Ch3 - Probability Theory - 2

(Random Variables and Probability Distributions)

Statistics For Business and Economics - I

Shaikh Tanvir Hossain

East West University, Dhaka

Last Updated December 24, 2023

Outline

Outline

- 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
- Calculating Probabilities and Distributions

2. Prob. Dist: PMF, PDF, CDF, $\mathbb{E}(\cdot)$, $\mathbb{V}\mathrm{ar}(\cdot)$

- Idea of PMF
- Idea of PDF
- Cumulative Distribution Function - CDF
- Summary Measures - Expectation $\mathbb{E}(\cdot)$
- Summary Measures - Variance $\mathbb{V}\mathrm{ar}(\cdot)$

3. Parametric Family of Probability Distributions

- Discrete Distribution - Bernoulli and Binomial Distribution
- Discrete Distribution - Poisson Distribution
- Continuous Distribution - Normal Distribution
- Continuous Distribution - Exponential Distribution

4. Rules of Expectation and Variance

Random Variables

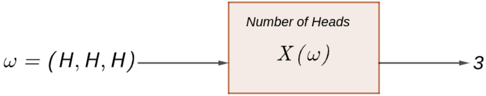
Random Variables

Definitions, Discrete and Continuous

- What is a random variable? Sometimes we are not interested in the experimental outcome directly, rather we are interested in some *function of the experimental outcome*. Here is an example.
- **Example 3.1**: (Three coin toss experiment) Suppose we are doing the experiment of tossing 3 independent coins. Then we know the experimental outcomes are (recall Ω is the sample space)

$$\Omega = \{(H, H, H), (H, H, T), (H, T, H), (T, H, H), \\ (T, T, H), (T, H, T), (H, T, T), (T, T, T)\}$$

- Now suppose we are NOT interested in these outcomes, rather we are interested in the *"total number of heads"* for each experimental outcome. For example, for the first outcome (H, H, H), the total number of heads is 3, for the second outcome (H, H, T), the total number of heads is 2, etc.
- We can represent this as function $X : \Omega \to \mathbb{R}$, where the domain is Ω and co-domain is $\mathbb{R}$.
- Here we can think $X(\omega)$ is a function where ω is the input that is coming from the sample space Ω, and the output will be total number of heads, which is going to be a number in $\mathbb{R}$.



- In this way we can think about all outcomes of this experiments and the value of the function $X(\omega)$

ω	(H, H, H)	(H, H, T)	(H, T, H)	(T, H, H)
$X(\omega)$	3	2	2	2

ω	(T, T, H)	(T, H, T)	(H, T, T)	(T, T, T)
$X(\omega)$	1	1	1	0

- *Here X is a random variable, and it's values are* $0, 1, 2, 3$.
- So a random variable is a function which takes input from the sample space Ω and gives us a number in the real line $\mathbb{R}$.

- Notice in the last example, the random variable *cannot take any values in* $\mathbb{R}$. It can only take 4 values, $0, 1, 2, 3$. So we can write the *range of the function* as $\mathcal{X} = \{0, 1, 2, 3\}$.
- Depending upon whether the range $\mathcal{X}$ is a *countable set or uncountable set*, we can classify the random variables into two types.
  - *Discrete Random Variable*: When the range of the random variable $\mathcal{X}$ is a countable set, for example $\mathcal{X} = \{0, 1, 2, 3\}$, we call it discrete random variable. Important $\mathcal{X} = \{0, 1, 2, 3, \ldots\}$, is also a countable set, we call it infinitely countable set.
  - *Continuous Random Variable*: When the range of the random variable $\mathcal{X}$ is an uncountable or infinite set, we call it a continuous random variable, for example $\mathcal{X} = [0, 1]$ or even $\mathcal{X} = \mathbb{R}$
- Now we can write the formal definition of a random variable,

Definition 3.2: (Random Variables, discrete and continuous)

A Random Variable $X(\omega)$ is a function, we write $X : \Omega \to \mathbb{R}$. If the range $\mathcal{X}$ of this function a countable set we call the random variable *a discrete random variable*, and if the range $\mathcal{X}$ is an uncountable set we call it *a continuous random variable*. Usually the random variables are denoted with uppercase letters X, Y, Z or W and often the input ω is suppressed.

- Again, recall, $X : \Omega \to \mathbb{R}$, means X is a function, Ω is the *domain* of the function, $\mathbb{R}$ is the *codomain*.
- Also as we mentioned before, we will denote $\mathcal{X}$ as the range of the function.

- In the last example, can you think about other random variables defined on the same sample space? ... (Ans: Yes we can, for example think about total number of tails, whether we have at least two heads, or whether we have at least one tail, etc.)
- In principle it is *possible to define many random variables* on the same sample space.
- As we wrote in the definition, we will use the uppercase letters such that X, Y, Z, etc., to denote the random variables. The lower case letters x, y, z, etc., will be used to denote the values of the random variables. For example it can be $X(\omega) = x$, this means if the experimental outcome is ω, the random variable X will take value x.
- 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*.
- Note that there is a difference between a mathematical function $f(x)$ (which is non-random) and a random variable $X(\omega)$ which is also a function?
- For a random variable the input is a random object, but when we are thinking function generally in mathematics we think about a fixed / non-random input (there is no experiment going on in the background!)

