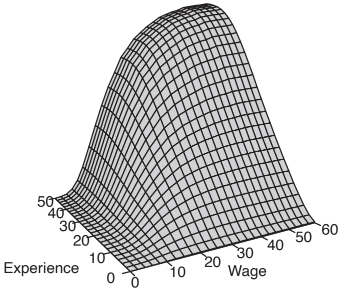
- We rarely care about a single random variable, but multiple random variables
- A random variable is a function maps from sample space S to the real line $\mathbb{R}$
- A **multivariate random variable** is a function maps from sample space S to $\mathbb{R}^n$

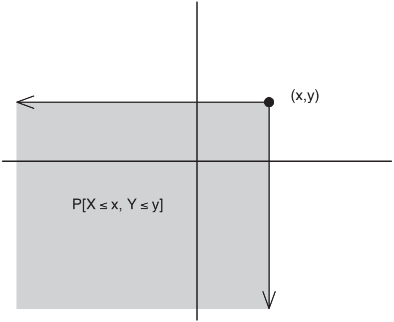


Joint Cumulative Distribution Function

- Joint Event $(A \cap B)$ or simply (A, B): The event that both A and B occur
- **Joint CDF**: The joint CDF of (X, Y) is

$$F_{X,Y}(x, y) = \mathbb{P}[X \leq x, Y \leq y] = \mathbb{P}[\{X \leq x\} \cap \{Y \leq y\}]$$





Joint Density and Mass Functions

- The joint distribution of (X, Y) is continuous if $F_{X,Y}(x, y)$ is continuous in (x, y)
- For **continuous** multivariate random variable (X, Y), its **Joint PDF** is

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

- The joint distribution of (X, Y) is discrete if $F_{X,Y}(x, y)$ is discrete in (x, y)
- For **discrete** multivariate random variable (X, Y), its **Joint PMF** is

$$f_{X,Y}(x, y) = \mathbb{P}[X = x, Y = y] \tag{1}$$

Marginal Density and Mass Functions

- We have the same **marginalization** properties for density and mass functions
- For **continuous** multivariate random variable (X, Y), its **Marginal PDFs** are

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx$$

- For **discrete** multivariate random variable (X, Y), its **Marginal PMFs** are

$$f_X(x) = \mathbb{P}[X = x] = \sum_{y \in \mathcal{Y}} \mathbb{P}[X = x, Y = y] = \sum_{y \in \mathcal{Y}} f_{X,Y}(x, y)$$

$$f_Y(y) = \mathbb{P}[Y = y] = \sum_{x \in \mathcal{X}} \mathbb{P}[X = x, Y = y] = \sum_{x \in \mathcal{X}} f_{X,Y}(x, y)$$

where $\mathcal{X} = \{x_1, x_2 \ldots\}$ denotes the support of X and $\mathcal{Y} = \{y_1, y_2 \ldots\}$ denotes the support of Y

Conditional Distribution, Density, and Mass

- We often want to know the distribution of Y given some variable $X = x$
  - E.g. How distribution of wage (Y) is different across gender (X)
- We can define the **Conditional Distribution** function of Y given $X = x$ as

$$F_{Y|X}(y|x) = \mathbb{P}[Y \leq y \,|\, X = x]$$

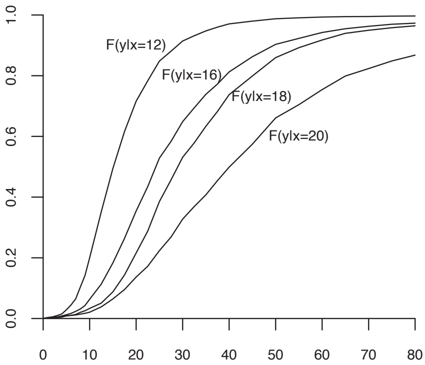
- We can also define the **Conditional Density/Mass** function of Y given $X = x$ as

