Chapter 2

Recap - Probability Theory and Sampling

Shaikh Tanvir Hossain

October 25, 2022

East West University

Outline

1. **Random Experiment**

2. **Random Variables**

3. **Distribution of Random Variables**

4. **Population Summary Measures**
   - Expectation
   - Variance

5. **Statistic and Sampling Distributions**

- In the language of probability, the random sampling is an example of a *random experiment*, so what is a random experiment?
  - ✓ A random experiment is any process where we know all possible outcomes but we do not know which outcome will come before we perform the experiment. Examples of random experiment are
    - Tossing a coin, or multiple coins
    - Throwing a die, or multiple dice
    - Picking a random data point, or a random sample.
  - ✓ Any random experiment has a *set of possible outcomes*, this set is called the *sample space* of the experiment. It's a set. Usually we use the notation Ω to represent the sample space and ω to represent a specific outcome. For the sampling problem that we discussed in the last chapter, you can think the sample space is the population.

1. Random Experiment

2. **Random Variables**

3. Distribution of Random Variables

4. Population Summary Measures
   - Expectation
   - Variance

5. Statistic and Sampling Distributions

- From a random experiment we can define different *random variables*. A random variable is a function (we use the notation $X(\omega)$), where the input ω is coming from the *sample space* (which is the domain of the function) and the output is going to be in a *countable* set or in an *uncountable set* (this is the range of the function).
- When the range is a countable set we call it *discrete random variable*. For example the range could be $\{0, 1\}$ or $\{1, 2, 3\}$ or $\{1, 2, 3, 4, \ldots, \}$. The last one is an example of countably infinite set.
- When the range is an uncountable set we call it a *continuous random variable*, for example in this case the range could be $\mathbb{R}$ or any intervals like $(0, \infty)$ or $(1, 2]$ in $\mathbb{R}$ (there are different kinds of intervals, we talked about this in the class!)

◇ Follwing are two examples of random variable.

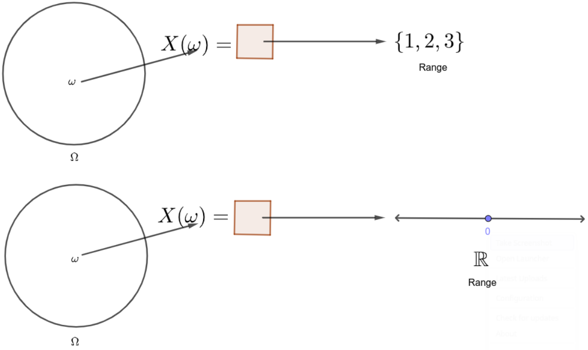

Figure 1: At the top we have a discrete random variable $X(\omega)$, the input is coming from the sample space and the range is a coutable set $\{1, 2, 3\}$. At the bottom we have a continuous random variable where the range is the whole real line $\mathbb{R}$.

- So a random variable is a function. What is the difference between the random variable $X(\omega)$ and the function that you have seen in your previous math courses, for example, $f(x) = 4x + 1$ (note for this function the domain is $\mathbb{R}$ and the range is also $\mathbb{R}$).
- The difference between the ordinary mathematical function $f(x)$ and the random variable $X(\omega)$ is that, for the random variable the input ω is *random object* (because this is a possible outcome of the experiment). So before performing the experiment we do not know which ω will appear, so this means for the function $X(\omega)$ we will not be able to say what value it will take..... or in other words we have the situation $X(?) =?$.

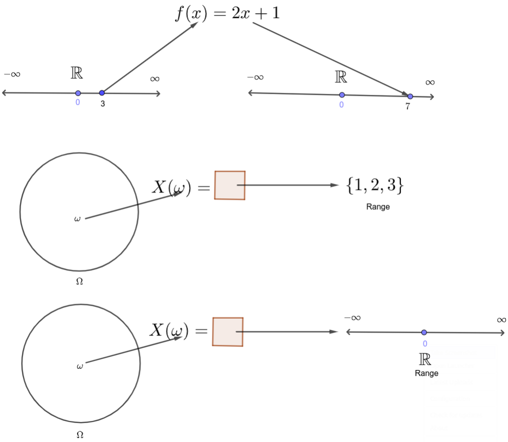

Figure 2: Comparison between the random variables and a mathematical function that you are familiar with. When a function $f(x)$ is not a random variable, there is no randomness in the input x. So you know the input x and you know the output $f(x)$. However for a random variable $X(\omega)$ the input ω is a random quantity (since this is an outcome of the experiment), hence the output is also unknown before we perform the experiment!