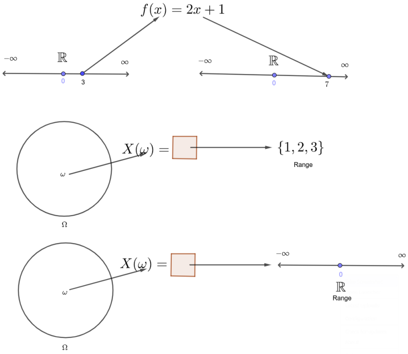


Figure 1: From the top, the first one is a mathematical function where the input is not random and the output is also not random (we often call this deterministic function). Here $f : \mathbb{R} \to \mathbb{R}$, and also the range is $\mathbb{R}$. The second one is a discrete random variable where the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

- Let's see an example of a continuous random variable (example taken from DeGroot and Schervish (2012))
- Suppose a contractor is building an office complex and needs to plan for water and electricity demand. After consulting with prospective tenants and examining historical data, the contractor decides that the demand for electricity will range somewhere between 1 million and 150 million kilowatt-hours per day and water demand will be between 4 and 200 (in thousands of gallons per day). All combinations of electrical and water demand are considered possible. In this case we have the following sample space

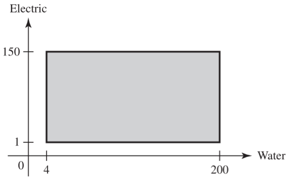


Figure 2: This is the sample space Ω of the experiment, notice the sample space is an infinite set

- The shaded region shows the sample space for the experiment, consisting of learning the actual water and electricity demands for the office complex.

- Here we can express the sample space as a Cartesian Product $[4, 200] \times [1, 150]$ or set of all ordered pairs (x, y) where $4 \leq x \leq 200$ and $1 \leq y \leq 150$,

$$\begin{aligned} \Omega &= [4, 200] \times [1, 150] \\ &= \{(x, y) : 4 \leq x \leq 200, 1 \leq y \leq 150\} \end{aligned}$$

- where x stands for water demand in thousands of gallons per day and y stands for the electric demand in millions of kilowatt-hours per day.
- We can also think about different events, which is going to be a subset of the sample space Ω.

$$\begin{aligned} A &= \{(x, y); x \geq 100, 1 \leq y \leq 150\} \text{ or maybe} \\ B &= \{(x, y) : 4 \leq x \leq 200, y \geq 115\} \end{aligned}$$

- A means water demand is *at least* 100,
- B means electricity demand *at least* 115.
- What if we write $C = \{(x, y) : x \geq 100, y \leq 35\}$, can you interpret this?
- This sample space has infinitely many points, and note carefully that the sample space is uncountable.

- Now we can think about different random variables defined here.
- For example, One random variable that is of interest is the demand for water. This can be expressed as $X(\omega) = x$ when $\omega = (x, y)$. The possible values of X are the numbers in the interval $[4, 200]$, so in this case the range is $\mathcal{X} = [4, 200]$. So this is a continuous random variable.
- Another random variable is Y where it's value will be the electricity demand, which can be expressed as $Y(\omega) = y$ when $\omega = (x, y)$. The possible values of Y are the numbers in the interval $[1, 150]$. Again this is also an example of continuous random variable.

- Another possible random variable could be Z, where Z is an *indicator of* whether or not at least one demand is high. Let A and B be the two events described before, that is, A is the event such that water demand is at least 100 , and B is the event such that electricity demand is at least 115 . Define

$$Z(\omega) = \begin{cases} 1 & \text{if } \omega \in A \cup B, \\ 0 & \text{if } \omega \notin A \cup B, \end{cases}$$

- The possible values of Z are the numbers 0 and 1 . The event $A \cup B$ is shown in the following figure, which represents the area of the sample space where at least one demand is high.

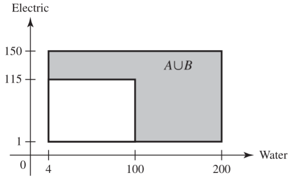


Figure 3: $A \cup B$ is the shaded area where water demand is at least 100 or electricity demand is at least 115

- 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	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

Random Variables

Calculating Probabilities and Distributions

- There is a very important point of thinking about random variables, that is

  Once we start thinking about random variables, now we have a new sample space $\mathbb{R}$, then we can think about different events in the new sample space $\mathbb{R}$, and forget about the original sample space Ω

- Since now we have a new sample space $\mathbb{R}$, we can think about different events in $\mathbb{R}$ (which are subsets of $\mathbb{R}$) and then maybe we can calculate probability in these events, for example

$$
\begin{aligned}
&\mathbb{P}([1, 1.5)) \\
&\mathbb{P}([4, 10000)) \\
&\mathbb{P}([1, 2))
\end{aligned}
$$

- Note that since $\mathbb{R}$ is sample space in this case, we will always have $\mathbb{P}(\mathbb{R}) = 1$.
- Also note if X is a random variable, for the probabilities we wrote above, we can write them as,

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

- *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!

- Now how do we calculate probabilities of these events or intervals in $\mathbb{R}$?
- Let's do one example, say we want to calculate the probability $\mathbb{P}([1, 1.5))$ for the random variable defined in Example 3.1.
- It turns out for this random variable in Example 3.1, if we know following probabilities,
  - $\mathbb{P}(\{0\}) = \mathbb{P}(X = 0)$ or probability that the random variable will take value 0 and
  - $\mathbb{P}(\{1\}) = \mathbb{P}(X = 1)$ or probability that the random variable will take value 1 and
  - $\mathbb{P}(\{2\}) = \mathbb{P}(X = 2)$ or probability that the random variable will take value 2 and
  - $\mathbb{P}(\{3\}) = \mathbb{P}(X = 3)$ or probability that the random variable will take value 3,
- then we can calculate probability of any interval in $\mathbb{R}$ for this random variable.
- Here is how we can do this, we can find the event associated with each value and then calculate the probability of that event using classical definition, for example