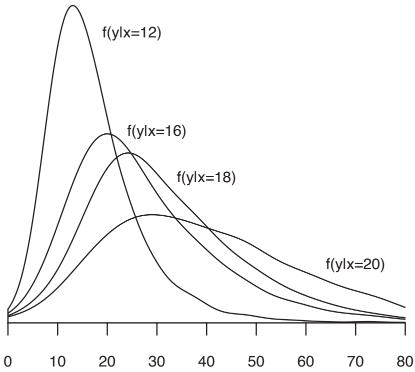
$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \qquad \text{for all } f_X(x) > 0$$

- **Product Rule** for density/mass:

$$f_{X,Y}(x, y) = f_{Y|X}(y|x) \cdot f_X(x) \qquad \text{for all } f_X(x) > 0$$

Conditional CDF and PDF/PMF





Independence of Random Variables

- We have defined the independence of two events, now for random variables
- Random variables X and Y are independent if and only if
  - Events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent; in other words

$$\mathbb{P}\left[\{X \leq x\} \cap \{Y \leq y\}\right] = \mathbb{P}[X \leq x] \; \mathbb{P}[Y \leq y] = F_X(x) \; F_Y(y)$$

- Independence between random variables X and Y (can be derived from above):
  - Based on CDFs:
    1. $F_{X|Y}(x \mid y) = F_X(x)$ for all x and y
    2. $F_{Y|X}(y \mid x) = F_Y(y)$ for all x and y
    3. $F_{X,Y}(x, y) = F_X(x) \; F_Y(y)$ for all x and y
  - Based on PDFs/PMFs:
    1. $f_{X|Y}(x \mid y) = f_X(x)$ for all x and y
    2. $f_{Y|X}(y \mid x) = f_Y(y)$ for all x and y
    3. $f_{X,Y}(x, y) = f_X(x) \; f_Y(y)$ for all x and y

Independence and Covariance

- Show that if X and Y are independent, then
  - $\mathbb{C}\mathrm{ov}[X, Y] = 0$
  - $\mathbb{V}\mathrm{ar}[X + Y] = \mathbb{V}\mathrm{ar}[X] + \mathbb{V}\mathrm{ar}[Y]$

Conditional Expectation Function (CEF)

- An important concept in regression is conditional expectation

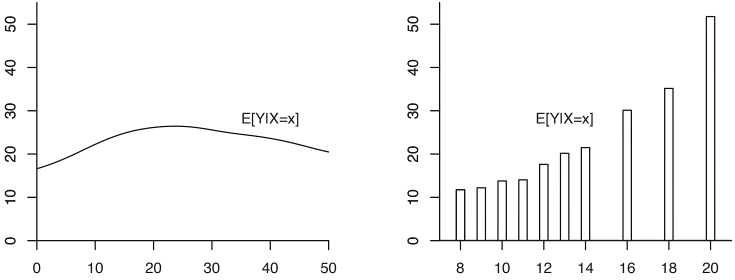
$$\mathbb{E}\left[Y \mid X_1, X_2, X_3\right] \approx \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

  - E.g. Y = wage, X_1 = gender, X_2 = race, X_3 = age
- The conditional expectation is the central tendency of a conditional distribution

$$\mathbb{E}[Y \mid X = x] = \begin{cases} \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|x)\, dy = \int_{-\infty}^{\infty} y \cdot \dfrac{f_{X,Y}(x,y)}{f_X(x)}\, dy & \text{if } X \text{ continuous} \\ \sum_{y \in \mathcal{Y}} y \cdot f_{Y|X}(y|x) = \sum_{y \in \mathcal{Y}} y \cdot \mathbb{P}[Y = y \mid X = x] & \text{if } X \text{ discrete} \end{cases}$$

- This tells us the average of Y given that X equals the **specific value** x

Conditional Expectation Function (CEF)



- When X is discrete, it is the expected value of Y within the sub-population for which $X = x$
  - ex. X is gender, $\mathbb{E}[Y|X = x]$ is the expected value of Y for men and women, separately
- When X is continuous, it is the expected value of Y within the infinitesimally small population for which $X \approx x$