Usually we suppress (or hide) the input ω when we write the random variables, so we will write simply X. So when we have a random sample of 10, we actually have 10 random variables $X_1, X_2, X_3, \ldots, X_{10}$

	Income (realized or observed)	Random variable
1.	20	$X_1 = \boxed{?}$
2.	60	$X_2 = \boxed{?}$
3.	20	$X_3 = \boxed{?}$
4.	-20	$X_4 = \boxed{?}$
5.	-30	$X_5 = \boxed{?}$
6.	-10	$X_6 = \boxed{?}$
7.	80	$X_7 = \boxed{?}$
8.	10	$X_8 = \boxed{?}$
9.	30	$X_9 = \boxed{?}$
10.	40	$X_{10} = \boxed{?}$

Table 1: avg. income/year

◇ It is important to mention again that when we think about a random sample, or the random variables $X_1, X_2, X_3, \ldots, X_{10}$ we are NOT thinking about fixed numbers, we are thinking functions. You can think the functions here are like empty boxes, since any 10 values might appear.

- After we have the sample at hand, the sampling is done and the random variables already took it's value and we call it *realized* or *observed* values of the random variables and there is no randomness after we have observed the values.
- So you should always keep in mind about this dual nature of the data - random sample vs. observed values. As a student of Statistics you should always view the sample as a random object, not just some numbers.

- *An important notational remarks* -
  - ✓ Often we will write lowercase letters to denote the non-random objects or values. So rather than writing 20, 60, 20, $-20, \ldots,$ we will write $x_1, x_2, x_3, x_4, \ldots,$. So When we write this you should understand *these are fixed values* for a fixed sample or data.
  - ✓ But when we will write uppercase letters $X_1, X_2, X_3, X_4, \ldots,$ you should understand these are random variables.

	Income (realized or observed)	Random variable
1.	x_1	$X_1 = \boxed{?}$
2.	x_2	$X_2 = \boxed{?}$
3.	x_3	$X_3 = \boxed{?}$
4.	x_4	$X_4 = \boxed{?}$
5.	x_5	$X_5 = \boxed{?}$
6.	x_6	$X_6 = \boxed{?}$
7.	x_7	$X_7 = \boxed{?}$
8.	x_8	$X_8 = \boxed{?}$
9.	x_9	$X_9 = \boxed{?}$
10.	x_{10}	$X_{10} = \boxed{?}$

Table 2: avg. income/year

- Now since $X_1, X_2, X_3, X_4, \dots, X_n$ are random variables or random objects, any function of them is also random. So when we write

$$\bar{X}_n = \frac{X_1 + X_2 + X_3 + X_4 + \ldots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  it means $\bar{X}_n$ is a function of variables, so it is also random. Or in other words $\bar{X}_n$ is also a random variable. Note that this random variable depends on the sample size n.
- But if we write

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  this means it is for a fixed sample and this will be a constant.
- $\bar{X}_n$ is a very important random variable in Statistics, and we will come back to this later.

1. **Random Experiment**
2. **Random Variables**
3. **Distribution of Random Variables**
4. **Population Summary Measures**
   - Expectation
   - Variance
5. **Statistic and Sampling Distributions**

- Once we understand what is a random variable, the next two topics to understand are
  - ✓ *distribution of a random variable*, and
  - ✓ *summary measures of the distribution*
- When we say we know the distribution of a random variable, roughly this means *we know all probabilities when the random variable takes different values* in it's range.
- Now let's see some ideas about how we can calculate these probabilities.

- **Distribution of a discrete random variable:** If we have a discrete random variable X where the range of possible values are $\{1, 2, 3\}$. Then knowing Probability distribution implies we know
  - ✓ $P(X = 1)$, $P(X = 2)$ and $P(X = 3)$
- How do we calculate these probabilities? One answer is using *probability mass function* (short PMF), notation $p(x)$. In this case

$$P(X = x) = p(x)$$

- So knowing PMF means we know all the probabilities.
- Let's see some examples.

Example 1.1 (*Discrete random variables, PMF and distribution - Bernoulli Distribution*)

Suppose someone told you that X is a discrete random variable which takes only two values 0 and 1. So the range of this random variable is $\{0, 1\}$ and magically you know the PMF of X is

$$p(x) = (0.7)^x \times (0.3)^{1-x}$$

- ◇ Again, note that it is a function of x, that is why we wrote $p(x)$. Here x is the value of this binary random variable X, and we can have $X = 0$ or $X = 1$. And $p(0) = 0.3$ and $p(1) = 0.7$. This means $P(X = 0) = p(0) = 0.3$ and $P(X = 1) = p(1) = 0.7$.
- ◇ Since $p(x)$ is a function, we can plot this, how will be the plot (see it later)?
- ◇ Note that we can also write.

$$p(x) = (\theta)^x \times (1 - \theta)^{1-x}$$

- ◇ Where we have $\theta = 0.7$, so now we have $p(1) = \theta$ and $p(0) = 1 - \theta$.
- ◇ What is θ? In Probability theory and Statistics we call it a *parameter*. What is a *parameter*?