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

- Now applying the classical definition with equally likely assumption we get,

$$\mathbb{P}\left(\{(H, T, T), (T, H, T), (T, T, H)\}\right) = \frac{3}{8}$$

- So this means, now we know

$$\mathbb{P}(X = 1) = \mathbb{P}\left(\{(H, T, T), (T, H, T), (T, T, H)\}\right) = \frac{3}{8}$$

- With the same idea applied we can calculate other probabilities,

x	0	1	2	3
$\mathbb{P}(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

- Here $\mathbb{P}(X = x)$ means, probability of X taking values x, where x can be 0, 1, 2 and 3. Please try to calculate the other values.
- So with this approach we can actually get the Probabilities of a random variable X taking different values.
- And now we can also calculate the following probabilities
  - $\mathbb{P}(X \in [1, 1.5))$ or
  - $\mathbb{P}(X \in [4, 10000))$ or
  - $\mathbb{P}(X \in [1, 2))$ or
  - any other interval in $\mathbb{R}$
- How? Just use the probabilities where X actually takes its values, for example,

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

- Why $\frac{3}{8}$? The reason is in this interval X does not take values other than 1, so all the other points in this interval have probability 0.
- Can you do the other calculation in page 16? Yes, but we can do more
- In fact with these 4 probabilities, for this random variable we can calculate the probability of *any interval in* $\mathbb{R}$.
- Here for a random variable X, the probability distribution means how the probabilities are distributed on the real line $\mathbb{R}$.

Distribution of a random variable

Let X be a random variable. The distribution of X is the collection of all probabilities of the form $\mathbb{P}(X \in C)$ for all sets C of real numbers such that $\{X \in C\}$ is an event.

- 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* and for continuous random variable this function is known as *probability density function* or in short *PDF*.
- Next we will see the discussion for Discrete Probability Distributions and Continuous Probability Distributions and we will talk about PMF and PDF and their use in detail.

Prob. Dist: PMF, PDF, CDF, $\mathbb{E}(\cdot)$, $\mathbb{V}\mathrm{ar}(\cdot)$

Prob. Dist: PMF, PDF, CDF, $\mathbb{E}(\cdot)$, $\mathbb{V}\text{ar}(\cdot)$

Idea of PMF

- The distribution of a discrete random variable is known as *discrete distribution*.
- We already saw that to know the probability distribution of a discrete random variable it is 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.
- 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 3.3: (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 here is the PMF for the random variable that we have defined in Example 3.1,

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

- So for all possible values other than in the support the function has output 0. Recall, this is actually a *piecewise function* in math.
- Suppose if someone gives you this function, then you can calculate the probabilities of different intervals in $\mathbb{R}$.
- Can we plot this function? Yes we can. Here is the plot.

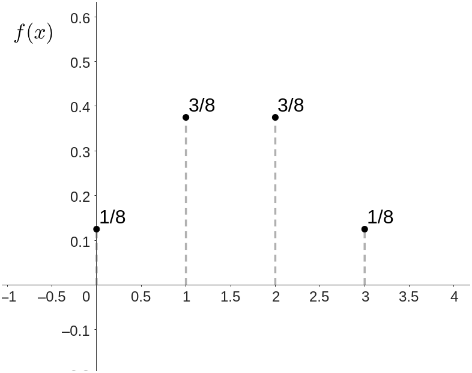


Figure 4: Plot of the PMF in (1)

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

Theorem 3.4: (Properties of PMF)

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

$$\sum_{i=1}^{\infty} f(x_i) = f(x_1) + f(x_2) + \ldots = 1$$

- **The probability of a subset in $\mathbb{R}$ can be calculated as the sum of PMF values in the set**: If A is any subset of the real line (for example any interval), then we can calculate $\mathbb{P}(X \in A)$ with

$$\mathbb{P}(X \in A) = \sum_{x_i \in A} f(x_i)$$

- The first one should be obvious, all it says if you sum the PMF functions value, then you should get 1.
- For example, in our 3 coin toss example, we have

$$f(0) + f(1) + f(2) + f(3) = 1/8 + 3/8 + 3/8 + 1/8 = 1$$

- 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 we will use PMF.
- Let's see a real life / empirical example of a PMF. Here the idea of the probability mass function is very similar to calculating *relative frequency*, but we need to calculate this using a population. Following example will illustrate this idea.

- **Example 3.5**: (Random Variable and PMF - Empirical Example from Anderson et al. (2020))
- Suppose our random variable X is the number of cars sold per day at DiCarlo Motors in Saratoga, New York and we know it can be $0, 1, 2, 3, 4$ or 5.
- Now over the past 300 days, DiCarlo has experienced -
  - 54 days with no (or 0) automobiles sold,
  - 117 days with 1 automobile sold,
  - 72 days with 2 automobiles sold,
  - 42 days with 3 automobiles sold,
  - 12 days with 4 automobiles sold and
  - 3 days with 5 automobiles sold.
- If we think this is our population then with this population we can write following probability mass function.

x	$f(x)$
0	$54/300 = .18$
1	$117/300 = .39$
2	$72/300 = .24$
3	$42/300 = .14$
4	$12/300 = .04$
5	$3/300 = .01$

- So the PMF is

$$f(x) = \begin{cases} .18 & \text{if } x = 0 \\ .39 & \text{if } x = 1 \\ .24 & \text{if } x = 2 \\ .14 & \text{if } x = 3 \\ .04 & \text{if } x = 4 \\ .01 & \text{if } x = 5 \\ 0 & \text{otherwise} \end{cases}$$

- Note the idea is very similar to calculating relative frequency! This way of thinking is helpful for understanding, but in reality often *we never know the population*, we actually have a sample.
- From a sample we can never get the probability mass function, what we can get is frequency distribution (this is what we did in chapter 1)
- So question is if we don't know the Population what is the solution? Ans - The idea is we will usually *assume the population is distributed according to some theoretical distribution* and then we will model the real life scenarios using some theoretical distribution.

