Ch3 - Probability Theory - 2

(Random Variables and Probability Distributions)

Statistics For Business and Economics - I

Shaikh Tanvir Hossain

East West University, Dhaka

Last Updated May 12, 2025

Outline

Outline

1. Random Variables

- Definitions, Discrete and Continuous RVs, and Prob. Distributions

2. Discrete Random Variables

- 1. Probability Distribution and the idea of PMF
- 2. CDF, quantiles and percentiles
- 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$
- 4. Rules of Expectation and Variance

3. Parametric Distributions : Discrete

- 1. Discrete Uniform
- 2. Bernoulli and Binomial Distribution

4. Continuous Random Variables

- 1. Probability Distribution and the idea of PDF
- 2. CDF, quantiles and percentiles
- 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$

5. Parametric Distributions : Continuous

- 1. Uniform Distribution
- 2. Normal Distribution

- In this chapter we start with the second part of the Probability Theory where we will start talking about *random variables* and *probability distributions*. Undoubtedly these two concepts are really the core part of Probability Theory and Statistics. In this chapter we will cover univariate random variables and some univariate probability distributions. These are theoretical distributions which are useful to *model* real life scenarios for one variable only. The next chapter will be about multivariate random variables and joint probability distributions (along with conditional distributions) which is more like an extension of these ideas to multivariate setting.
- So let's start... 🚶 🚶 🚶 .

1. Random Variables

- Definitions, Discrete and Continuous RVs, and Prob. Distributions

2. Discrete Random Variables

- 1. Probability Distribution and the idea of PMF
- 2. CDF, quantiles and percentiles
- 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$
- 4. Rules of Expectation and Variance

3. Parametric Distributions : Discrete

- 1. Discrete Uniform
- 2. Bernoulli and Binomial Distribution

4. Continuous Random Variables

- 1. Probability Distribution and the idea of PDF
- 2. CDF, quantiles and percentiles
- 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$

5. Parametric Distributions : Continuous

- 1. Uniform Distribution
- 2. Normal Distribution

Random Variables

Random Variables

Definitions, Discrete and Continuous RVs, and Prob. Distributions

- What is a random variable? Sometimes we are not interested in the experimental outcomes directly, rather we are interested in some *kind of numerical representation of the experimental outcomes*, the idea of the random variable is essentially this....Here is a rough definition,

Definition 3.1: (Random Variables, discrete and continuous)

A Random Variable is a numerical representation of the outcomes of any random experiment. We often use uppercase or capital letters, for example X, Y, or Z, to denote random variables.

Here are some examples.

- **Example 1: Experiment - Tossing a single coin**

  Sample Space: $\Omega = \{H, T\}$

  Random Variable: $X = 1$ for Head and $X = 0$ for Tail.

- **Example 2: Experiment - Tossing two coins**

  Sample Space: $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$

  Random Variable: Here it is possible to think about different random variables, for example we can think about *X is a random variable that counts the number of heads*, then in this case

$$\text{when we have outcome } (H, H), \text{ then } X = 2,$$
$$\text{when we have outcome } (H, T), \text{ then } X = 1,$$
$$\text{when we have outcome } (T, H), \text{ then } X = 1,$$
$$\text{when we have outcome } (T, T), \text{ then } X = 0.$$

  So here the random variable X takes values $0, 1, 2$.

  Note we can also think about different random variables in this setting, for example we can think about *total number of tails*, or *whether we have at least one head*, etc.

- **Example 3: Experiment - Tossing three coins**

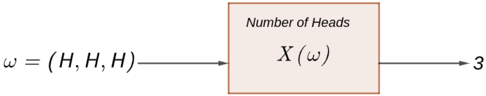
  Sample Space:
  $\Omega = \{(H,H,H),(H,H,T),(H,T,H),(T,H,H),(T,T,H),(T,H,T),(H,T,T),(T,T,T)\}$

  Random Variable: here again we can think *X is a random variable that counts the number of heads*, then in this case

$$\begin{aligned}
&\text{when we have outcome } (H,H,H), \text{ then } X = 3,\\
&\text{when we have outcome } (H,H,T), \text{ then } X = 2,\\
&\text{when we have outcome } (H,T,H), \text{ then } X = 2,\\
&\text{when we have outcome } (T,H,H), \text{ then } X = 2,\\
&\text{when we have outcome } (T,T,H), \text{ then } X = 1,\\
&\text{when we have outcome } (T,H,T), \text{ then } X = 1,\\
&\text{when we have outcome } (H,T,T), \text{ then } X = 1,\\
&\text{when we have outcome } (T,T,T), \text{ then } X = 0.
\end{aligned}$$

  So here the random variable X takes values $0, 1, 2, 3$.

- You get the idea of the random variable, actually mathematically a random variable is a function, that takes inputs from the sample space Ω and gives outputs to the real line $\mathbb{R}$. For example in the experiment where we are tossing three coins, we can think a function $X(\omega)$ where the input ω is coming from the sample space Ω, and the output will be *total number of heads*, which is going to be a real number in the set $\mathbb{R}$.



- Like the above picture for any input ω, you can get an output $X(\omega)$, which is going to be a number in $\mathbb{R}$. When we write the random variable X, we don't write the input ω.
- So again the random variable is actually a function from the sample space Ω to the real line $\mathbb{R}$. .....

Discrete and Continuous Random Variables:

Depending upon whether the set of values of a random variable is a *countable set* (it could be finite or infinite), or an *uncountable set (which is always infinite)* we can classify the random variables in two types / categories ...

- *Discrete Random Variable*: When the random variable takes values in a countable set for example $\{0, 1, 2, 3\}$, or even infinitely countable set like $\{0, 1, 2, 3, \ldots\}$, we call it a *discrete random variable*.
- *Continuous Random Variable*: When the random variable takes values in an uncountable or infinite set, for example $[0, 1]$ or $[0, 30]$ or even $(-\infty, \infty)$, we call it a *continuous random variable*.

- You might be wondering why we call this *"random"* variable? Any guess? This is because before performing the experiment, the *input of the function is random*. So the *output of the function is also random*. Following picture is useful,

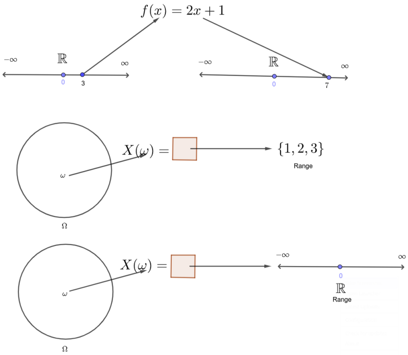


Figure 1: From the top, the first one is a mathematical function where the input is not random and the output is also not random (we often call this *deterministic function*). The second one is a *discrete random variable* where the output set is $\{1, 2, 3\}$. And the third one is a *continuous random variable* where the output set is whole $\mathbb{R}$. Note that for the random variables there is a blank box, this means before performing the experiment we don't know the output, since the input is random, so does the output...

- Let's see some real life examples of random variables. It's important that in many cases the random variable and the values are clear but the sample space is probably not clear. So when we start thinking about random variables, we don't actually think about the actual sample space, rather we think about the random variable and the values it can take.
- Following are some examples of discrete random variable (some examples are taken from Anderson et al. (2020)).

Random Experiment	*RV - X*	*Possible Values, x*
Toss a coin	1 - head, 0 - tail	$1, 0$
Roll a die	# dots in upper face	$1, 2, 3, 4, 5, 6$
Contact a single customers	1 - receives, 0 - ignores	$1, 0$
Contact 5 customers	# customers who receives	$0, 1, 2, 4, 5$
Operate a hospital for a day	# patients who arrive	$0, 1, 2, 3, \ldots$
Offer a customer the choice of products	product chosen by customer	0 - None, 1 - A, 2 - B
Randomly pick 10 EWU students	their occupation status	1 - job, 0 - no job

Table 1: Examples of Discrete Random Variables

Random Experiment	*RV - X*	*Possible Values, x*
Customer visits a web page	time customer spends (in min)	$[0, \infty)$
Taking a bus to Uni	time you have to wait (in min)	$[0, \infty)$
Randomly pick 10 EWU students	each of their heights (in cm)	$[120, 210]$
A flight from Dhaka to Chittagong	time need to travel (in min)	$[60, 90]$

Table 2: Examples of Continuous Random Variables

- Our next question is *how do we calculate probabilities for these random variables*, for example, if the *random experiment is tossing two coins* and X is a random variable that counts *the number of heads*, then we know X can take three values $0, 1$ and 2. Question is how do we calculate $\mathbb{P}(X = 0)$ or $\mathbb{P}(X = 1)$ or $\mathbb{P}(X = 2)$?
- In this case, we can actually go back to the events associated with it and then calculate the probability of that event using classical definition, for example

$$\begin{aligned} X = 0 &\text{ is associated with the event } \{(T, T)\} \\ X = 1 &\text{ is associated with the event } \{(H, T), (T, H)\} \\ X = 2 &\text{ is associated with the event } \{(H, H)\} \end{aligned}$$

- Then using the classical definition we can calculate the probabilities, for example

$$\mathbb{P}\left(\{(H, T), (T, H)\}\right) = \frac{2}{4} = \frac{1}{2}$$

- Doing similarly we get

Probabilities of Events:	$\mathbb{P}\left(\{(T,T)\}\right)$	$\mathbb{P}\left(\{(H,T),(T,H)\}\right)$	$\mathbb{P}\left(\{(H,H)\}\right)$
Probabilities:	$1/4$	$1/2$	$1/4$

Now we can map the same probabilities with the random variables since they are connected with the events, so this gives us,

Probabilities of the Values of X :	$\mathbb{P}(X=0)$	$\mathbb{P}(X=1)$	$\mathbb{P}(X=2)$
Probabilities :	$1/4$	$1/2$	$1/4$

- Here $\mathbb{P}(X=x)$ means, *probability of X taking values x*, where x can be 0, 1, and 2. Note that we will write the random variables with uppercase letters and the values with lowercase letters.