- A parameter is an object which has one to one relationship with the distribution, this means changing the parameter will change the distribution. This gives us a way to think about how we could control the probabilities. Here it says we can do it with a single object θ. Sometimes there are more than one parameters for a distribution (for example for normal distribution, you wills see the examples later!).
- In notation we will write parameters inside the PMF function *like* $p(x;\theta)$. This means $p()$ is a function of x, but there is an external element which must be fixed for a fixed distribution. So changing θ means changing the PMF and this means changing the distribution.

$$p(x) = (0.7)^x \times (1-0.7)^{1-x} \text{ (here } \theta = 0.7)$$
$$p(x) = (0.5)^x \times (1-0.5)^{1-x} \text{ (here } \theta = 0.5)$$
$$p(x) = (0.3)^x \times (1-0.3)^{1-x} \text{ (here } \theta = 0.3)$$

- Can you calculate $P(X=0)$ and $P(X=1)$? Yes we can, $P(X=0) = p(0) = 1-\theta$ and $P(X=1) = p(1) = \theta$
- Let's see the effect of changing θ in plots.

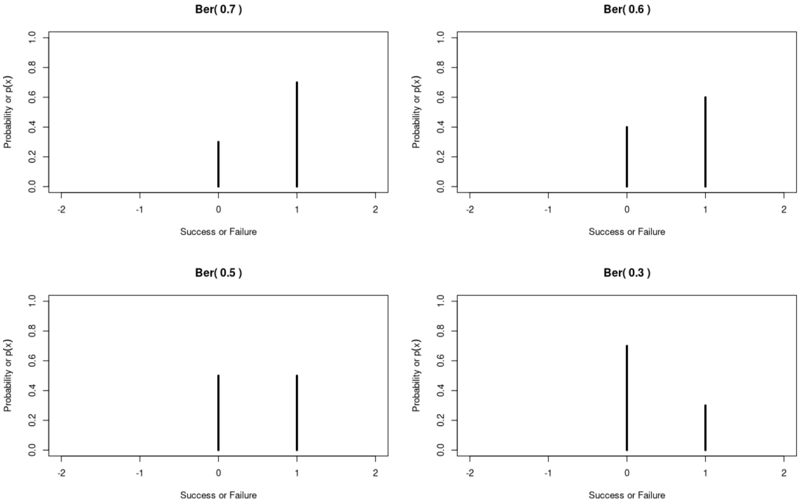

Figure 3: Changing θ means changing $p(1)$ and $p(0)$, this means changing $P(X = 1)$ or $P(X = 0)$, this means changing the distribution. So the parameter θ controls the distribution (These plots have been created using R)

- The last example is an example of a random variable called *Bernoulli random variable*, and the distribution of a Bernoulli random variable is called *Bernoulli distribution*.
- So the PMF of Bernoulli distribution is

$$p(x;\theta) = (\theta)^x \times (1-\theta)^{1-x} \text{ (for } x \in \{0,1\})$$

- If a random variable X follows Bernoulli distribution with parameter θ, then we will write $X \sim \text{Bernoulli}(\theta)$. This means $P(X=1)=\theta$ and $P(X=0)=1-\theta$

- **For Continuous random variables** If we have a continuous random variable then we can use *probability density function* (short PDF), notation $f(x)$, to calculate the probabilities. In this case if a and b are two real numbers then we have

$$P(X \in [a,b]) = \int_a^b f(x)dx$$

- So this means knowing this function means we can calculate all probabilities in these kinds of intervals. Here we can have different kinds of intervals, e.g., open, closed (we talked about this in the class)
- Note that if the random variable X is continuous then $P(X \in [a,b]) = P(X \in (a,b)) = P(X \in [a,b)) = P(X \in (a,b])$, so there is no difference whether we write closed interval or one sided closed interval or open interval.
- The reason is, if we have a continuous random variable, and x is a possible value, we will always have $P(X = x) = 0$
- Let's see a famous example of a continuous random variable.

Example 1.2 (Continuous random variables, PDF and distribution)

Suppose someone told you that X is a continuous random variable which takes values in the real line $\mathbb{R}$, and following is the density function of X

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

- Can we calculate $P(X \in [2, 3])$? Yes we can,

$$P(X \in [2, 3]) = \frac{1}{\sqrt{2\pi}} \int_2^3 f(x)dx = \int_2^3 e^{-\frac{1}{2}x^2} dx$$

- So in this case calculating probabilities is an *integration problem*, don't worry we won't do integration but you need to understand that calculating an integral within some interval means finding an area under the curve. Ques: What does the area under the curve within some interval mean here? Ans: This is the probability that X takes value in that interval.
- So again, always remember *Area = Probability* in the case of a continuous random variable.