- There are many theoretical discrete distributions, but we will learn 4 of them,
  - *Discrete Uniform Distribution*,
  - *Bernoulli distribution*,
  - *Binomial distribution* and
  - *Poisson distribution*.
- Now we will start with *Discrete Uniform Distribution*, and will talk about other distributions in coming sections.

- *Discrete Uniform Distribution* is very simple and you can think this is similar to the idea of equally likely outcomes but this is for the values of the random variable.

Definition 3.6: (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 $\{x_1, x_2, \ldots, x_k\}$. The PMF of X is

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

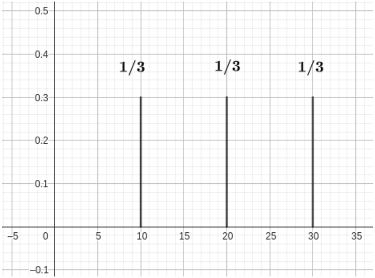
We write this as $X \sim \mathrm{DUnif}\{x_1, x_2, \ldots, x_k\}$.

- 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 the discrete 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 if we represent the random variable X as the amount of money you will get everyday, then X follows discrete uniform distribution with parameter $\{10, 20, 30\}$, we write $X \sim \mathrm{DUnif}\{10, 20, 30\}$ and the PMF of X is

$$\begin{aligned} f(10) &= 1/3, \\ f(20) &= 1/3 \\ \text{and } f(30) &= 1/3. \end{aligned}$$

- 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.

Prob. Dist: PMF, PDF, CDF, $\mathbb{E}(\cdot)$, $\mathbb{V}\mathrm{ar}(\cdot)$

Idea of PDF

- So we understood the idea of PMF, now let's talk about continuous distributions. How do we extend the idea of PMF 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?
- Following picture might be helpful to understand the idea what happens.
- Now let's talk about continuous distributions. 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,

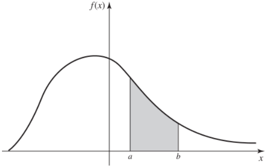


Figure 5: Here the shaded area is the probability of a random variable X taking value between a and b, so this means the shaded area is $\mathbb{P}(X \in [a, b]) = \int_a^b f(x)dx$

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

Definition 3.7: (Probability Density Function (PDF))

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

$$\mathbb{P}(X \in \mathcal{I}) = \int_{\mathcal{I}} f(x)dx$$

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

- For example, for interval $[a, b]$ we can calculate
$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx$$
- Similarly, $\mathbb{P}(X \geq a) = \int_a^{\infty} f(x)dx$ and $\mathbb{P}(X \leq b) = \int_{-\infty}^{b} f(x)dx$.
- We see that the density function f characterizes the distribution of a continuous random variable X in much the same way that the probability mass function characterizes the distribution of a discrete random variable.
- In this chapter we will see 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

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

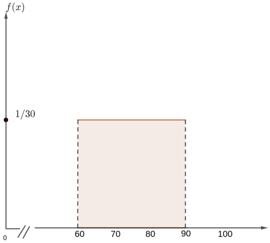
We denote this by $X \sim \text{Unif}(a, b)$.

- 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?

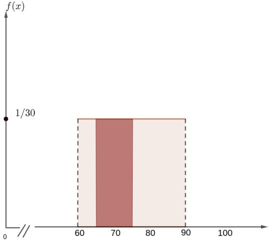


Probability Distributions
Idea of PDF

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

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

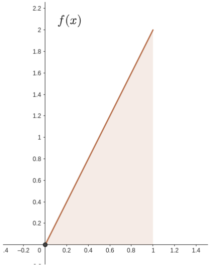
- So this means there is $1/3$ probability that the flight time will be between 65 minutes and 75 minutes.
- Note that this also shows probability is an area under the curve, in the following this is the shaded area,



- 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 $\mathbb{P}(X \in [70, 80]) = \frac{1}{3}$, $\mathbb{P}(X \in [75, 85]) = \frac{1}{3}$ and so on.

- Here is another example, suppose we have a random variable X takes value in $[0,1]$, so $\mathcal{X} = [0,1]$, 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 $\mathbb{P}(X \in [0.5, 0.7]) =?$ Yes we can...

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

- So now we know $\mathbb{P}(X \in [0.5, 0.7]) = 0.24$
- 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 $\mathbb{P}(X \in [a,b])$?

Prob. Dist: PMF, PDF, CDF, $\mathbb{E}(\cdot)$, $\mathbb{V}\mathrm{ar}(\cdot)$

Cumulative Distribution Function - CDF

- If we know PMF or PDF of a random variable, we can also calculate *cumulative probabilities*. Cumulative Probability means probabilities upto a certain value.
- For example, $\mathbb{P}(X \leq 2)$ is a cumulative probability, it means for the random variable X what is the probability of taking value 2 or less than 2.
- For the PMF given in page 25, we can calculate

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

- Often cumulative probabilities are represented using a function called *cumulative distribution function* in short CDF. CDF is a function defined on the real line, where for any value x, it gives the cumulative probability upto that value.