- Here is another example, when we have tossed three coins, here the sample space is $\Omega = \{(H,H,H),(H,H,T),(H,T,H),(T,H,H),(T,T,H),(T,H,T),(H,T,T),(T,T,T)\}$
  Now we can think about a random variable X that counts the number of heads, then in this case

$$
\begin{aligned}
X &= 0 \text{ is associated with the event } \{(T,T,T)\}\\
X &= 1 \text{ is associated with the event } \{(H,T,T),(T,H,T),(T,T,H)\}\\
X &= 2 \text{ is associated with the event } \{(H,H,T),(H,T,H),(T,H,H)\}\\
X &= 3 \text{ is associated with the event } \{(H,H,H)\}
\end{aligned}
$$

- With the same idea in this case we can calculate,

Probabilities of Events:	$\mathbb{P}(\{(T,T,T)\})$	$\mathbb{P}(\{(H,T,T),(T,H,T),(T,T,H)\})$	$\mathbb{P}(\{(H,H,T),(H,T,H),(T,H,H)\})$	$\mathbb{P}(\{(H,H,H)\})$
Probabilities:	1/8	3/8	3/8	1/8

- Now we can map the same probabilities with the random variables since they are connected with the events, for example

Probabilities of the Values of X :	$\mathbb{P}(X=0)$	$\mathbb{P}(X=1)$	$\mathbb{P}(X=2)$	$\mathbb{P}(X=3)$
Probabilities:	1/8	3/8	3/8	1/8

- Here again $\mathbb{P}(X=x)$ means, probability of X taking values x, where x can be 0, 1, 2 and 3.

- Interestingly with this information we can also calculate probabilities when our random variable X takes values in different kinds of intervals (recall the open and closed intervals we saw before, like $[2, 3]$ and $(2, 3]$, etc)
- For example we can calculate

$$\mathbb{P}(X \in [1, 1.5))$$
$$\mathbb{P}(X \in [4, 10000))$$
$$\mathbb{P}(X \in [1, 2))$$

- The idea is pretty simple, we will *use the probabilities where X actually takes its values*, for example, if we have

	$\mathbb{P}(X = 0)$	$\mathbb{P}(X = 1)$	$\mathbb{P}(X = 2)$
Probabilities	$1/4$	$1/2$	$1/4$

then we can calculate

$$\mathbb{P}(X \in [1,2]) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = 1/2 + 1/4 = 3/4$$
$$\mathbb{P}(X \in [2,3]) = \mathbb{P}(X = 2) = 1/4$$
$$\mathbb{P}(X \in [0,2]) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = 1/4 + 1/2 + 1/4 = 1$$

- You can try to calculate ...

$$\mathbb{P}(X \in [1, 1.5))$$
$$\mathbb{P}(X \in [4, 10000))$$
$$\mathbb{P}(X \in [1, 2))$$

- for the probabilities we wrote above, we can write them slightly differently, for example

$\mathbb{P}(X \in [1, 1.5))$ or we write $\mathbb{P}(1 \leq X < 1.5)$
$\mathbb{P}(X \in [4, 10000))$ or we write $\mathbb{P}(4 \leq X < 10000)$
$\mathbb{P}(X \in [1, 2))$ or we write $\mathbb{P}(1 \leq X < 2)$

- Finally always remember since $\mathbb{R}$ includes everything or all the possible values of all kinds of random variables, we will always have $\mathbb{P}(\mathbb{R}) = 1$ or $\mathbb{P}(X \in (-\infty, \infty)) = 1$
- *CAVEAT ☕:* It turns out that it is not easy to calculate probabilities in $\mathbb{R}$, and there might be some issues, there are some bad sets / intervals where we cannot calculate probabilities with a consistent way. To explain it fully we need to talk about measurability issues, which is beyond the scope of this course, so we will simply assume that it is possible to calculate probabilities in the interval of $\mathbb{R}$, and you can ignore this comment if you want!

- **Probability Distributions of a Random Variable**
- The last topic for this part is talking about *probability distribution of a random variable*. Here is an informal definition,

Probability Distribution of a Random Variable

If X is a random variable then the probability distribution of X is the collection of *all probabilities* of different possible intervals in $\mathbb{R}$, for example $\mathbb{P}(X \in [2,3])$ or $\mathbb{P}(X \in [1,5])$, etc. If X is a *discrete random variable* then the probability distribution means *how the probabilities are distributed on the values of X*, since these probabilities are enough to calculate the probabilities of different intervals in $\mathbb{R}$ for X.

- So when X is a continuous random variable the idea of probability distribution is a bit different, we will see that later, but when X is a discrete random variable, the probability distribution means all the probabilities of the form $\mathbb{P}(X = x)$ for all possible values x that X can take. For example, here is a probability distribution a random variable that we already saw

	$\mathbb{P}(X=0)$	$\mathbb{P}(X=1)$	$\mathbb{P}(X=2)$
Probabilities	$1/4$	$1/2$	$1/4$

- Do you think we always have to find probabilities by going back to original sample space? The answer is NO. There are actually two nice functions on the real line $\mathbb{R}$, which helps to calculate probabilities when the random X takes values in different kinds of intervals.
- For discrete random variable this function is known as *probability mass function* or in short *PMF*. The idea of this function is same as the probability distribution we just talked about, but here we will have a function that gives us the probabilities of the form $\mathbb{P}(X = x)$ for all possible values x that X can take.
- and for continuous random variable this function is known as *probability density function* or in short *PDF*.
- Next we will start talking about discrete and continuous random variables and their distributions.

1. **Random Variables**
   - Definitions, Discrete and Continuous RVs, and Prob. Distributions

2. **Discrete Random Variables**
   - 1. Probability Distribution and the idea of PMF
   - 2. CDF, quantiles and percentiles
   - 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$
   - 4. Rules of Expectation and Variance

3. **Parametric Distributions : Discrete**
   - 1. Discrete Uniform
   - 2. Bernoulli and Binomial Distribution

4. **Continuous Random Variables**
   - 1. Probability Distribution and the idea of PDF
   - 2. CDF, quantiles and percentiles
   - 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$

5. **Parametric Distributions : Continuous**
   - 1. Uniform Distribution
   - 2. Normal Distribution

Discrete Random Variables

Discrete Random Variables

1. Probability Distribution and the idea of PMF

- The distribution of a discrete random variable is known as *discrete probability distribution*. We already saw examples of probability distribution of a discrete random variable. And we know that for a discrete random variable it is always enough to know the *probabilities at all values x* that the discrete random variable can take. This means we need to know $\mathbb{P}(X = x)$ at all x in the range of X.
- Again here is another example of a discrete probability distribution when X is a random variable that counts the number of heads when we toss three coins. We can write the probability distribution of this random variable as follows,

$$\mathbb{P}(X = x) = \begin{cases} 1/8 & \text{when } x = 0 \\ 3/8 & \text{when } x = 1 \\ 3/8 & \text{when } x = 2 \\ 1/8 & \text{when } x = 3 \\ 0 & \text{for any other numbers in } \mathbb{R} \end{cases} \tag{1}$$

- The idea of *Probability mass function or PMF* is just a function of x which gives us $\mathbb{P}(X = x)$ directly, it's like if you know the PMF of a discrete random variable, you know the distribution of the random variable. Here is the formal definition of PMF,

Definition 3.2: (Probability Mass Function (PMF))

If X is a discrete random variable then the probability mass function (in short PMF) of X, denoted by f, is defined as the function such that *for all $x \in \mathbb{R}$*,

$$f(x) = \mathbb{P}(X = x)$$

The set $\{x : f(x) > 0\}$ is called the *support of (the distribution of) X*.

- It's important to note that this function is defined for all real numbers but when it has positive values we call the set of those points *the support of the distribution*. For example the distribution that we saw, we can think about following PMF function,

$$f(x) = \begin{cases} 1/8 & \text{if } x = 0 \\ 3/8 & \text{if } x = 1 \\ 3/8 & \text{if } x = 2 \\ 1/8 & \text{if } x = 3 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- You will understand the importance of PMF when we start talking about theoretical distributions, but for now you should think *if someone gives you the PMF of a discrete random variable, then you can calculate any probabilities for the random variable can possibly take*.

- Since it's a function we can actually plot this function, here is a plot

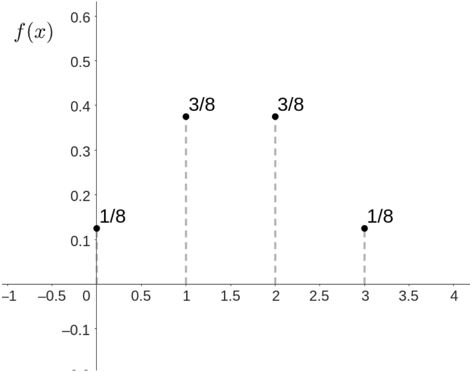


Figure 2: Plot of the PMF in (2)

- The PMF has two important properties, both are easy to understand

Theorem 3.3: (Properties of PMF)

- **The Sum of all PMF Values is 1**: This means if $x_1, x_2, \ldots x_k$ are all the possible values of a discrete random variable X, then

$$\sum_{i=1}^{k} f\left(x_i\right) = f(x_1) + f(x_2) + \ldots f(x_k) = 1$$

- **The Probability of an Interval in $\mathbb{R}$ can be Calculated as the Sum of PMF Values in the Set**: For example if we are talking about an interval $[2, 3]$, then we can calculate $\mathbb{P}(X \in [2, 3])$ with

$$\mathbb{P}(X \in [2, 3]) = \mathbb{P}(2 \leq X \leq 3) \sum_{x_i \in [2,3]} f\left(x_i\right)$$

- The first one is obvious, it says if you sum all the PMF values, then you should get 1. You already know this since probabilities have to be summed to 1 for all values when we are talking about discrete random variables.
- The second one says the probability of any interval in $\mathbb{R}$ can be calculated by adding the PMF values in that interval. We already saw the application of this rule in page 19/38. Recall we calculated

$$\mathbb{P}(X \in [1, 1.5)) = \mathbb{P}(X = 1) = f(1) = \frac{3}{8}$$

- The key thing to understand here is *knowing PMF is enough to calculate probabilities of different intervals in* $\mathbb{R}$.