Expectation of Conditional Expectation

- Notice that $m(x) = \mathbb{E}[Y \mid X = x]$ is a function of x
  - Once X is observed, $\mathbb{E}[Y \mid X = x]$ is a **known fixed number**
  - Before X is observed, $\mathbb{E}[Y \mid X] = m(X)$ is a **random variable**
- We can average $m(\cdot)$ across X (take expectation): $\mathbb{E}[m(X)] = \mathbb{E}[\mathbb{E}[Y \mid X]]$
- **Law of Iterated Expectations**:

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}_X[\mathbb{E}_{Y|X}[Y \mid X]] = \mathbb{E}[Y]$$

  - Intuition: Weighted average of $\mathbb{E}[Y \mid X = x]$, using $\mathbb{P}[X = x]$ as weights
  - The average across group averages is the grand average

Law of Iterated Expectations

- Special case when X is discrete:

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \sum_x \mathbb{E}[Y \mid X = x]\, \mathbb{P}[X = x] = \mathbb{E}[Y]$$

  - Can think of it as the **product rule** for conditional expectations
- Show that

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}_X[\mathbb{E}_{Y|X}[Y \mid X]] = \mathbb{E}[Y]$$

Properties of Conditional Expectation

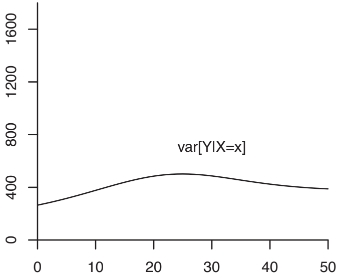
- $\mathbb{E}[\, g(X)\, Y \,|\, X \,] = g(X)\; \mathbb{E}[\, Y \,|\, X \,]$

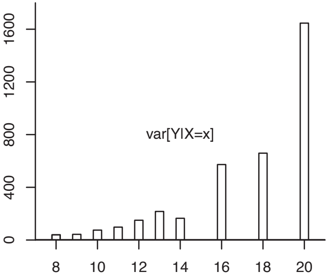
- $\mathbb{E}[\mathbb{E}[\, Y \,|\, X] \,|\, X \,] = \mathbb{E}[\, Y \,|\, X \,]$

Conditional Variance

- What about the variance of a conditional distribution?
- Similarly, we define the **Conditional Variance** as

$$\mathbb{V}\mathrm{ar}[Y \mid X = x] = \mathbb{E}\left[(Y - \mathbb{E}[Y \mid X = x])^2 \mid X = x\right]$$
$$= \mathbb{E}[Y^2 \mid X = x] - \mathbb{E}[Y \mid X = x]^2$$



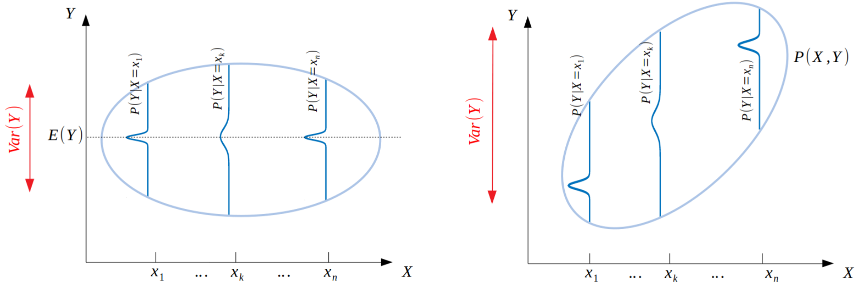


Conditional Variance

- **Law of Total Variance**:

$$\mathbb{V}\text{ar}[Y] = \mathbb{E}\left[\mathbb{V}\text{ar}\left[Y \mid X\right]\right] + \mathbb{V}\text{ar}\left[\mathbb{E}\left[Y \mid X\right]\right]$$

  - We can decompose the variability of a random variable Y into two parts:
    - Average variability "within" each values of X: $\mathbb{E}\left[\mathbb{V}\text{ar}\left[Y \mid X\right]\right]$
    - Variability of means "across" values of X: $\mathbb{V}\text{ar}\left[\mathbb{E}\left[Y \mid X\right]\right]$