- So for a continuous random variable knowing density function means we can calculate the probabilities and this then means we know the distribution.
- The density function in the last example is a very famous density function, and we call it the *density function of a standard normal distribution*
- Can we plot $f(x)$? Yes we can!

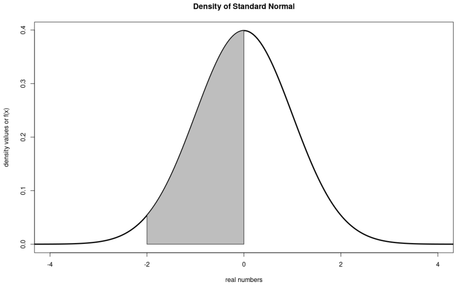

Figure 4: This is the density of the standard normal distribution. If X is a random variable which follows standard normal distribution, then the shaded are is means $P(X \in (-2, 0) = \int_{-2}^{0} f(x)dx = 0.4772499$. How did we calculate this probability? We can do this using Excel or R

- Recall for the Bernoulli distribution, there was a parameter. Is there any parameter here? The answer is both Yes and No. The idea is, the *density of a Normal distribution* can be written as

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Here we have two parameters μ and σ. When we wrote the standard Normal, we just plugged $\mu = 0$ and $\sigma = 1$ ($\sigma = 1$ means $\sigma^2 = 1$, there is a reason we will often use σ^2 rather than σ, this will be clear later!)
- Why we call them parameters? because changing them will change the shape of the density function, and this will change the distribution.
- Here changing μ will have the effect of changing location, this is why it is called the *location parameter* and changing σ will have the effect of changing dispersion (how dispersed the values are), this is why it is called the *dispersion parameter*.

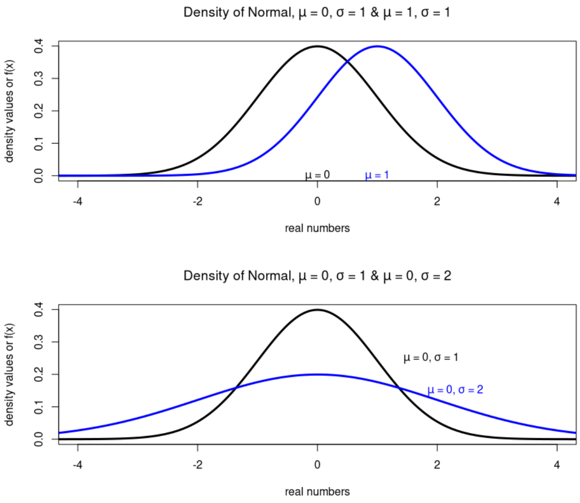

Figure 5: The top figure shows the effect of changing μ while keeping the σ same, note that this changes the location of the density. The bottom one shows the effect of changing σ while keeping the μ same. Clearly the calculated area will change and hence the probability!

- So, random variables can be discrete and continuous. And we saw the two most famous examples of a discrete (Bernoulli distribution) and a continuous (Normal distribution) random variable.
- Here again we write the PMF of the Bernoulli distribution, and PDF of the Normal distribution

$$p(x;\theta) = (\theta)^x \times (1-\theta)^{1-x} \text{ (for } x \in \{0,1\})$$
$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

- These are often often called *parametric distributions*, since the distribution depends on parameters.
  - ✓ If the random variable X follows Bernoulli distribution with parameter θ, we will write $X \sim \text{Bernoulli}(\theta)$
  - ✓ If the random variable X follows Normal distribution with parameter μ and σ^2, we will write $X \sim \mathcal{N}(\mu, \sigma^2)$

- Note that any random variable will have a distribution.
- Some examples of discrete random variables are *Bernoulli, Binomial, Poisson, Geometric, Negative Binomial and Hypergeometric*.
- Some examples of continuous random variables are *Normal, Uniform, Exponential, Gamma and Beta*.
- Some important continuous distributions are also derived using Normal distributions, these are *t, Chi-square* and *F* distributions.
- Do these distributions play any role in real life scenarios? The answer is, Yes, we can model different real life scenario using different theoretical parametric distributions. Since these are all parametric distributions, we call this approach *parametric modeling*!

- For example,
  - ✓ if we have a data of income, maybe we can use Normal distributions. So if we have random sample of size n, then we assume

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$$

    Here "iid" means all random variables in the random sample are *identically and independently distributed*, and $\overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ means all random variables are independent and they all follow Normally distributed with parameters μ and σ^2 (this means the distributions are identical).