- It is important to mention that Anderson et al. (2020) uses the terminology *probability function* for PMF. But essentially they are same thing. So if you are reading Chapter 5.2 in Anderson et al. (2020), then *probability function* means *probability mass function*. Because probability mass function or PMF is more common in the literature I will use PMF.

- A question remains, how do we use the idea of PMF in real life examples, actually we will later see that in real life we will model probability using some theoretical distributions, e.g., for discrete random variables we have Discrete Uniform, Bernoulli, Binomial, Poisson, etc. And similarly for continuous random variables we have Uniform, Normal, Exponential, etc.
- For now let's understand the idea of PMF using an empirical data. In this case we can call this *empirical probability mass function*, notice this is very similar to calculating *relative frequency*. However you need to be careful, since this is not the actual probability mass function, since the actual probability mass function can be calculated only using the population.

- **Example 3.4**: (Random Variable and empirical PMF - Soft Drink example from Anderson et al. (2020)). Recall the soft drink example we saw before. The data was available for a soft drink company for *last* 50 *sales*, here is the data

Sale	Brand Purchased	Sale	Brand Purchased
1	Coca-Cola	26	Coca-Cola
2	Diet Coke	27	Coca-Cola
3	Pepsi	28	Coca-Cola
4	Diet Coke	29	Pepsi
5	Coca-Cola	30	Coca-Cola
6	Coca-Cola	31	Sprite
7	Dr. Pepper	32	Dr. Pepper
8	Diet Coke	33	Pepsi
9	Pepsi	34	Diet Coke
10	Pepsi	35	Pepsi
11	Coca-Cola	36	Coca-Cola
12	Dr. Pepper	37	Coca-Cola
13	Sprite	38	Coca-Cola
14	Coca-Cola	39	Pepsi
15	Diet Coke	40	Dr. Pepper
16	Coca-Cola	41	Coca-Cola
17	Coca-Cola	42	Diet Coke
18	Sprite	43	Pepsi
19	Coca-Cola	44	Pepsi
20	Diet Coke	45	Pepsi
21	Coca-Cola	46	Pepsi
22	Diet Coke	47	Coca-Cola
23	Coca-Cola	48	Dr. Pepper
24	Sprite	49	Pepsi
25	Pepsi	50	Sprite
		51	X = ?

Figure 3: Note that 51*st* sale is a question mark, so we can think about a random variable in that place

Here we can think about a random experiment, that is *sales of the 51st soft drink*, Now we can think about a random variable X that represents soft drink brand. And here is how we can think about a random variable (Keep in mind that a random variable is always a number!)

$$X = 1 \text{ for Coca-Cola}$$
$$X = 2 \text{ for Diet Coke}$$
$$X = 3 \text{ for Dr. Pepper}$$
$$X = 4 \text{ for Pepsi}$$
$$X = 5 \text{ for Sprite}$$

- Since we already have the frequency distribution, we can write,

Brand	RV X	Frequency	Relative Frequency and f(x)
Coca-Cola	$X = 1$	19	0.38
Diet Coke	$X = 2$	8	0.16
Dr. Pepper	$X = 3$	5	0.1
Pepsi	$X = 4$	13	0.26
Sprite	$X = 5$	5	0.1
Grand Total		50	1

We can also simply write the *empirical PMF* as

x	$f(x)$
1	$19/50 = .38$
2	$8/50 = .16$
3	$5/50 = .1$
4	$13/50 = .26$
5	$5/50 = .1$

- *Again be careful:* We have constructed an *empirical PMF*, and this is not the *true PMF of a random variable X is for the population*,
- *Question: What is the population here?*, *Ans:* The data of all sales starting from the opening of the store till it ends the store... so if we can get the population data then we can calculate the *true PMF* of the random variable X
- But in this case is impossible to get the population and calculate the true PMF, so we can only calculate the empirical PMF using a sample, and one can say it's an estimate of the true PMF.
- Later we will see that for the population we will usually use the idea of theoretical distributions.. .... So we will assume our true PMF follows some *known form of distributions*.... but more on this later...

Discrete Random Variables

2. CDF, quantiles and percentiles

Probability Distributions

CDF for Discrete R.V.

- If we know probability distribution of a discrete random variable, we can also calculate *cumulative probabilities*. Cumulative Probability means *probabilities up-to a certain value*. For example, following is a cumulative probability (you already know this!)

$$\mathbb{P}(X \leq 2)$$

- For a random variable X, this means *the probability of X taking value less than or equal to 2*.
- For example if the distribution and PMF of a random variable is given as,

$$\mathbb{P}(X = x) = f(x) = \begin{cases} 1/8 & \text{if } x = 0 \\ 3/8 & \text{if } x = 1 \\ 3/8 & \text{if } x = 2 \\ 1/8 & \text{if } x = 3 \\ 0 & \text{otherwise} \end{cases}$$

- Then from this we can calculate,

$$\begin{aligned} \mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\ &= f(0) + f(1) + f(2) \\ &= 1/8 + 3/8 + 3/8 = 7/8 \end{aligned}$$

- Like PMF represents probabilities in terms of a function. For cumulative probabilities we have a function called *cumulative distribution function* in short CDF.
- So CDF simply gives cumulative probabilities. It is a function defined on the real line, where for any value x it gives the cumulative probability upto that value. Here is the formal definition,

Definition 3.5: (The cumulative distribution function (CDF))

The Cumulative Distribution Function or CDF of of a random variable X is the function $F(x)$ defined as

$$F(x) = \mathbb{P}(X \leq x), \quad \text{for } -\infty < x < \infty$$

If X takes values $x_1, x_2, \ldots, x_k$, and $f(x)$ is the PMF, then

$$F(x) = \sum_{x_i \leq x} f(x_i)$$

- For example, if this is the PMF

$$f(x) = \begin{cases} 1/8 & \text{if } x = 0 \\ 3/8 & \text{if } x = 1 \\ 3/8 & \text{if } x = 2 \\ 1/8 & \text{if } x = 3 \\ 0 & \text{otherwise} \end{cases}$$

then we can easily find the CDF for different x. Here we need to find for 0, 1, 2 and 3

$$\begin{aligned} F(0) &= f(0) = 1/8 \\ F(1) &= f(0) + f(1) = 4/8 \\ F(2) &= f(0) + f(1) + f(2) = 7/8 \\ F(3) &= f(0) + f(1) + f(2) + f(3) = 1 \end{aligned}$$

- We can also think about what happens in the interval $(0, 1)$, note that in this interval X does not take any values, so the cumulative probabilities in this interval is 0
- Following is the CDF plot,

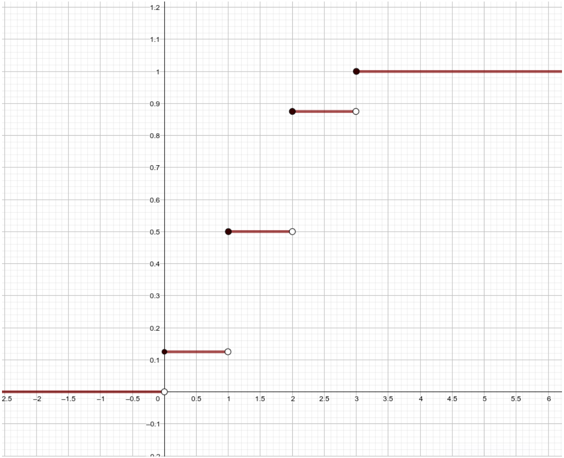


Figure 4: Notice there is a jump when the random variable takes its value, and the difference where there is a jump is the probability. Also note that the CDF function is defined for the whole $\mathbb{R}$

The last figure can be written as a piecewise function

$$F(x) = \begin{cases} 0 & : x < 0 \\ \frac{1}{8} & : 0 \le x < 1 \\ \frac{4}{8} & : 1 \le x < 2 \\ \frac{7}{8} & : 2 \le x < 3 \\ 1 & : x \ge 3 \end{cases}$$

There are 3 important properties for the CDF,

- 1) **Always Non-Decreasing:** If $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
  This makes sense since as we go to the right on the real line, the cumulative probabilities will increase or stay the same, it cannot decrease.
- 2) **Right Continuous:** $F(a) = \lim_{x \to a^+} F(x)$.
  This means the right hand limit of the CDF at a is equal to the value of the CDF at a (Do you understand what is a limit?)
- 3) **At infinity the limits are** 0 **and** 1 $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.
  This means as we go to the left of the real line, the CDF will approach 0, and as we go to the right of the real line, the CDF will approach 1. This makes sense since at $-\infty$ we have no probabilities, and at $+\infty$ we have all probabilities.

Discrete Random Variables

3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$

- Now we will learn two important summary measures of the distribution, namely *expected value* and *variance*.
- Essentially a summary measure is a single number that summarizes the probability distribution of the random variable.
- Let's start with *Expected value* or *Expectation* of a random variable.
- *An Expected value is a number that gives us an idea about the central value of the distribution*, or where most of the values of the random variables are concentrated.
- Calculating an Expected Value is very easy, here is the definition,

Definition 3.6: (Expected Value)

If X is a *discrete random variable* with values $x_1, x_2, \ldots, x_k$, and it has PMF $f(x)$ then the *Expectation* or the *Expected Value* of X is defined as

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

We will usually use the notation $\mathbb{E}(\cdot)$ to write the *Expectation* on X, and often the number or the expected value is represented with μ

- The formula says for a discrete random variable, if we know the PMF, then calculating Expected value is just *multiplying x with $f(x)$ and then adding them altogether*.
- Again if we use the following PMF of a random variable X,

$$f(x) = \begin{cases} 1/8 & \text{if } x = 0 \\ 3/8 & \text{if } x = 1 \\ 3/8 & \text{if } x = 2 \\ 1/8 & \text{if } x = 3 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

- then we can calculate the expected value as,

$$\begin{aligned} \mathbb{E}(X) &= (0 \times f(0)) + (1 \times f(1)) + (2 \times f(2)) + (3 \times f(3)) \\ &= (0 \times 1/8) + (1 \times 3/8) + (2 \times 3/8) + (3 \times 1/8) = 1.5 \end{aligned}$$

- So calculation is very easy, now we may ask *what does expected value mean*?