Definition 3.8: (The cumulative distribution function (CDF))

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

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

- So the function just gives the cumulative probabilities for each $x \in \mathbb{R}$. For example, using the PMF in page 29, 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,

Probability Distributions

CDF for Discrete R.V.

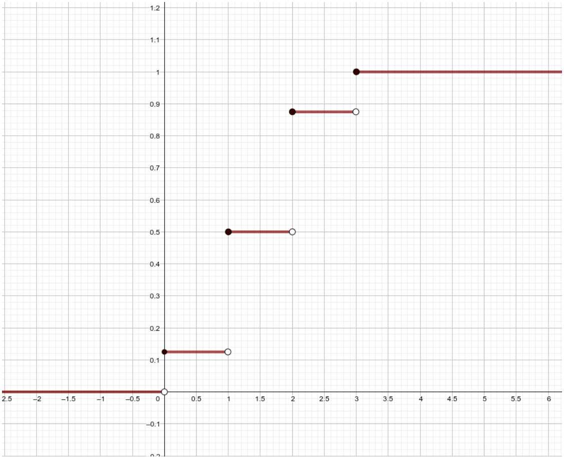


Figure 6: Notice there is a jump when the random variable takes its value, and the difference where there is a jump is the probability. Also note that the CDF function is defined for the whole $\mathbb{R}$

The last figure can be written as a piecewise function

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

There are 3 important properties for the CDF,

- 1) **Always Non-Decreasing:** If $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
- 2) **Right Continuous:** $F(a) = \lim_{x \to a^+} F(x)$.
- 3) **At infinity the limits are** 0 **and** 1 $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

- 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 *

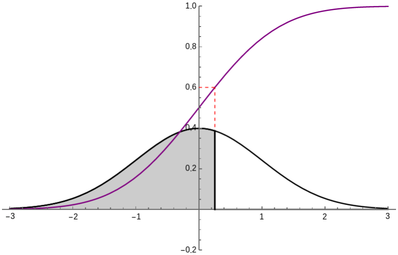


Figure 7: The density function $f(x)$ is the Bell-Shaped curve, the shaded are is $\mathbb{P}(X < .253) = \mathbb{P}(X \in (-\infty, .253)) = .60$. The function in the purple color is the cumulative distribution function (CDF) $F(x)$.

- 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/

Prob. Dist: PMF, PDF, CDF, $\mathbb{E}(\cdot)$, $\mathbb{V}\mathrm{ar}(\cdot)$

Summary Measures - Expectation $\mathbb{E}(\cdot)$

- Calculating Expected Value and Expectation is very easy, let's first see the definition of Expectation and we will do an example right away!

Definition 3.9: (Expected Value)

If X is a *discrete random variable* with PMF $f(x)$ then the *Expectation* or the *Expected Value* of X is defined as

$$\mathbb{E}(X) = \sum_{\text{all values } x} x f(x)$$

If X is a *continuous random variable* whose PDF is $f(x)$ then it is defined as follows:

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

We will usually use the notation $\mathbb{E}(\cdot)$ to denote that we are performing *Expectation* on X.

- So let's break it down ... For a discrete random variable, if we know the PMF, then calculating Expected value is not difficult, it's just multiplying x with $f(x)$ and then adding them altogether. Let's calculate the Expectation for the random variable X from Example 3.1.
- Recall the PMF of Example 3.1,

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

- So 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 like average, but it is for a population, so it's a *population average* we call it also *population mean*, let's explain how....

- We will use the Example 3.5 again to understand the idea of Expectation. Suppose here is the population data (you can think about an excel file with 300 rows)

$$\underbrace{0,0,0,0,0,0,0,0,\ldots,0}_{\text{54 days so 54 rows}},\underbrace{1,1,1,1,1,1,\ldots,1}_{\text{117 days so 117 rows}},\underbrace{2,2,2,2,2,2,\ldots,2}_{\text{72 days so 72 rows}},\underbrace{3,3,3,3,3,3,\ldots,3}_{\text{42 days so 42 rows}},$$
$$\underbrace{4,4,4,4,4,4,\ldots,4}_{\text{12 days so 12 rows}},\underbrace{5,5,5}_{\text{3 days so 3 rows}}$$

- If someone asks what is the average number of selling in last 300 days? You might take the average of these numbers and if you do this then the average will be 1.5.
- But now using the PMF in page 29, if we calculate the expected value, we will get the same answer,

$$(0\times f(0))+(1\times f(1))+(2\times f(2))+(3\times f(3))+(4\times f(4))+(5\times f(5))$$
$$=(0\times 54/300)+(1\times 117/300)+(2\times 72/300)+(3\times 42/300)+(4\times 12/300)+(5\times 3/300$$
$$=1.5$$

- So the Expected Value or Expectation is actually a population average. It's just we are not doing the average directly, rather we are weighting values with their probabilities.
- You may ask why we are learning this formula? What's the benefit?
- There are at least three reasons why?

- First, we don't have to take average of huge number of values, rather we can just use the PMF and calculate the expectation.
- Second, we can calculate the expectation even if we don't have the population data but only know the PMF.
- Third, we can extend this idea to continuous case, the idea is replace the $\sum$ with $\int$

- Homework: Can you calculate the expected value of the Discrete Uniform in page 35, where $X \sim \mathrm{DUnif}\{10, 20, 30\}$
- So *Expectation* is the population average, and roughly you can say it's a *central value* such that more probabilities (or weights) are near this value. This is why it is often called a measure for *central tendency*!
- Notice for the expected value we calculated 1.5 (page 54, coin toss example) is between 1 and 2, and 1 and 2 are the two points where the probability is highest for this random variable.