Standard Normal Distribution

- $Z \sim \text{Normal}(0, 1)$ if Z has PDF

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \mathbb{1}(-\infty < z < \infty)$$

- The support of Z is $(-\infty, \infty)$
- The CDF of standard normal disribution is

$$\Phi(x) = F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}z^2}\, dz$$

- The normal distribution is the most commonly-used distribution
- The standard normal density function is typically written as $\phi(x)$, and the distribution function as $\Phi(x)$

Standard Normal Distribution

- $Z \sim$ Normal$(0, 1)$ if Z has PDF

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \mathbb{1}(-\infty < z < \infty)$$

- What is $\mathbb{E}[Z]$?
  - Hint: Show that $\int_0^{\infty} z\phi(z)\,dz = -\int_{-\infty}^{0} z\phi(z)\,dz$

Standard Normal Distribution

- $Z \sim \text{Normal}(0, 1)$ if Z has PDF

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \mathbb{1}(-\infty < z < \infty)$$

- What is $\mathbb{V}\text{ar}[Z]$?
  - Hint: $\mathbb{V}\text{ar}[Z] = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$, let $u = z, dv = ze^{-\frac{1}{2}z^2} dz$

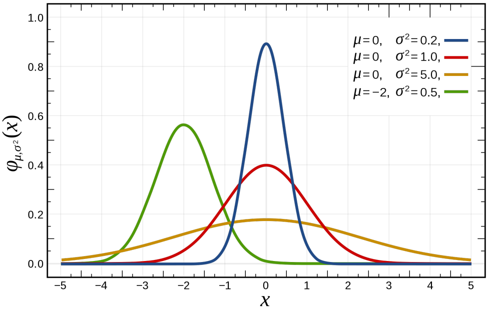
Normal Distribution

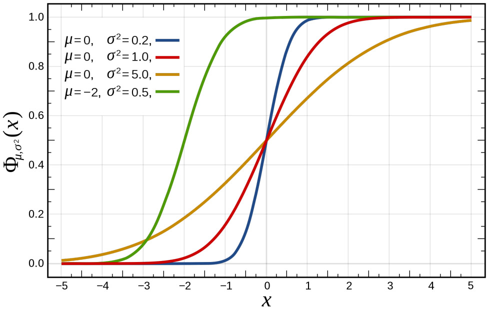
- Show that if $Z \sim N(0,1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$
  - This means that you can "standardize" any normal random variable by $\frac{x-\mu}{\sigma}$
  - Consider $\mathbb{P}[X \leq x]$

Normal Distribution

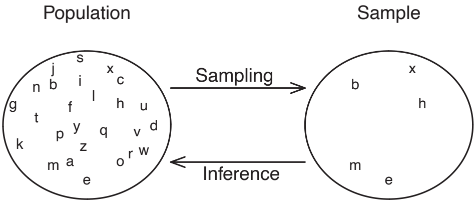
- Derive the pdf of $N(\mu, \sigma^2)$
  - Take derivative of CDF: $F_Z(\frac{x-\mu}{\sigma})$

Normal PDF and CDF





Random Sample



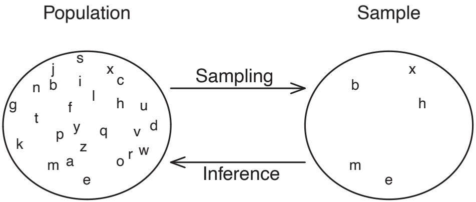
- Statistics/Inference: Learning properties of the population from samples
- Sample/Data: A collection of random variables from a population

$$\{X_i\}_{i=1}^{n} = \{X_1, \dots, X_n\}$$

- **Random Sample**: A sample that is independent and identically distributed (i.i.d.), i.e, they are **mutually independent** with **identical marginal distributions** F_X