- Actually Expectation (or Expected value) is a *population average*, or *population mean*, so if we have a population of size N, then we can calculate the expected value as,

$$\frac{\text{sum of all values in the population}}{N}$$

- In expected value we are calculating the same number but the idea is we are *weighting values with their probabilities*... let's explain this with a concrete example

- **Example 1: Idea of Expectation**
  Suppose we have following *population data* (Careful: It;s not a sample, it's a population)

$$\text{Population Data} = \{1, 1, 1, 2, 1, 1, 2, 2, 3, 1, 3, 3, 3, 1\}$$

Then in this case population size is $N = 14$, and we can calculate the population average or population mean as

$$\frac{1+1+1+2+1+1+2+2+3+1+3+3+3+1}{14} = \frac{24}{14}$$

Now if we think about a random variable X that can take values $1, 2$ and 3 then we can think about following *true PMF*,

x	$f(x)$
1	$7/14$
2	$3/14$
3	$4/14$

Now with this PMF we can calculate the expected value as

$$\begin{aligned}\mathbb{E}(X) &= (1 \times f(1)) + (2 \times f(2)) + (3 \times f(3)) = (1 \times 7/14) + (2 \times 3/14) + (3 \times 4/14)\\ &= \frac{7+6+12}{14} = \frac{25}{14}\end{aligned}$$

which gives us the same result as before.

- You may ask why we are learning this formula? Why not directly take average of all values? Two reasons,
  - Almost always we don't have access to the population data, so we cannot calculate the average of all values, but the nice thing about the formula for the expected value is *we can apply the formula when we only know true PMF*. So the formula for expected value gives us the population average even if we don't have all population data and somehow only know true PMF.
  - We can extend this idea easily to the continuous case, the idea is replace the $\sum$ with $\int$ (We will see this later)

- *Homework Question:* Suppose from the population data that we have just seen in the last example we take a *random sample* of 5 numbers (we did *sampling with replacement)*, below is the Population data and the Sample data,

$$\text{Population Data} = \{1, 1, 1, 2, 1, 1, 2, 2, 3, 1, 3, 3, 3, 1\}$$
$$\text{Sample Data} = \{1, 1, 2, 2, 3\}$$

- Note here the population size $N = 14$ and the sample size is $n = 5$, now answer following question.
- Calculate *True PMF and Empirical PMF* (We already solved one)
- Calculate the Expected Value using the *True PMF* and Direct average? (We already did this)
- Calculate the sample average using the formula for the sample average.
- What is the relationship between True PMF and Empirical PMF?
- What is the relationship between Expected Value (or Population Average) and Sample Average?

- Like Expectation, variance is also a summary measure, where the expectation gives an idea of the central value or average, variance gives the idea how *dispersed the values are* (you already know sample variance, but we will learn the population variance formula now)

Definition 3.7: (Variance)

If X is a *discrete random variable* with values $x_1, x_2, \ldots, x_k$, and it has PMF $f(x)$ then the *Variance* of X is defined as

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left(\left(X - \mathbb{E}\left(X\right)\right)^2\right) = \mathbb{E}\left(\left(X - \mu\right)^2\right) = \sum_{i=1}^{k} (x_i - \mu)^2 f(x_i)$$

where we used $\mu = \mathbb{E}(X)$. Also for the calculated variance we often use the symbol σ^2.

- First note that, Variance is also an Expectation, but it is an *Expectation of* $(X-\mu)^2$, NOT X.
- So what is $(X-\mu)^2$? Or what is $(X-\mu)$? Ans: $X-\mu$ is the *deviation of the random variable from its Mean* and $(X-\mu)^2$ is the *squared deviation*.
- So Population Variance is the *Expectation of the squared deviation* (or the population average of the squared deviation!)
- Let's calculate $\mathbb{V}\mathrm{ar}(X)$ for the random variable X from the example that we have been using,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(X) &= \left((0-1.5)^2 \times f(0)\right) + \left((1-1.5)^2 \times f(1)\right) + \\
&\quad \left((2-1.5)^2 \times f(2)\right) + \left((3-1.5)^2 \times f(3)\right) \\
&= \left((-1.5)^2 \times 1/8\right) + \left((-0.5)^2 \times 3/8\right) + \left((0.5)^2 \times 3/8\right) + \left((1.5)^2 \times 1/8\right) \\
&= (2.25 \times 1/8) + (0.25 \times 3/8) + (0.25 \times 3/8) + (2.25 \times 1/8) \\
&= 0.75
\end{aligned}$$

- Like population average, population variance can be also calculated, using

$$\frac{\text{sum of all squared deviations in the population}}{N}$$

- But the formula is using the PMF, so same idea like population mean - here we are weighting the squared deviations using the PMF.

- So calculating Variance is really easy, if we know PMF we can easily calculate the variance.
- The interpretation of the variance is how dispersed the values are.
- As you already know the square root of the variance is called *Standard Deviation*, so if σ^2 is the Variance, then σ is the standard deviation.

- *Homework Question:* Continuing from the last Homework Question...
- Calculate the Variance using the *True PMF* and Direct average of the squared deviations? Are they same?
- Calculate the sample variance using the formula for the sample variance.
- What is the relationship between Population Variance and Sample Variance?

Discrete Random Variables

4. Rules of Expectation and Variance

Rules of Expectation and Variance

- We already saw the idea of Expectation, you should always keep the following picture in your mind, that expectation works on random variables, not on number, and the result of Expectation is a constant.

Expected Value or Mean is a constant

$$\mathbb{E}(X) = 2.5$$

Random Variable (which is NOT a constant,
and it possible that it takes many values).
But if we take expectation, we are asking
for its Mean value. The Mean or Expected value is always a constant.

Rules of Expectation and Variance

- Now we consider a slightly different problem, we ask what is

$$\mathbb{E}(X^2) \text{ or } \mathbb{E}(X^3) \text{ or } \mathbb{E}(3+2X) \text{ ???}$$

- Note that X^2, X^3 or $3+2X$ are all functions of random variable X.
- So now our question is *for functions X^2, X^3 or $3+2x$, what are the expected values*. First of all note that *any function of a random variable is also a random variable*.
- It turns out that for a function $g(X)$ in this case we can calculate $\mathbb{E}(g(X))$ by using the distribution of X, the idea is the following,

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

- This rule has an interesting name, it is called - *Law of the unconscious Statistician* or in short *LOTUS*
- Why this name? Since we just used the distribution of X blindly...

Rules of Expectation and Variance

- Applying this rule we can easily calculate expectation of All functions of X.

$$\mathbb{E}(X^2) = \sum_{i=1}^{k} x_i^2 f(x_i)$$

$$\mathbb{E}(X^3) = \sum_{i=1}^{k} x_i^3 f(x_i)$$

$$\mathbb{E}(3 + 2X) = \sum_{i=1}^{k} (3 + 2x_i) f(x_i)$$

- In the last case, we have a simpler way to calculate, we can use the following rule,

$$\mathbb{E}(3 + 2X) = 3 + 2\mathbb{E}(X)$$

- This is called the *linearity of expectation*, and it is a very important rule. In general the rule says,

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

- where a and b are any constants.
- We will see the proof of this rule, but before let's learn some rules for the summation...

Rules of Expectation and Variance

Rules 3.8: Algebra Rules for Summation

- *Sum and Difference Rule*:

$$\sum_{i=1}^{n}(x_i \pm y_i) = \sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} y_i$$

- *Constant Multiple Rule:* For any constant or fixed number c

$$\sum_{i=1}^{n} c\, x_i = c \cdot \sum_{i=1}^{n} x_i$$

- *Constant Value Rule:* For any constant or fixed number c

$$\sum_{i=1}^{n} c = n \cdot c$$

- These rules are very useful when it comes to doing summation. Let's see some examples, you will see more examples in the homework.

- **Example 3.9**:
  (a) $\sum\limits_{k=1}^{n}\left(3k-k^2\right)=3\sum\limits_{k=1}^{n}k-\sum\limits_{k=1}^{n}k^2$
  (b) $\sum\limits_{k=1}^{n}(-a_k)=\sum\limits_{k=1}^{n}(-1)\cdot a_k=-1\cdot\sum\limits_{k=1}^{n}a_k=-\sum\limits_{k=1}^{n}a_k$
  (c)

$$\begin{aligned}\sum_{k=1}^{3}(k+4)&=\sum_{k=1}^{3}k+\sum_{k=1}^{3}4\\&=(1+2+3)+(3\cdot 4)\\&=6+12=18\end{aligned}$$

  (d) $\sum\limits_{k=1}^{n}\frac{1}{n}=n\cdot\frac{1}{n}=1$
  (e) (Do this!)

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2$$

notice $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is also a formula for sample variance, in fact when n is very large, there is not so much difference between $\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ and $\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

Rules of Expectation and Variance

Now let's apply the rules to see whether the linearity of expectation is correct...

$$\mathbb{E}[a + bX] = \sum_{i=1}^{k}(a + bx_i)f(x_i) \tag{4}$$

$$= \sum_{i=1}^{k} af(x_i) + \sum_{i=1}^{k} b\, x_i\, f(x_i) \tag{5}$$

$$= a\sum_{j=1}^{k} f(x_i) + b\sum_{j=1}^{n} x_i f(x_i) \tag{6}$$

$$= a + b\mathbb{E}[X] \tag{7}$$

Notice the *linearity of expectation* rule simply says, 1) Expectation of a constant is always constant and 2) If constant is multiplied with a random variable, we can always pull it out from the Expectation.

Linearity is actually remarkable property of Expectation, later we will see that we can apply this property for many random variables together.

Rules of Expectation and Variance

Notice interestingly if you look at the formula for the Variance, it already has the idea of LOTUS, since

$$\begin{aligned}\mathbb{V}\mathrm{ar}(X) &= \mathbb{E}\left[(X-\mu)^2\right] \\ &= \sum_{i=1}^{k}(x_i-\mu)^2 f(x_i)\end{aligned}$$

But interestingly using *linearity of expectation* we can get another formula for variance,

$$\begin{aligned}\mathbb{V}\mathrm{ar}(X) &= \mathbb{E}\left[(X-\mu)^2\right] \\ &= \mathbb{E}\left[X^2-2X\mu+\mu^2\right] \\ &= \mathbb{E}\left[X^2\right]-2\mu\mathbb{E}\left[X\right]+\mu^2 \\ &= \mathbb{E}\left[X^2\right]-2\mu^2+\mu^2 \\ &= \mathbb{E}\left[X^2\right]-\mu^2 = \mathbb{E}\left[X^2\right]-(\mathbb{E}\left[X\right])^2\end{aligned}$$

So both are valid formulas, and you can use any of them.