- Now let's see how to calculate the expected value of a continuous random variable X.
- We do it for the random variable given in page 42, 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 $X \sim \text{Unif}(a, b)$

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

- 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 $X \sim \text{Unif}(60, 90)$
- **A Notational Remark:** You should always think about following picture if you think about Expectation of a random variable X. Often the constant that we get after calculating the expectation is denoted by μ. So from now on, if you see μ (although sometimes this is a bit misleading because of normal random variable, we will see it later, but it's been used), you should understand this is an expected value like 1.5.

Expectation operates on random variable, not on a fixed number

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

is a random variable, has many possible values

The result of the expectation will be a fixed number,
for example it was 1.5 for the random variable which counts number of heads.

- Finally always remember *Population Mean, Expectation, Expected Value*, they are all same thing!

Prob. Dist: PMF, PDF, CDF, $\mathbb{E}(\cdot)$, $\mathbb{V}\text{ar}(\cdot)$

Summary Measures - Variance $\mathbb{V}\text{ar}(\cdot)$

- 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 3.10: (Variance)

If X is a discrete random variable with PMF $f(x)$ then the *Variance* of X is defined as

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left(\left(X - \mathbb{E}\left(X\right)\right)^2\right) = \mathbb{E}\left(\left(X - \mu\right)^2\right) = \sum_{\text{all values } x} (x-\mu)^2 f(x)$$

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

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

- 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 variance is the Expectation of the squared deviation.

- Let's calculate $\mathbb{V}\mathrm{ar}(X)$ for the random variable X, which counts the number of heads (PMF in page 27).

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

- 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.
- **A Notational remark:** Often the constant that we get after calculating the variance is denoted by σ^2. So from now on, if you see σ^2, you should understand this is a variance like .75.
- The square root of the variance is called *Standard Deviation*, so if σ^2 is the Variance, then σ is the standard deviation.
- Like discrete random variables we can also calculate Expected values and Variance for a continuous random variable.
- 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 ✍ ✍ ✍)

Parametric Family of Probability Distributions

Parametric Family of Probability Distributions

Discrete Distribution - Bernoulli and Binomial Distribution

Bernoulli & Binomial Random Variables

- If a random variable X has only two values 0 and 1, we call the random variable a *Bernoulli Random variable*, and its distribution is known as *Bernoulli distribution*, some examples,
  - Random Experiment - Toss a coin. Random Variable $X = 1$ means head, $X = 0$ means tail.
  - Random Experiment - Pick a random person from a Population of people. Random Variable $X = 1$ means female, and $X = 0$ means male
  - Random Experiment - Calling a customer. Random Variable $X = 1$ means picked up the call, $X = 0$ means didn't pickup.
- 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 a 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)

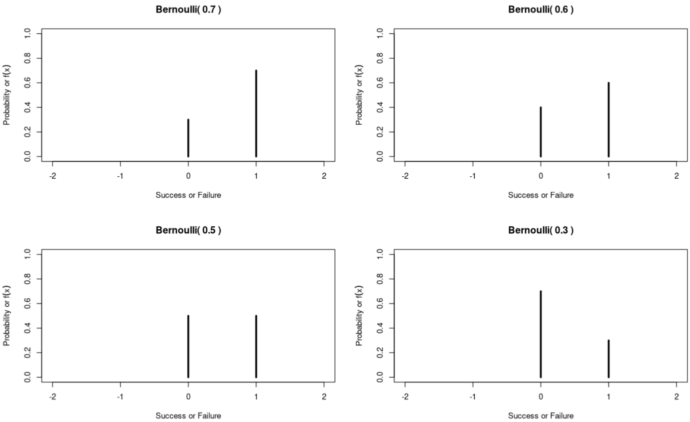
Definition 3.12: (Bernoulli distribution). A random variable X is said to have the Bernoulli distribution with parameter p (we wrote $X \sim \text{Bernoulli}(p)$), if $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$. The PMF of X can be written as,

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

where $0 < p < 1$

- 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 $X \sim \text{Bernoulli}(.3)$, 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 $\mathbb{V}\text{ar}(X) = p(1-p)$ (please do the calculation!)

Bernoulli & Binomial Random Variables

- The Binomial distribution comes when we perform *n independent Bernoulli experiment*.
- Here is the story - Suppose now we perform n independent Bernoulli experiments (or Bernoulli trials) 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, if we toss a single coin and represent X as a Bernoulli random variable which can take values 1 (success) or 0 (failure) with probability p, then when we toss $n = 3$ independent coins and represent X as total number of heads (total number of success), then X is said to follow Binomial with parameter $n = 3$ and probability p.

Definition 3.13: (Binomial Distribution)

Suppose that n independent Bernoulli trials are performed, each with the same success probability p. Let X be the number of successes. The distribution of X is called the Binomial distribution with parameters n and p. 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 $0 < p < 1$. The PMF of X can be written as

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

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 $\left(\begin{array}{c} n \\ x \end{array} \right) = {}^nC_x = \frac{n!}{x!(n-x)!}$. 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 ${}^nC_x$ 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!)