  - ✓ If we have a data of gender, maybe we can use Bernoulli distributions. This means

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(\theta)$$

    Here $\overset{iid}{\sim}$ Bernoulli(θ) means all random variables are independent and they all follow Bernoulli distributed with parameter θ (this means the distributions are identical).
- Other than parameters is there any other interpretation of μ and σ^2? Or θ? More on this later...
- Can we follow a non-parametric approach? Yes we can (more on this later!)

Population Summary Measures

Expectation

- There are many possible summary measures of the distribution. The idea of the summary measures are - with one/two numbers we get a rough idea about the distribution.
- Two most common ones are *Expectation* and *Variance*. If we have a random variable X, the Expectation of the random variable X is defined as follows,
  - ✓ When X is discrete, $p(\cdot)$ is the PMF and the range of X is $\{x_1, x_2, x_3, \ldots, x_k\}$, then the Expectation of X is

$$\mathbb{E}(X) = \sum_{i=1}^{k} x_i \, p(x_i),$$

  - ✓ When X is continuous and $f(\cdot)$ is the PMF and range of f X is $\mathbb{R}$, , then the Expectation of X is

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

- Notice the notation for the Expectation of the random variable X is $\mathbb{E}(X)$. You can think Expectation is like a function operating on another function X. This is why Expectation is also called an *operator*.

- Now let's see how can we calculate *Expectation* if we know the distribution,
- If X is a (discrete) random variable taking values 10, 20, and 30 and we know the PMF $p(10) = 1/4, p(20) = 1/4, p(30) = 2/4$, then

$$\begin{aligned}\mathbb{E}(X) &= (10 \times p(10)) + (20 \times p(20)) + (30 \times p(30)) \\ &= (10 \times 1/4) + (20 \times 1/4) + (30 \times 2/4) \\ &= 90/4 = 22.5\end{aligned}$$

- Is it possible to calculate the Expectation of a Bernoulli or a Normal Random Variables that you saw? Yes we can.
  - ✓ If $X \sim$ Bernoulli(θ), we have $\mathbb{E}(X) = \theta$ (try this! you should be able to do this!)
  - ✓ If $X \sim \mathcal{N}(\mu, \sigma^2)$, we have $\mathbb{E}(X) = \mu$ (this needs some integration skills, but you can skip this calculation!)

- *Notational Remark:* Often the expectation of a random variable is denoted with μ even though it is NOT normally distributed. This is just a notational abuse. So if we explicitly mention $X \sim \mathcal{N}(\mu, \sigma^2)$, then you know $\mathbb{E}(X) = \mu$, where μ is the location parameter. But if we do not specify the distribution of X and write $\mathbb{E}(X) = \mu$. then you should understand X follows some unknown distribution with Expectation μ. At the end Expectation is going to be a fixed value.

Theorem 1.1 (Rules for Expectation)

✓ 1. The expected value of a constant is the constant itself. Thus, if b is a constant, $\mathbb{E}(b) = b$.

✓ 2. If a and b are constants,

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

This can be generalized. If $X_1, X_2, \ldots, X_n$ are n random variables and $a_1, a_2, \ldots, a_n$ and b are constants, then

$$\mathbb{E}\left(a_1 X_1 + \cdots + a_n X_n + b\right) = a_1 \mathbb{E}\left(X_1\right) + \cdots + a_n \mathbb{E}\left(X_n\right) + b$$

✓ 3. If X and Y are *independent random variables*, then

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

That is, the *expectation of the product XY is the product of the (individual) expectations of X and Y.*

✓ 4. If X is a random variable with PDF $f(x)$ and if $g(X)$ is any function of X, then

$$E[g(X)] = \sum_x g(x)\, p(x) \quad (X \text{ discrete }), \quad E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx \quad (X \text{ continuous })$$

- How do you interpret Expectation? Ans: Expectation is similar to the idea of a *weighted average* (we talked about this in the class).
- Also note Expectation is a Population object, (there is no sampling going on here), to see this observe that *to calculate expectation we need to know the true population probabilities*, or true distribution.
- Take the example in page 30. For the random variable X, when we wrote

$$p(10) = P(X = 10) = 1/4$$
$$p(20) = P(X = 20) = 1/4$$
$$p(30) = P(X = 30) = 2/4$$

  You can think this means 25% of the population have income 10 thousands, 25% of the population have income 20 thousands and remaining 50% have 30 thousands. These are all 100% population. So here X means income *.
- Now when we calculate *Expectation*, we get 22.5, then this is the weighted average of the 3 different income levels. So this is a *Population Mean*

---

*Note that probabilities have to sum to 1 because at the end we need 100% population, this is why PMF or PDF values will always be sum to / integrated to 1

- Do we know Population mean in real-life scenarios? Answer is - most of the times NO!
- In general we use μ to denote this unknown population mean.
- Lastly we would like to point out that, expectation is often called a measure of *central tendency*, it gives information about the center of the *population*. The idea is very similar to *sample average* but be careful this is not for a sample, there is no sampling going on here.