Rules of Expectation and Variance

So this means for the population variance, we have two equivalent ways of writing variance, and they are,

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

and for the sample variance with $\frac{1}{n}$, we have

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

Rules of Expectation and Variance

- Now we can also ask what is $\mathbb{V}\mathrm{ar}(a + bX)$?
- It turns out that in this case, we can use the following rule,

$$\mathbb{V}\mathrm{ar}(a + bX) = \mathbb{V}\mathrm{ar}(a) + \mathbb{V}\mathrm{ar}(bX) = 0 + b^2\mathbb{V}\mathrm{ar}(X) = b^2\mathbb{V}\mathrm{ar}(X)$$

- where $\mathbb{V}\mathrm{ar}(a) = 0$, since a is not a random variable, and it has a fixed value, so no variance.
- What about $\mathbb{V}\mathrm{ar}(bX)$? This is also easy, we can check this using the definition of variance and the linearity of expectation,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(bX) &= \mathbb{E}\left[\left(bX - \mathbb{E}\left[bX\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(bX - b\mathbb{E}\left[X\right]\right)^2\right] \\
&= \mathbb{E}\left[\left(b\left(X - \mathbb{E}\left[X\right]\right)\right)^2\right] \\
&= \mathbb{E}\left[b^2\left(X - \mathbb{E}\left[X\right]\right)^2\right] \\
&= b^2\mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)^2\right] \\
&= b^2\mathbb{V}\mathrm{ar}(X)
\end{aligned}$$

Rules of Expectation and Variance

- This gives us two following rules for the linear function of X, that is

$$\mathbb{E}(a + bX) = a + b\mathbb{E}(X)$$
$$\mathbb{V}\mathrm{ar}(a + bX) = b^2\mathbb{V}\mathrm{ar}(X)$$

Rules of Expectation and Variance

Homework Question:

- Suppose we have a random variable X that takes 3 values, 1, 2, and 3 with following PMF,

x	$f(x)$
1	1/4
2	1/2
3	1/4

- Calculate Expected Value and Variance using the PMF.
- Now using Linearity of Expectation and LOTUS, calculate $\mathbb{E}(2+X^2)$, $\mathbb{E}(3X+X^3)$ and $\mathbb{E}(3+2X)$.
- Calculate $\mathbb{V}\mathrm{ar}(3+2X)$ using the rules of variance.

1. Random Variables

- Definitions, Discrete and Continuous RVs, and Prob. Distributions

2. Discrete Random Variables

- 1. Probability Distribution and the idea of PMF
- 2. CDF, quantiles and percentiles
- 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$
- 4. Rules of Expectation and Variance

3. Parametric Distributions : Discrete

- 1. Discrete Uniform
- 2. Bernoulli and Binomial Distribution

4. Continuous Random Variables

- 1. Probability Distribution and the idea of PDF
- 2. CDF, quantiles and percentiles
- 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$

5. Parametric Distributions : Continuous

- 1. Uniform Distribution
- 2. Normal Distribution

Parametric Distributions : Discrete

Parametric Distributions : Discrete

1. Discrete Uniform

- There are many theoretical discrete distributions, these are often called parametric distributions.
- The idea of parameter is - it's simply a number (or a set of numbers) such that if we change the number, the distribution will change.
- There are many parametric discrete distributions, and we will see some of them in this course.
  - *Discrete Uniform Distribution*,
  - *Bernoulli distribution*,
  - *Binomial distribution* and
  - *Poisson distribution*.
- Let's start with *Discrete Uniform Distribution*, and will talk about other distributions in coming sections.

- *Discrete Uniform Distribution* is similar to the idea of equally likely outcomes but this is for the values of the random variable.

Definition 3.10: (Discrete Uniform Distribution)

If X can take values $x_1, x_2, \ldots, x_k$ then we say X follows *Discrete Uniform* distribution with parameter $\{x_1, x_2, \ldots, x_k\}$ if the PMF of X can be written as,

$$f(x) = \left\{ \begin{array}{ll} \frac{1}{k} & \text{for } x = x_1, x_2, \ldots, x_k \\ 0 & \text{otherwise} \end{array} \right\}$$

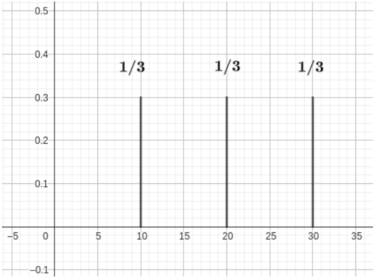
We write this as $X \sim \mathrm{DUnif}\{x_1, x_2, \ldots, x_k\}$.

- Here parameter determines which specific Uniform Distribution we have and changing the parameter will change the distribution. It will be still a discrete uniform distribution but the distribution will change (see example in the next slide).
- Note that X can take values in any finite set, the idea of a discrete uniform random variable is the *values will all have equal probabilities*. Let's see a real life example.

- Suppose everyday your brother can give you 10, 20 or 30 taka and he might give you any of the three amounts with equal probability.
- So Let X be a random variable that represents the amount of money, then X follows a discrete uniform distribution with parameter $\{10, 20, 30\}$, we write $X \sim \mathrm{DUnif}\{10, 20, 30\}$ and the PMF of X is

x	$f(x)$
10	1/3
20	1/3
30	1/3

- How does it look like?



- Notice here parameter is $\{10, 20, 30\}$, so if we change the parameter, for example, if we change the parameter to $\{10, 20, 30, 40\}$, then the distribution will change.
- For the following Discrete Uniform PMF, We can easily calculate the Expected Value and Variance.

x	$f(x)$
10	1/3
20	1/3
30	1/3

- Using the formula for the expected value we can calculate,

$$\begin{aligned}
\mathbb{E}(X) &= (10 \times f(10)) + (20 \times f(20)) + (30 \times f(30)) \\
&= (10 \times 1/3) + (20 \times 1/3) + (30 \times 1/3) \\
&= (10 + 20 + 30) \times \frac{1}{3} = \frac{60}{3} = 20
\end{aligned}$$

- Using the formula for the variance we can calculate,

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}(X) &= \mathbb{E}\left(\left(X - \mathbb{E}\left(X\right)\right)^2\right) = \mathbb{E}\left(\left(X - 20\right)^2\right) \\
&= \left(\left(10 - 20\right)^2 \times f(10)\right) + \left(\left(20 - 20\right)^2 \times f(20)\right) + \\
&\quad \left(\left(30 - 20\right)^2 \times f(30)\right) \\
&= \left(\left(-10\right)^2 \times 1/3\right) + \left(\left(0\right)^2 \times 1/3\right) + \left(\left(10\right)^2 \times 1/3\right) \\
&= (100 \times 1/3) + (0 \times 1/3) + (100 \times 1/3) \\
&= (100 + 0 + 100) \times \frac{1}{3} = \frac{200}{3} = 66.67
\end{aligned}
$$

Parametric Distributions : Discrete

2. Bernoulli and Binomial Distribution

Bernoulli & Binomial Random Variables

- If a random variable X only has two values 0 and 1, we call the random variable a *Bernoulli Random variable*, and its distribution is known as *Bernoulli distribution*, some examples could be,
  - When we toss a coin, a random variable $X = 1$ means head, $X = 0$ means tail.
  - When we sample then Gender of a person, So $X = 1$ means female, and $X = 0$ means male
  - When we call a customer, a Random Variable X could be such that $X = 1$ means picked up the call, $X = 0$ means didn't pickup.
  - And so on....
- In practice or in real life scenario, when you have possible data with $0, 1, 0, 1, 0, 1$, you can think about these are values of some Bernoulli random variables. So in any experiment, when we have only two possible outcomes we can think about modeling that experiment using a Bernoulli random variable. For a Bernoulli random variable, if $X = 1$, we often call it *"success"* and if $X = 0$, we often call it "failure". Here is the formal definition.

Definition 3.11: (Bernoulli Distribution)

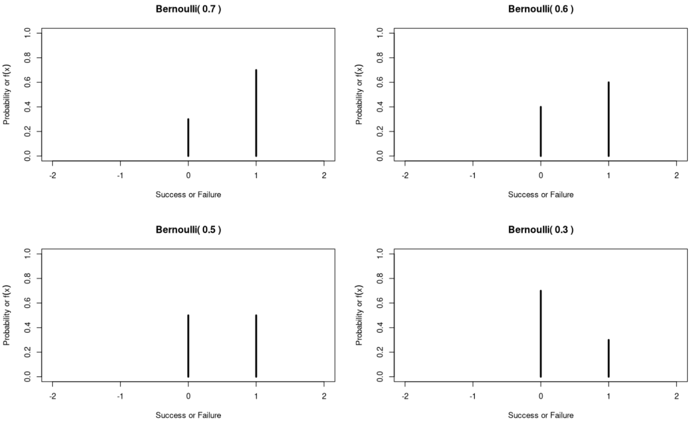
A random variable X is said to follow Bernoulli distribution with parameter p if the PMF of X can be written as,

$$\begin{aligned} f(x) &= p^x(1-p)^{1-x}, \quad \text{when } x = 0, 1 \\ &= 0, \quad \text{otherwise} \end{aligned}$$

where $0 < p < 1$. We write $X \sim \text{Bern}(p)$ to represent X follows Bernoulli Distribution with parameter p

- Note that, because of this PMF, we have $\mathbb{P}(X = 1) = f(1) = p$ and $\mathbb{P}(X = 0) = f(0) = 1 - p$.
- Notice, the parameter p controls the probability and hence controls the distribution of the random variable. For example, if $X \sim \text{Bern}(0.3)$, this automatically means $\mathbb{P}(X = 1) = 0.3$ and $\mathbb{P}(X = 0) = 0.7$. Here is how the PMF will look like for different parameters p,

Bernoulli & Binomial Random Variables



- Because we have the PMF we can also calculate the Expected Value and Variance of a Bernoulli random variable.
- If you do the calculation, then you should get $\mathbb{E}(X) = p$ and $\mathbb{V}\mathrm{ar}(X) = p(1-p)$ (please do the calculation!)