$$X_1, \dots, X_n \sim \text{ i.i.d. } F_X$$

Statistic and Estimation



- **Parameter** θ: A measured quantity of the population F_X e.g. p, $\mathbb{E}[X]$, $\mathbb{V}\text{ar}[X]$
- **Statistic**: Any function of the sample $\{X_1, \ldots, X_n\}$ e.g. $\frac{1}{n}\sum_{i=1}^{n} X_i$
- **Sampling Distribution**: The distribution of a statistic
- **Estimator** for a parameter θ: A statistic intended as a guess about θ

$$\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$$

- **Estimate**: Realized value of the estimator **on a specific sample**

$$\hat{\theta}(x_1, \ldots, x_n) = \hat{\theta}(X_1 = x_1, \ldots, X_n = x_n)$$

Some Possible Statistics

- The sample mean is a statistic:

$$\overline{X}_n = \overline{X}_n(X_1, \dots, X_n) = \frac{1}{n}(X_1 + \cdots + X_n)$$

- Another possible, but quite naive, statistic can be:

$$\widehat{X}_1 = \widehat{X}_1(X_1, \dots, X_n) = X_1$$

- You can define whatever statistic you want, but some are better than others
  - Note that **any statistic** is also a **random variable** with its own distribution
  - The distribution of a statistic is called its **sampling distribution**
- Suppose we're interested in estimating parameter from the population F_X

$$\mu = \mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx$$

Sample Mean and Bias

- If $X_1, \ldots, X_n \sim$ i.i.d. F_X and $\mathbb{E}[X] = \mu$, then
$$\mathbb{E}\left[\overline{X}_n\right] = \frac{1}{n}(\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]) = \frac{1}{n}(\mu + \cdots + \mu) = \mu$$
- The **bias** of an estimator $\hat{\theta}$ is defined as the difference between the expected value of the estimator and the true value of the parameter
$$\text{Bias}[\hat{\theta}, \theta] = \mathbb{E}[\hat{\theta}] - \theta$$
- If drawn from a random sample, the sample mean $\overline{X}_n$ is an **unbiased estimator for population mean** μ because
$$\text{Bias}[\overline{X}_n, \mu] = \mathbb{E}[\overline{X}_n] - \mu = \mu - \mu = 0$$

Weak Law of Large Numbers (WLLN)

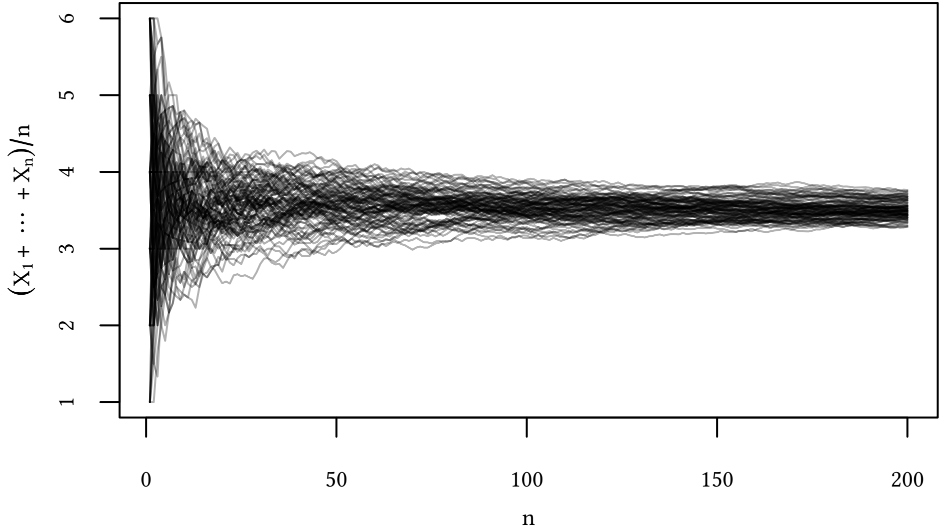
Weak Law of Large Numbers

Let $X_1, \dots, X_n \sim$ i.i.d. F_X and $\mathbb{V}\text{ar}[X] < \infty$, then for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left[\underbrace{\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}[X] \right|}_{\text{Distance between } \overline{X}_n \text{ and } \mathbb{E}[X]} \geq \varepsilon \right] = 0 \qquad \left(\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{p} \mathbb{E}[X] \right)$$