Population Summary Measures

Variance

- ✓ Variance of a random variable X, denoted by $\text{Var}(X)$ is defined as follows,

$$\text{Var}(X) = \mathbb{E}\left(\left(X - \mathbb{E}\left(X\right)\right)^2\right),$$

- ✓ This means Variance is also an Expectation. In fact we have applied 4. of Theorem 1.1, where $g(X) = \left(X - \mathbb{E}\left(X\right)\right)^2$ (can you see this?)

- ✓ Regardless of the distribution if we denote $\mathbb{E}(X) = \mu$, then applying Expectation as an operator we get,

$$\text{Var}(X) = \sum_{i=1}^{k} (x_i - \mu)^2 p(x_i) \text{ when } X \text{ is discrete}$$

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x), \text{ when } X \text{ is continuous}$$

◇ Like Expectation, *Variance* is also a summary measure of the Population. It gives us a measure of the spread in the population. This is why we call it Population Variance.

◇ Now let's see how can we calculate *Variance* if we know the distribution, this example is a continuation of the example in page 30, we already calculated $\mathbb{E}(X) = 22.5$, so here $\mu = 22.5$

$$\begin{aligned}\mathrm{Var}(X) &= \sum_{i=1}^{3}(x_i - \mu)^2 p(x_i) \\ &= \Big((10 - 22.5)^2 \times p(10)\Big) + \Big((20 - 22.5)^2 \times p(20)\Big) + \Big((30 - 22.5)^2 \times p(30)\Big) \\ &= (156.25 \times 1/4) + (6.25 \times 1/4) + (56.25 \times 2/4) \\ &= 68.75\end{aligned}$$

◇ Is it possible to calculate the Variance of a Bernoulli or a Normal Random Variable that you saw? Yes we can.

✓ If $X \sim \text{Bernoulli}(\theta)$, we have $\mathrm{Var}(X) = \theta \times (1 - \theta)$ (try this! you should be able to do this!)

✓ If $X \sim \mathcal{N}(\mu, \sigma^2)$, we have $\mathrm{Var}(X) = \sigma^2$ (this needs some integration skills, but you can skip this calculation!)

- Like Population Mean, we usually don't know Population Variance and regardless of the distribution often we use σ^2 to denote the Population variance of a random variable X.
- *Remarks on Notation regarding Variance and Standard Deviation:* Note that $\sigma = \sqrt{\sigma^2}$, which is called *Standard Deviation*. So *Standard Deviation* is just the square root of the Variance.
- *Remarks on Notation regarding Normal Distribution:* As we mentioned if we write $X \sim \mathcal{N}(\mu, \sigma^2)$, this means X is Normally distributed with two parameters μ and σ^2. But now we can say this also means X is Normally distributed with Population Mean μ and Population Variance σ^2 (or Population Standard Deviation σ.)
- Note that parameters are not always same as mean and variance. Parameters are objects which will control the distribution. Summary measures are coming from some calculations. For Normal distributions they are same, but for other distributions they could be different. Although there is always going to be a connection between these two type of objects, but they can be different.
- There are some rules to calculate Variance, these are easy to apply (the proofs are also easy but I am skipping it!)

◇ Similarly there are some rules to calculate Variance.

Theorem 1.2 (Rules for Variance)

✓ 1. The variance of a constant b is zero, so $\mathrm{Var}(b) = 0$

✓ 2. If a and b are constants, then

$$\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$$

✓ 3. If X and Y are *independent* random variables, then

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$
$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

This can be generalized to more than two independent variables.

✓ 4. If X and Y are independent random variables, and a and b are constants, then

$$\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y)$$

✓ 5. $\mathbb{E}\left((X-\mu)^2\right) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2$, this is an alternative formula for Variance.

1. Random Experiment

2. Random Variables

3. Distribution of Random Variables

4. Population Summary Measures
   - Expectation
   - Variance

5. Statistic and Sampling Distributions

◇ Let's start our journey with *Statistical Inference*.