Bernoulli & Binomial Random Variables

- Now let's talk about Binomial distribution, The Binomial distribution comes when we perform *more than one independent Bernoulli experiments*.
- Here is the story - Suppose now we perform n independent Bernoulli experiments (or *Bernoulli trials*, like tossing a coin) with parameter p.
- If X is a random variable which represents the total number of success out of the n trials, then we say the random variable X follows Binomial distribution with parameter n and p.
- You already know the example, ... number of heads ... recall..
- Here is the formal definition

Definition 3.12: (Binomial Distribution)

Suppose that n independent Bernoulli trials are performed, each with the same success probability p, we say X follows Binomial distribution with parameters n and p if the PMF of X can be written as

$$
\begin{aligned}
f(x) &= \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, 2, \ldots, n \\
&= 0, \ \text{otherwise}
\end{aligned}
$$

We write $X \sim \text{Bin}(n, p)$ to mean that X has the Binomial distribution with parameters n and p, where n is a positive integer and $0 < p < 1$.

Bernoulli & Binomial Random Variables

- Notice here x is the value of the random variable where x can be $0, 1, 2, \ldots, n$. The PMF looks very similar to Bernoulli PMF, except we have a combination term, recall

$$\binom{n}{x} = {}^{n}C_{x} = \frac{n!}{x!(n-x)!}$$

- Question is why is this coming?
- Here is a short answer, the experiment consisting of n independent Bernoulli trials produces a sequence of successes and failures. The probability of any specific sequence of x successes and $n-x$ failures is $p^{x}(1-p)^{n-x}$. There are ${}^{n}C_{x}$ such sequences.
- We can also calculate the Mean and the Variance of the Binomial distribution. *If $X \sim Bin(n, p)$, then $\mathbb{E}(X) = np$ and $\mathbb{V}\mathrm{ar}(X) = np(1-p)$*, where do we get this? You can see the proof in the next page. However there is an easy trick that is you remember Binomial is the sum n independent Bernoulli trials (Let's do it using easy trick!)
- The easy trick is applying linearity of expectation... and for variance applying the idea of independence.... from Bernoulli...

Bernoulli & Binomial Random Variables

- If $X \sim \mathrm{Bin}(n, p)$, then applying the formula for Expectation
$$\mathbb{E}(X) = \sum_{x=0}^{n} x f(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}.$$
- Note here we wrote $q = (1-p)$. Also note that we have $x\binom{n}{x} = n\binom{n-1}{x-1}$, so
$$\begin{aligned}\sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} &= n \sum_{x=0}^{n} \binom{n-1}{x-1} p^x q^{n-x} \\ &= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{n-x} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} \\ &= np.\end{aligned}$$
Then *we get $\mathbb{E}(X) = np$, similarly we can derive the variance is $\mathbb{V}\mathrm{ar}(X) = np(1-p)$ (I am skipping the proof). Again the easy trick is to apply Binomial - Bernoulli relation*.
- Here is how the PMF will look like for two Binomial distributions, with same $n = 3$ but different p.

Bernoulli & Binomial Random Variables

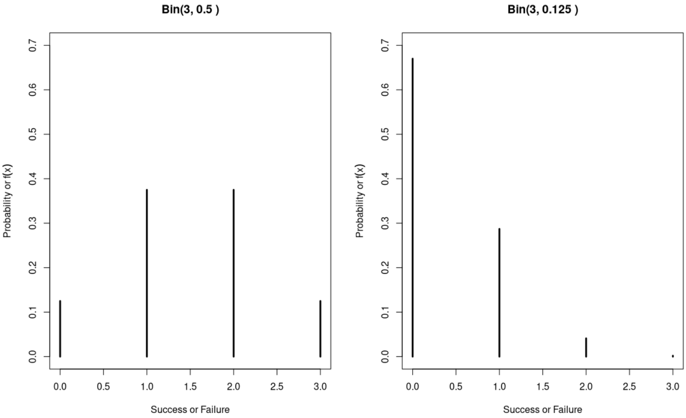


Figure 5: On the left we have the PMF of $X \sim \mathrm{Bin}(3, 0.5)$ and on the right we have $X \sim \mathrm{Bin}(3, 0.125)$.

Bernoulli & Binomial Random Variables

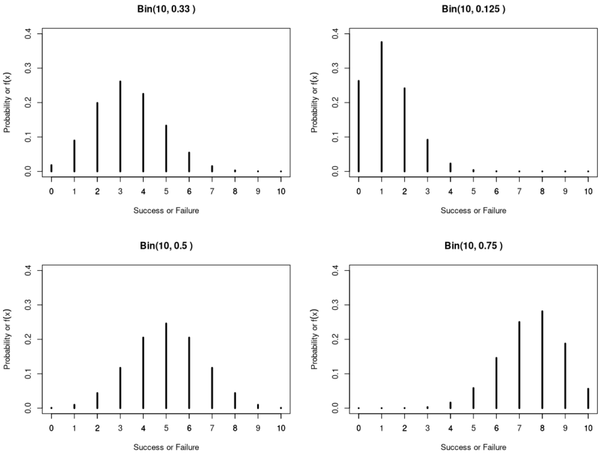


Figure 6: From top left, we have the PMF of $X \sim \mathrm{Bin}(10, 0.33)$, then right $X \sim \mathrm{Bin}(10, 0.125)$, then bottom left $X \sim \mathrm{Bin}(10, 0.5)$ and right $X \sim \mathrm{Bin}(10, 0.75)$

Bernoulli & Binomial Random Variables

- Binomial distribution comes in many forms in real life, you should always remember the essential structure - *that is tossing* n *independent coins and then the random variable is the number of success out of* n.
- Here are some examples where we can think about a Binomial random variable.
  - Random experiment: Calling n people. Random variable X will represent how many people will answer the call.
  - Random experiment: n students registered for a course. Random variable X will represent how many students will finish it.
  - Random experiment: Randomly asking 5 people whether they are satisfied with the transportation system of Bangladesh. Random variable is number of people who said "yes"!
  - And there are more examples in Anderson et al. (2020).
- Notice two important assumptions for the Binomial random variable is, 1) all trials are independent and 2) all trials happens with same probability. Only in these cases you can think about the random variable is Binomial.

1. **Random Variables**
   - Definitions, Discrete and Continuous RVs, and Prob. Distributions

2. **Discrete Random Variables**
   - 1. Probability Distribution and the idea of PMF
   - 2. CDF, quantiles and percentiles
   - 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$
   - 4. Rules of Expectation and Variance

3. **Parametric Distributions : Discrete**
   - 1. Discrete Uniform
   - 2. Bernoulli and Binomial Distribution

4. **Continuous Random Variables**
   - 1. Probability Distribution and the idea of PDF
   - 2. CDF, quantiles and percentiles
   - 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$

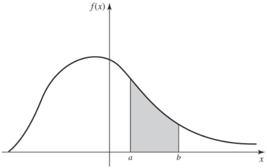
5. **Parametric Distributions : Continuous**
   - 1. Uniform Distribution
   - 2. Normal Distribution

Continuous Random Variables

Continuous Random Variables

1. Probability Distribution and the idea of PDF

- Now let's talk about continuous distributions. How do we extend the idea of PMF, Expectation and Variance for continuous random variables?
- Recall histogram, the idea of histogram is very similar to PMF, but the difference is PMF is for discrete random variable and histogram is for continuous random variable where we have a lot of data points.
- In the histogram, we have some bins and we count the number of data points in each bin and then we divide by the total number of data points to get the relative frequency in each bin.
- What if we have a lot of data points and make the size of the bins very small?
- Can you visualize what happens? We might get a very smooth curve like this. it's called PDF



- Similar to PMF, for continuous random variable we have another function called *probability density function (or PDF)* to calculate the probabilities.
- Here the idea is if we know the *pdf of the random variable*, then we can calculate the probability of any interval in $\mathbb{R}$ with integral. For example here is a PDF, it's a function of x,

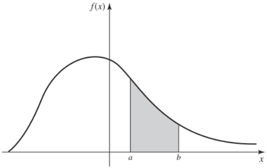


Figure 7: Here the shaded area is the probability of a random variable X taking value between a and b, so this means the shaded area is $\mathbb{P}(a \leq X \leq b]) = \int_a^b f(x)dx$

- Now how do we calculate probability of X takes value in the interval $[a, b]$, the idea is we can simply integrate, so $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$, since integration meas finding area under the curve, so the probability of X takes value in $[a, b]$ = area under the curve in the interval $[a, b]$.

- So you should remember for a continuous random variable X with density function $f(x)$

$$
\begin{aligned}
\text{Probability of } X \text{ taking value in the interval } [a,b] &= \mathbb{P}(a < X < b) \\
&= \int_a^b f(x)dx \\
&= \text{area under the density function}
\end{aligned}
$$

Here is the definition of a density function ..

Definition3.13: (Probability Density Function (PDF))

If X is a continuous random variable then a *nonnegative* function $f(x)$ on $\mathbb{R}$ is called the probability density function (in short PDF) of X if for any interval, for example $[a, b]$, we have

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

and it satisfies $\int_{-\infty}^{\infty} f(x)dx = 1$ (the density function should be integrated to 1).

- Note we can also calculate, $\mathbb{P}(X \geq a) = \int_a^{\infty} f(x)dx$ and $\mathbb{P}(X \leq b) = \int_{-\infty}^{b} f(x)dx$.
- Notice a PDF must satisfy two conditions,
  - 1. $f(x) \geq 0$, it is always non-negative for any value of x
  - 2. $\int_{-\infty}^{\infty} f(x)dx = 1$ (if we integrate for all values it will be 1)

- The idea of a density function for a continuous random variable X is same as the the probability mass function for a discrete random variable.
- For continuous random variable we will learn 3 important continuous distributions,
  - *Uniform distribution*,
  - *Normal distribution* and
  - *Exponential distribution*
- Let's see the uniform distribution now. Notice this is a continuous uniform distribution, the idea is very similar to discrete, but it's for a continuous random variable.