- As N gets large, the sample mean becomes increasingly likely to approximate $\mathbb{E}[X]$ to any arbitrary degree of precision
- This ensures the **consistency** of sample mean: $\frac{1}{n} \sum_{i=1}^{n} X_i \to \mathbb{E}[X]$ as $n \to \infty$

Weak Law of Large Numbers (WLLN)



The Variance of Sample Mean

- $X_1, \ldots, X_n \sim$ i.i.d. F_X, $\mathbb{E}[X] = \mu$, $\mathbb{V}\mathrm{ar}[X] = \sigma^2$
- Show that $\mathbb{V}\mathrm{ar}\left[\overline{X}_n\right] = \frac{\sigma^2}{n}$
- Show that $\mathbb{E}\left[\left(\overline{X}_n\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] = \frac{\sigma^2}{n} + \mu^2$

Sample Variance

- The k-th moment of X is $\mathbb{E}[X^k]$
- **Plug-in Principle**: $\frac{1}{n}\sum_{i=1}^{n} X_i^k$ is often a good estimator for $\mathbb{E}[X^k]$
- Show that $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] = \sigma^2 + \mu^2$

- In principle, we can use $\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mu^2$ to estimate σ^2, but often μ is **unknown**

Sample Variance

- What about

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

  - $\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i^2] = \sigma^2 + \mu^2$
  - $\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] = \frac{\sigma^2}{n} + \mu^2$
  - So we have that $\mathbb{E}\left[\hat{\sigma}^2\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\frac{1}{n}\sum_{i=1}^{n} X_i)^2\right] = \frac{n-1}{n}\sigma^2$
- We define sample variance s^2 such that $\mathbb{E}[s^2] = \sigma^2$

$$s^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

Common Statistics and Sampling Distributions

- Let $X_1, \dots, X_n \sim$ i.i.d. $N(\mu, \sigma^2)$, and let

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2$$

- We can show that $\overline{X}_n$ and s^2 are independent, and $\overline{X}_n \sim N(\mu, \sigma^2/n)$
- We can define the t-statistic

$$t = \frac{\overline{X}_n - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

  - Studentized sample mean follows **t-distribution** with $n-1$ degrees of freedom
  - Used for test of mean of a population or two populations

Common Statistics and Sampling Distributions

- We can compare the sample variance to the population variance

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2$$

$$\frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}_n}{\sigma} \right)^2 \sim \chi_{n-1}^2$$

  - The sample variance divided by population variance follows **Chi-squared distribution**
  - Used for test of goodness-of-fit with respect to a population

Common Statistics and Sampling Distributions

- We can compare the variability of two populations
- Let $X_1, \ldots, X_n \sim$ i.i.d. $N(\mu, \sigma^2)$, and $Y_1, \ldots, Y_m \sim$ i.i.d. $N(\mu_Y, \sigma_Y^2)$, then

$$F = \frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} = \frac{\chi_{n-1}^2/(n-1)}{\chi_{m-1}^2/(m-1)} \sim F_{n-1,m-1}$$

  - The ratio of two sample variance divided by population variance follows **F-distribution**
  - Used for comparing the variability of two populations

Central Limit Theorem (CLT)

Central Limit Theorem

Let $X_1, \ldots, X_n \sim$ i.i.d. F_X, $\mathbb{E}[X] = \mu$, and $\mathbb{V}\text{ar}[X] = \sigma^2 < \infty$

$$\lim_{n\to\infty} \underbrace{\mathbb{P}\left[\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \le x\right]}_{\text{CDF of standardized sample mean}} = \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}z^2}\, dz}_{\text{CDF of standard normal distribution}}$$

$$\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)\right)$$

- If n is large, the sampling distribution of the sample mean will tend to be approximately normal no matter how weirdly the population under consideration distributed
- It will allow us to rigorously quantify uncertainty when n is large

Central Limit Theorem (CLT)