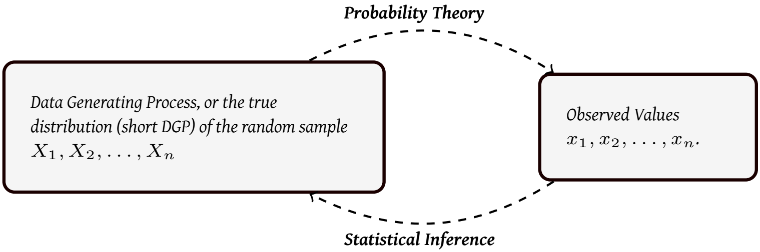

Figure 6: Interplay between Probability Theory and Statistical Inference

◇ Although *Probability theory* tells us about the Population distributions, different summary measures (e.g., Mean or Variance), in practical scenarios we cannot calculate these objects using a sample. So this is a theory which deals with Population.

◇ When we study different methods in *Statistical Inference / Inferential Statistics*, essentially we study how to make different conclusions about population objects using a data. This is what we call *inference*.

- Recall we mentioned that in Statistics a data is a random object, and we call it a *random sample*, typically we use the notation $X_1, X_2, \ldots, X_n$. So this means we have n random variables.
- Now these random variables could be dependent / independent. Also they can have same distributions / different distributions.
- Unless we specify explicitly, from now on we will always assume we have an *iid random sample of size n*, this means the random variables $X_1, X_2, \ldots, X_n$ have identical distributions (i.e., same PDF or PMF) and they are mutually independent.
- For example, it could be that

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$$

  means all random variables are *independent* and they all follow Normal distribution with same mean μ and same variance σ^2.
- It does not have to be always Normal, the random variables could follow different distributions, but the idea of the identical distributions is they have to follow the *same distribution*.

- The idea of *independence* is also important. If we say we have an independent random samples, this means when we are picking samples from a population we must have
  - ✓ 1) *Either*, a finite population and we are sampling with replacement. So every unit in the population have exactly same probability of being in the sample or not.
  - ✓ 2) *Or*, an infinite population, so when we pick one sample, it doesn't affect the probability of the next unit being in the sample or not.
- Make sure you understand the iid assumption clearly.
- A set of iid random variables can make our life really easy, and we will see some benefit of this assumption later (Ques: Is this always a good assumption? Can you think about a setup where it fails?)
- Almost all of the discussions in Inferential Statistics is centered around *Estimation* and *Testing*.
- To understand Estimation and Testing we need to understand two objects
  - ✓ Sample Statistic
  - ✓ Distribution of a Statistic

- Sample Statistic (in short often it is called *Statistic*) is simply a *function of the random sample*, so $f(X_1, X_2, \ldots, X_n)$. For example following are examples of Statistic
  - ✓ 1) Average of the random variables $\frac{1}{n}\sum_{i=1}^{n} X_i$.
  - ✓ 2) Variance of the random variables $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ (also $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$)
  - ✓ 3) Median of the random variables $\mathrm{Med}(X_1, X_2, \ldots, X_n)$
  - ✓ 4) Maximum of the random variables $\mathrm{Max}(X_1, X_2, \ldots, X_n)$
  - ✓ And there could be other functions, we will see more examples later
- For now, the two most important functions for us, are 1) and 2) (you already know them!)

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

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

- Recall what we have discussed in slide 12, $\bar{X}_n$ is a random variable since it is a function of n random variables, but when we fix a sample and calculate sample average $\bar{x}$ for a fixed data, then it's just a number.
- Similarly S^2 is also a random variable, and if we calculate sample variance s^2 for a fixed sample it is going to be a fixed number.
- Since $\bar{X}_n$ is a random variable, we may ask *how can we calculate the distribution of $\bar{X}_n$?*
- The answer to this question is not that straight forward. Essentially it depends on whether we know the distribution of $X_1, X_2, \ldots, X_n$. We will come back to the answer shortly, but first let's calculate

$$\mathbb{E}(\bar{X}_n) \text{ and } \operatorname{Var}(\bar{X}_n)$$

- It turn out if we only know the mean and variance of $X_1, X_2, \ldots, X_n$, regardless of the fact that whether we know distribution of $X_1, X_2, \ldots, X_n$, we can calculate $\mathbb{E}(\bar{X}_n)$ and $\operatorname{Var}(X_n)$
- How? Use iid assumption and Theorem 1.1 and Theorem 1.2

Theorem 1.1 Mean and Variance of $\bar{X}_n$ with an iid random sample

If we have an iid random sample $X_1, X_2, \ldots, X_n$ with $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ for $i = 1, 2, 3, \ldots, n$ (this means all random variables have same mean μ and same variance σ^2), then,

$$\mathbb{E}(\bar{X}_n) = \mu$$

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

- The proof is very easy, we need to use iid assumption and Theorem 1.1. and 1.2
- So just having the iid assumption, we can calculate the mean and the variance of $\bar{X}_n$, even though we don't know the distribution of $\bar{X}_n$.
- Now let's answer a deeper question, *what is the distribution of $\bar{X}_n$?* As we said it depends on whether we know the distribution of $X_1, X_2, X_3, \ldots, X_n$.
- In general the distribution of a statistic is called *sampling distribution*. So the distribution of $\bar{X}_n$ is also called *sampling distribution*. Another terminology - The standard deviation of a statistic is called *Standard Error*.
- Following theorem is useful to calculate the sampling distribution of $\bar{X}_n$.