- Intuitively, a Uniform random variable on the interval (a, b) is a completely random number between a and b. Here is the formal definition,

Uniform Distribution

A continuous random variable X is said to follow the *Uniform Distribution* on the interval (a, b) if its PDF is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

We denote this by $X \sim \text{Unif}(a, b)$, and we have

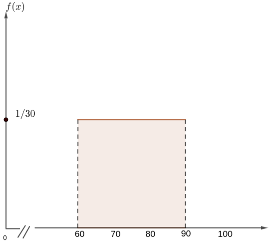
$$\mathbb{E}(X) = \frac{a+b}{2}$$
$$\mathbb{V}\text{ar}(X) = \frac{(b-a)^2}{12}$$

- Let's see a real life example, (this is adapted from Anderson et al. (2020))

- Think about a random variable X that represents the flight time of an airplane traveling from Dhaka to Chittagong. Suppose the flight time can be any value in the interval from 60 minutes to 90 minutes. With every 1-minute interval being equally likely, we can think the random variable X follows a uniform probability distribution, so $X \sim \text{Unif}(60, 90)$.
- Now because $\frac{1}{90-60} = \frac{1}{30}$, the PDF of X can be written as

$$f(x) = \begin{cases} \dfrac{1}{30} & \text{if } 60 \leq x \leq 90, \\ \\ 0 & \text{otherwise} \end{cases}$$

- How does it look like?

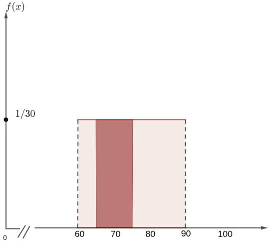


- Now with this density we can calculate the probability of X taking value in any interval, for example,

$$\mathbb{P}(65 < X < 75) = \int_{65}^{75} f(x)dx = \int_{65}^{75} \frac{1}{30} dx = \frac{1}{30}[x]_{65}^{75} = \frac{1}{30}(75 - 65) = \frac{10}{30} = \frac{1}{3}$$

- So this means there is $1/3$ probability that the flight time will be between 65 minutes and 75 minutes.

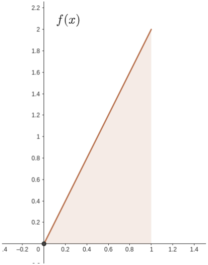
- Note that this also shows probability is an area under the curve, in the following this is the shaded area,



- Since in this case the area is a rectangle, we can apply the formula for area of a rectangle to calculate the area, which is $10 \times \frac{1}{30} = \frac{10}{30} = \frac{1}{3}$.
- Similarly, we can calculate $\mathbb{P}(70 \leq X \leq 80) = \frac{1}{3}$, $\mathbb{P}(75 \leq X \leq 85) = \frac{1}{3}$ and so on.
- In general for $X \sim \text{Unif}(a, b)$ we can calculate $\mathbb{P}(c \leq X \leq d) = \frac{1}{b-a} \times (d - c)$

- Here is another example of a PDF (this is not uniform), suppose we have a random variable X takes value in $[0, 1]$, with following PDF,

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$



- Is this a valid PDF? YES! note it satisfy two conditions
  - $f(x) \geq 0$ for all x
  - $\int_0^1 f(x)dx = \int_0^1 2xdx = 1$
- You should compare and contrast these two conditions with the conditions for a PMF.

- Can we calculate $\mathbb{P}(0.5 \leq X \leq 0.7) =$? Yes we can...

$$\int_{0.5}^{0.7} f(x)dx = \int_{0.5}^{0.7} 2xdx = [x^2]_{0.5}^{0.7} = 0.7^2 - 0.5^2 = 0.49 - 0.25 = 0.24$$

- So now we know $\mathbb{P}(0.5 \leq X \leq 0.7) = 0.24$
- Note in this case we cannot apply the formula for rectangle, since the area is not a rectangle. Uniform distribution is a special case where we can apply the formula for rectangle, but generally we need to use integration to calculate the area.
- so always remember integral = area = probability.

Some Important Remarks PDF

- When we calculate $f(x)$ for any x, *is this a probability, the answer is no*, it's just a function (look at the last example in some points it is more than 1). So the value of the density function $f(x)$ is not a probability, but it helps to to calculate probabilities when we do integration. Notice! this is an important difference with PMF: Unlike PMF, any PDF does not directly give us probabilities, we need to integrate this in a range and then we get a probability.
- Note: We calculated $\mathbb{P}(X \in [a, b]) = \int_a^b f(x)dx$, then using this you might want to calculate $\mathbb{P}(X = a) = \int_a^a f(x)dx$. Clearly this is 0, since $\int_a^a f(x)dx = 0$, so we get $\mathbb{P}(X = a) = 0$. Now this means for a continuous random variable X, for any constant, we will always have 0 probability. For example $\mathbb{P}(X = 3) = 0$, $\mathbb{P}(X = 100) = 0$ or $\mathbb{P}(X = 3.5) = 0$ and so on.
- From this you might conclude that *$X = a$ is impossible* because it happens with 0 probability. But isn't this strange or counter-intuitive? Because if this is impossible then X will not take any value at all, since we will always have 0 probability.
- So what is happening here? The last conclusion is actually not correct. It's not that *$X = a$ is impossible* rather what happens here is that the probability X is *spread so thinly* that we fail to calculate it precisely. This is why for a continuous random variable we can only calculate probabilities on any intervals or sets, *NOT for any fixed value*, so we write $\mathbb{P}(X = a) = 0$.
- Bonus Question: For a continuous random variable is there any difference between $\mathbb{P}(X \in (a, b))$ and $\mathbb{P}(X \in [a, b])$?

Continuous Random Variables

2. CDF, quantiles and percentiles

- If we know the PDF of a random variable, we can also actually easily calculate the cumulative distribution function or CDF. Notice $\mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x))$.
- Also we can see that $F(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \in (-\infty, x)) = \int_{-\infty}^{x} f(t)dt$. Here we used t in PDF because we have x in the limit.
- CDF is just a function we can find by taking these probabilities. A picture might be helpful here. Here is a PDF, and the associated CDF *

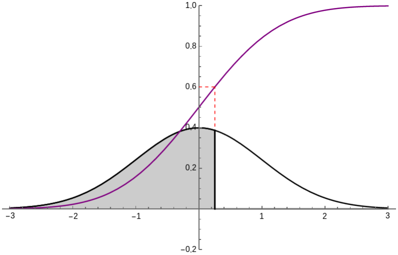


Figure 8: The density function $f(x)$ is the Bell-Shaped curve, the shaded are is $\mathbb{P}(X < .253) = \mathbb{P}(X \in (-\infty, .253)) = .60$. The function in the purple color is the cumulative distribution function (CDF) $F(x)$.

- So the CDF or the cumulative distribution function $F(x)$ simply gives us the cumulative probabilities at each x.
- Once we understand what is cumulative probabilities, we can understand *quantiles* or *percentiles*.
- In the last figure we showed

$$P(X \leq 0.253) = F(0.253) = 0.6$$

- In this case we say the number 0.253 is 0.6^{th} *quantile* of the distribution.
- Notice this actually means 60% values of the random variable is below 0.253.
- We also say 0.253 is the 60^{th} *percentile* of the distribution.
- So quantiles and percentiles are same things, when it come to quantiles we write in decimals, for example 0.6, 0.7, etc. However for percentile we write 60%, 70%.
- So if someone asks you what is 65% percentile or $.65^{th}$ quantile of the distribution, you should say this is a value below which there are 65% values of the random variable.

*You can use this nice Wolfram Demonstration, to have a clear idea, click here https://demonstrations.wolfram.com/PercentilesOfCertainProbabilityDistributions/

Continuous Random Variables

3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$

- Now let's see how to calculate the expected value of a continuous random variable X.

Definition 3.14: (Expectation)

If X is a continuous random variable with PDF $f(x)$, then the *Expectation* of X is defined as

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

- We do it for the random variable we saw already, recall the PDF is given by,

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

- We can calculate the expected value (just by *replacing sum with integration!*)

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x(2x)dx = \int_0^1 2x^2 dx \quad \ldots$$

$$\ldots = 2\int_0^1 x^2 dx = 2 \times \left[\frac{x^2}{3}\right]_0^1 = \frac{2}{3} \times \left[x^3\right]_0^1 \quad \ldots$$

$$\ldots = \frac{2}{3} \times \left[1^3 - 0^3\right] = \frac{2}{3} \times 1 = \frac{2}{3}.$$

- Let's do another example.
- Let's calculate the expected value of the Uniform distribution where $X \sim \mathrm{Unif}(a, b)$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\int_a^b xdx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$$

$$= \frac{1}{2(b-a)}\left[x^2\right]_a^b = \frac{1}{2(b-a)}\left[b^2 - a^2\right] = \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)(b-a)}{2(b-a)} = \frac{b+a}{2}$$

- This means Expected value of a Uniformly distributed random variable is the average of the two end points of the interval. This seems very intuitive, since the probability is same for all values, so the expected value should be the average of the two end points.
- Question: What is the expected value of the random variable X from the flight example? Recall $X \sim \mathrm{Unif}(60, 90)$

- Like Expectation, variance is also a summary measure, where the expectation gives an idea of the central value, variance gives the idea how *dispersed the values are*.