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

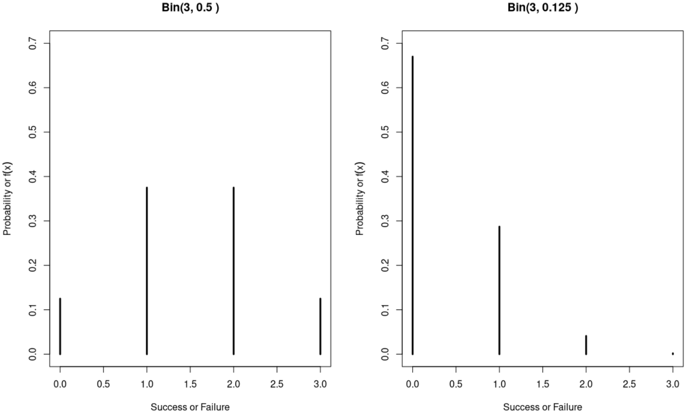


Figure 8: On the left we have the PMF of $X \sim \text{Bin}(3, 0.5)$ and on the right we have $X \sim \text{Bin}(3, 0.125)$.

Bernoulli & Binomial Random Variables

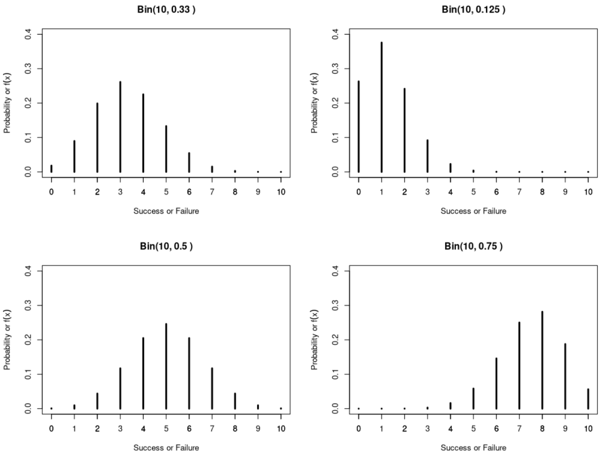


Figure 9: From top left, we have the PMF of $X \sim \mathrm{Bin}(10, 0.33)$, then right $X \sim \mathrm{Bin}(10, 0.125)$, then bottom left $X \sim \mathrm{Bin}(10, 0.5)$ and right $X \sim \mathrm{Bin}(10, 0.75)$

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.

Parametric Family of Probability Distributions

Discrete Distribution - Poisson Distribution

Poisson Distribution

- Now we will consider another distribution known as Poisson distribution.
- The situation when you can use Poisson distribution is very similar to Binomial distribution, but the difference is in Binomial you have often a small number of trials, but in Poisson you may have many (in theory infinite) independent trials, with success probability very small.
- For example, the number of arrivals at a car wash in one hour, the number of repairs needed in 10 miles of highway, or the number of leaks in 100 miles of pipeline (these examples are from Anderson et al. (2020)). Notice, n can be very large, often there is no upper limit, and also the values x also has no upper limit, so it can be $0, 1, 2, 3, \ldots$.

Definition 3.14: (Poisson Distribution)

An random variable X has the Poisson distribution with parameter λ if the PMF of X is

$$f(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

We write this as $X \sim \text{Pois}(\lambda)$.

- This is a valid PMF because we can show that $\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$.

Poisson Distribution

- How does Poisson PMF come? The idea is we can think Poisson distribution is a *limiting case* (or limit) of Binomial distribution. More specifically we need $n \to \infty$ and $p \to 0$, then Binomial PMF will converge to Poisson PMF.
- This means starting from $\left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x}$ and we take $n \to \infty$ and $p \to 0$ we get $\frac{e^{-\lambda}\lambda^x}{x!}$
- It's a mathematical problem with limit, so we will avoid it, but you should understand that Poisson distribution appears as a limit of Binomial.

Poisson Distribution

- In real life context it is often used in situations where we are counting the number of successes in a particular region or interval of time, and there are a large number of trials, each with a small probability of success. The random variable X is again the number of success and but in this case number of success can be $0, 1, 2, 3, \ldots$.
- Here are some more examples,
  - The number of emails you receive in an hour. There are a lot of people who could potentially email you in that hour, but it is unlikely that any specific person will actually email you in that hour.
  - The number of earthquakes in a year in some region of the world. At any given time and location, the probability of an earthquake is small, but there are a large number of possible times and locations for earthquakes to occur over the course of the year.
- There are more example in Anderson et al. (2020).
- If $X \sim \mathrm{Pois}(\lambda)$, then we can again calculate (skip the proof)

$$\mathbb{E}(X) = \lambda \quad \text{and} \quad \mathbb{V}\mathrm{ar}(X) = \lambda$$

- Here λ is the parameter, which is also the mean or expected value. It is often it is called the *rate of occurrence* in the rare events.

Poisson Distribution

- There are other discrete distributions, e.g., Geometric and Negative Binomial, which we won't cover here. If you are interested to read about them you can read DeGroot and Schervish (2012).

Parametric Family of Probability Distributions

Continuous Distribution - Normal Distribution

Normal Distribution

- We already saw one continuous distribution, which is Uniform distribution, now we will see another one, known as *Normal distribution*.
- *Normal distribution* is possibly one of the most important continuous probability distributions of all time.
- 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.