Theorem 1.2 (Exact distribution of $\bar{X}_n$ with Normality Assumption)

If $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ then $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$.

- The proof of Theorem 1.2 is a little bit technical so we will skip it, but the idea is pretty simple, it says - *if all the random variables in the random sample are independent and all are Normally distributed with mean μ and variance σ^2, then the sampling distribution of $\bar{X}_n$ is also Normally distributed with mean μ and variance σ^2/n.*
- Now, let's define a new random variable Z with following transformation,

$$Z = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)$$

- The last step is called *Standardization*, why this name? Here is the answer - it is possible to show that

$$Z \sim \mathcal{N}(0, 1), \text{ note that this means } \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$$

- Note that the distribution depends on the sample size n, this is why it is called the *exact distribution* of X_n. We can calculate this distribution for any finite n.

- We will see that this distribution plays a very important role in both estimation and testing.
- Again note that using $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ is same as $\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$
- Suppose we want to us this distribution (later we will see why), then it is clear that we need to know μ and σ^2, otherwise we don't know the distribution.
- In a real life setting "maybe" we can *assume* all of the random variables $X_1, \ldots, X_n$ are normally distributed, but it's not realistic that we assume we know both mean μ and variance σ^2.
- Let's forget about μ for a moment. Suppose we replace σ with S, where $S = \sqrt{S^2}$ and recall

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

- With this we actually get a new random variable. Under the same conditions of Theorem 1.2 we can show that

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

- Here $t_{(n-1)}$ means t distribution with parameter $(n-1)$, this parameter has a special name, it's called *degrees of freedom*!

- The benefit of using t distribution is, we don't need the knowledge of σ, we can just calculate S using the data and we can use t distribution.
- So far we have used Exact distributions, either Normal or t distributions.
- There are at least two problems with exact distributions,

  - ✓ We need to know the distributions of $X_1, X_2, \ldots, X_n$. Recall, we used the idea that they all follow Normal distribution with mean μ and variance σ^2. Even if we forget about σ and use S, then we get $t_{(n-1)}$ distribution but still the *Normality assumption is there*. So bottomline to use the sampling distribution of $\bar{X}_n$ as Normal or t, the random variables $X_1, X_2, \ldots, X_n$ must follow Normal distribution.

  - ✓ Even if we know the distributions of $X_1, X_2, \ldots, X_n$, it's not always simple to figure out the distribution of $\bar{X}_n$.

- So bottom line - exact distributions can be problematic, they depend on strong assumption that we know the distributions of $X_1, X_2, \ldots, X_n$.
- Is this the only way to get the sampling distributions? The answer is NO - there are two other ways
  - ✓ Assume we have a large sample (so we will assume $n \to \infty$) and use *asymptotic distributions* - this involves using Central Limit Theorems (CLTs) and other large sample techniques.
  - ✓ Resampling techniques (Jackknife or Bootstrap)
- We will now see the idea of using *asymptotic distribution*.
- Although resampling techniques are also very interesting or useful, but we will not cover this topic in this course (sorry :( )

Theorem 1.3 (Central Limit Theorem)

Let $X_1, X_2, \ldots$ be a sequence of *iid random variables* with $\mathbb{E}X_i = \mu$ and $\text{Var}X_i = \sigma^2 > 0$. (Assume both μ and σ^2 are finite). Define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$, then $\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)$ has a *limiting Standard Normal Distribution*. This is sometimes written as

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$$

- Note that we use the idea $n \to \infty$, is what we call *"asymptotic"* idea, and $\xrightarrow[n\to\infty]{d}$ means if $n \to \infty$ then the distribution is convering to". Also not that this is an abstract idea, in reality there is nothing called infinite n.
- Also, note that

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- So thie means the standardized version of X_n is asymptotically normally distributed.
- So CLT says - *the distribution of the standardized version of the sample mean will asymptotically follow Standard Normal distribution*

What are the benefits of asymptotic distribution?

- Note that we didn't use any assumptions on the distributions of the random variables.
- Even if we don't know σ and use S, then asymptoically we still get Normal distribution

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{S}\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1)$$

- So this means in practice when our sample size is large then we can always use normal distribution as a sampling distribution.