Definition 3.15: (Variance)

If X is a continuous random variable with PDF $f(x)$, then the *Variance* of X is defined as

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}((X-\mu)^2) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx.$$

- The expectation and the variance of a continuous random variable can be calculated the same way we did for discrete, however, we need *Integration* here (DIY ✎ ✎ ✎)

1. **Random Variables**
   - Definitions, Discrete and Continuous RVs, and Prob. Distributions

2. **Discrete Random Variables**
   - 1. Probability Distribution and the idea of PMF
   - 2. CDF, quantiles and percentiles
   - 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$
   - 4. Rules of Expectation and Variance

3. **Parametric Distributions : Discrete**
   - 1. Discrete Uniform
   - 2. Bernoulli and Binomial Distribution

4. **Continuous Random Variables**
   - 1. Probability Distribution and the idea of PDF
   - 2. CDF, quantiles and percentiles
   - 3. Summary Measures of a Distribution - Expectation $\mathbb{E}(\cdot)$ and Variance $\mathbb{V}(\cdot)$

5. **Parametric Distributions : Continuous**
   - 1. Uniform Distribution
   - 2. Normal Distribution

Parametric Distributions : Continuous

Parametric Distributions : Continuous

1. Uniform Distribution

Uniform Distribution

- Again we will see some parametric distributions, but now for continuous random variables. Recall *parametric* means, there are one or two numbers in the PDF, that will specify the distributions, important is - *knowing the parameters of the distribution means we know everything related to the distribution.....* We have already seen the details of the Uniform Distribution, so I won't repeat here,
- But here we give the definition again

Uniform Distribution

Uniform Distribution

A continuous random variable X is said to follow the *Uniform Distribution* on the interval (a, b) if its PDF is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

We denote this by $X \sim \text{Unif}(a, b)$, where a and b are the parameters of the distribution and we have

$$\mathbb{E}(X) = \frac{a+b}{2}$$

$$\mathbb{V}\text{ar}(X) = \frac{(b-a)^2}{12}$$

- Recall, we can also calculate

$$\mathbb{P}(c < X < d) = \int_c^d f(x)\, dx = \frac{1}{b-a} \times (d-c)$$

Parametric Distributions : Continuous

2. Normal Distribution

Normal Distribution

- Now we will see possibly one of the most important continuous probability distributions of all time.... that is *Normal Distribution*

Normal Distribution

A continuous random variable X is said to follow the *Normal Distribution* with location parameter μ and dispersion parameter σ^2 if its PDF is

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

We denote this by $X \sim \mathcal{N}(\mu, \sigma^2)$, and we have

$$\mathbb{E}(X) = \mu$$
$$\mathbb{V}\mathrm{ar}(X) = \sigma^2$$

Normal Distribution

- When a random variable X is normally distributed then we write $X \sim \mathcal{N}(\mu, \sigma^2)$. Here μ and σ^2 are the *two parameters* of the distribution, which controls the shape of the density function $f(x)$. The density of the normal distribution looks like following.

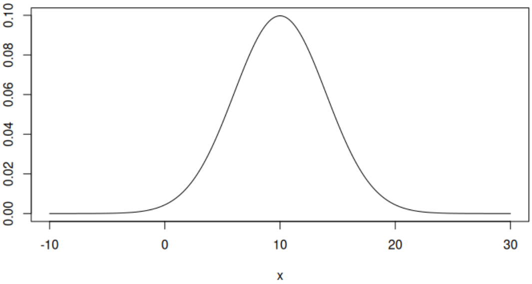


Figure 9: density function of a normal distribution when $\mu = 10$ and $\sigma^2 = 16$

- And the algebraic form of this function is

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

Normal Distribution

- Since we plotted the density in Figure 9 when $\mu = 10$ and $\sigma^2 = 16$, so this means the density function in Figure 8 is,

$$f(x) = \frac{1}{\sqrt{2\pi \times 16}} e^{-\frac{1}{2}\left(\frac{x-10}{4}\right)^2}$$

- The range of a normal distributed random variable is the whole real line or $\mathbb{R}$, so it takes values from ∞ to $+\infty$.
- μ is often called the *location* parameter and σ^2 is called the *dispersion* parameter, why this name? This is because If we change μ and σ^2, then we can shift the location of the density and also change the spread of the density.... Following picture will help to understand this ...

Normal Distribution

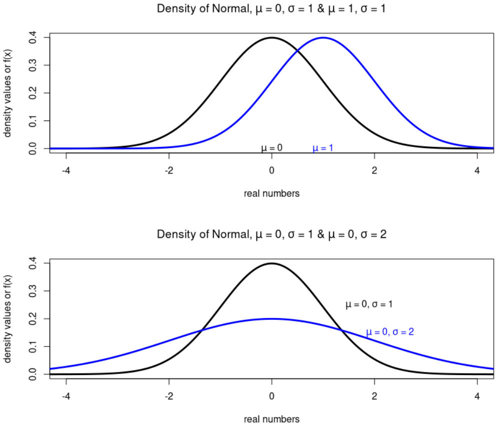


Figure 10: Effect of changing μ and σ^2 on the density function.

Normal Distribution

- So for each combination of μ and σ^2, we will get a different density function. Recall we can use the density function to calculate different probabilities. So you should keep in mind - *if parameters change then this means density changes and this then also means the probability distribution changes*.

Normal Distribution

- For example, following is a density function with parameters $\mu = 0$, $\sigma^2 = 1$, so the function is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. This normal distribution in particular is called *Standard Normal Distribution*, so in this case $X \sim \mathcal{N}(0, 1)$

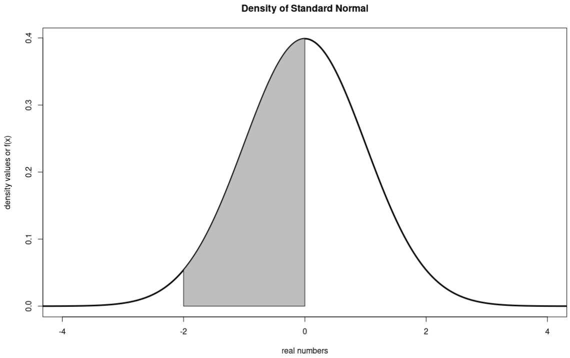


Figure 11: This is a density function for $X \sim \mathcal{N}(\mu, \sigma^2)$, so here $\mu = 0$, and $\sigma^2 = 1$. The shaded are is a probability, this is $\mathbb{P}(-2 < X < 0) = \int_{-2}^{0} f(x)dx = 0.4772499$

Normal Distribution

- We can use the density function to calculate the probabilities. Recall probability in this case is the area under the curve within some interval, right?
- Notice for each combination of μ and σ^2, we will get a different density function, this means different probability distribution (why?)

Normal Distribution

- Normal distribution has some amazing properties, even if you cannot remember the crazy looking density function, you should always remember these properties.

Some Properties of Normal Distribution

- **Knowing Mean and Variance:** If $X \sim \mathcal{N}(\mu, \sigma^2)$ then, $\mathbb{E}(X) = \mu$ and $\mathbb{V}\mathrm{ar}(X) = \sigma^2$. Notice always in the figure the Mean (or expected value) μ will be at the center of the Normal distribution.
- **3-σ Rule:** You should also remember (look at the figure below, taken from Anderson)

$$\begin{aligned}\mathbb{P}(\mu - \sigma < X < \mu + \sigma) &= 0.683\\ \mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma) &= 0.954\\ \mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma) &= 0.997\end{aligned}$$

- **Standardization Rule:** Finally if we know the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, then we can *transform* X to Z (or standardize) with

$$Z = \frac{X - \mu}{\sigma}$$

where $Z \sim \mathcal{N}(0, 1)$ (It's also possible to go from Z to X if you know μ and σ^2)

Normal Distribution

- First property is obvious...
- For the second property, look at the picture below,

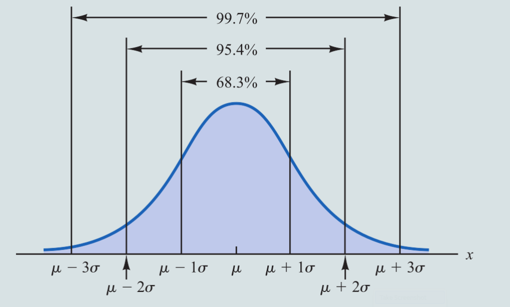


Figure 12: Picture taken from Anderson et al. (2020), this is called the famous $3 - \sigma$ rule, which says 68.3%, 95.4% and 99.7% of the values are within 1, 2 and 3 standard deviations from the mean μ.

- This means if we know the mean μ and variance σ^2, then we can figure out certain probabilities

Normal Distribution

- The last property is what we call *Z-transformation, or standardization or normalization*. All it says is -

  if we have a normal random variable with mean μ and variance σ^2, then we can transform this and get a standard normal random variable with mean 0 and variance 1. Here transformation means taking each of the values and then subtracting the mean μ and dividing by the standard deviation σ. This is also called standardization.

- This is a very useful rule because we can go back and forth from $\mathcal{N}(\mu, \sigma^2)$ to $\mathcal{N}(0, 1)$. In the back of your book you have a table of Standard Normal Distribution, we will do some problems then you will understand why this is useful.
- I think now we are ready to do some problems in Anderson et al. (2020), we will use the standard normal table at the back of the book

- **An Interesting point**: Interestingly we can do the standardization for sample data coming from any distribution, not necessary normal... so from $x_1, \ldots . x_n$, calculate

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

and then we can calculate the standardized value for x_i as

$$z_i = \frac{x_i - \bar{x}}{s_x}$$

where z_i is the standardized value for x_i, then we will also get

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = 0 \quad s_z^2 = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})^2 = 1$$

References

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J., and Ohlmann, J. W. (2020). *Statistics for Business & Economics*. Cengage, Boston, MA, 14th edition.

Bertsekas, D. and Tsitsiklis, J. N. (2008). *Introduction to probability*. Athena Scientific, 2nd edition.

Blitzstein, J. K. and Hwang, J. (2015). *Introduction to Probability*.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Thomson Learning, Australia ; Pacific Grove, CA, 2nd edition.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th edition.

Hansen, B. (2022). *Econometrics*. Princeton University Press, Princeton.

Newbold, P., Carlson, W. L., and Thorne, B. M. (2020). *Statistics for Business and Economics*. Pearson, Harlow, England, 9th, global edition.

Pishro-Nik, H. (2016). *Introduction to probability, statistics, and random processes*.

Ramachandran, K. M. and Tsokos, C. P. (2020). *Mathematical Statistics with Applications in R*. Elsevier, Philadelphia, 3rd edition.

Rice, J. A. (2007). *Mathematical Statistics and Data Analysis*. Duxbury advanced series. Thomson/Brooks/Cole, Belmont, CA, 3rd edition.