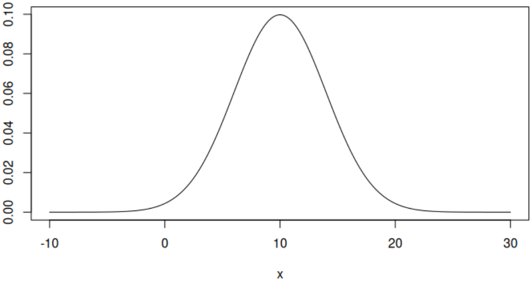


Figure 10: density function of a normal distribution when $\mu = 10$ and $\sigma^2 = 16$

Normal Distribution

- What is the algebraic form of this function? Here it is
$$f\left(x;\mu,\sigma^2\right)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$
- Since we plotted the density in Figure 10 when $\mu=10$ and $\sigma^2=16$, so this means the density function in Figure 8 is,
$$f(x;10,16)=\frac{1}{\sqrt{2\pi\times 16}}e^{-\frac{1}{2}\left(\frac{x-10}{4}\right)^2}$$
- The range of a normal distributed random variable is the whole real line or $\mathbb{R}$. This means it takes values from ∞ to $+\infty$.
- μ is often called the *location* parameter and σ^2 is called the *dispersion* parameter (why this name, we will see in a minute)
- Again notice this is a function of x, but we also wrote the two parameters μ,σ^2 after ";", always remember when we think about a density function now it's a function of x (this is similar to PMF) but we will write parameters to show that these parameters controls the function.
- 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?
- For example, following is a density function with parameters $\mu=0$, $\sigma^2=1$, so the function is $f(x;0,1)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$.

Normal Distribution

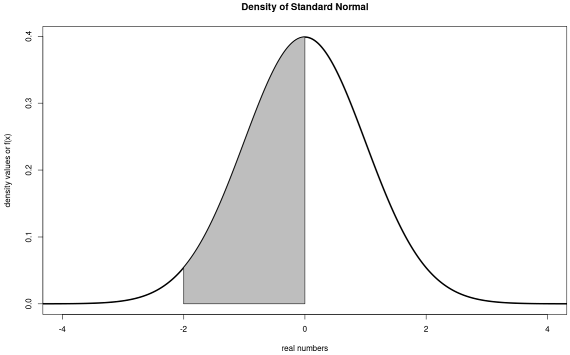


Figure 11: This is a density function of a normal distribution with $\mu = 0$, and $\sigma^2 = 1$. The shaded are is a probability, this is $\mathbb{P}(X \in (-2, 0)) = \int_{-2}^{0} f(x; 0, 1)dx = 0.4772499$

- Notice for each combination of μ and σ^2, we will get a different density function, this means different probability distribution (why?.
- Why we call them *location* and *dispersion* parameter. 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.

Normal Distribution

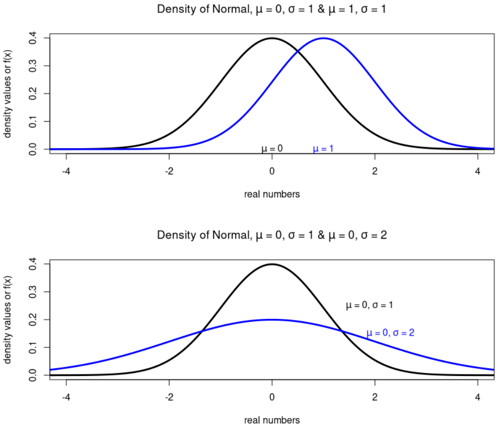


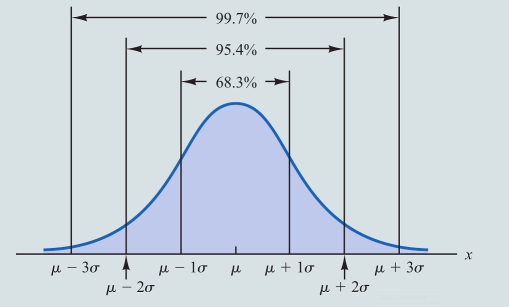
Figure 12: Effect of changing μ and σ^2 on the density function.

Normal Distribution

- So for each combination of μ and σ^2, we will get a different density function. Recall we use the density function to calculate different probabilities. So density change is equivalent to distribution change.

Normal Distribution

- Normal distribution has some amazing properties, even if you cannot remember the crazy looking density function, you should always remember these properties.
  - If $X \sim \mathcal{N}(\mu, \sigma^2)$ We can calculate the Expected value and Variance. We will get $\mathbb{E}(X) = \mu$ and $\mathbb{V}\text{ar}(X) = \sigma^2$.
  - Notice the figure the Mean (or expected value) will be always at the center of the Normal distribution.
  - Then you should remember following picture (this is taken from Anderson et al. (2020))



- This means if we know the mean μ and variance σ^2, then we can figure out
  - the two points $\mu + \sigma$, $\mu - \sigma$, and we know that within these two points there will be 68.3% probability.

Normal Distribution

- the two points $\mu + 2\sigma$, $\mu - 2\sigma$, and we know that within these two points there will be 95.4% probability.
- the two points $\mu + 3\sigma$, $\mu - 3\sigma$, and we know that within these two points there will be 99.7% probability.

- Finally if we know the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, then we can transform this random variable and get a new random variable Z, by

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

- This what we call *Z-transformation, or standardization or normalization*.
- The interesting thing is it can be proven that $Z \sim \mathcal{N}(0, 1)$.
- This is sometimes very helpful because we can go back and forth from $\mathcal{N}(\mu, \sigma^2)$ to $\mathcal{N}(0, 1)$.
- 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

Parametric Family of Probability Distributions

Continuous Distribution - Exponential Distribution

Exponential Distribution

Rules of Expectation and Variance

Rules of Expectation and Variance

- We already saw the way to calculate the Expected value[†] of a random variable $\mathbb{E}(X)$, when X is discrete or continuous. Here are the two formulas again

$$\mathbb{E}(X) = \sum_{\text{all values } x} xf(x) \quad \text{[when } X \text{ is discrete]}$$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx \quad \text{[when } X \text{ is continuous]}$$

- The Variance is also an expectation, but it's the expectation of the squared deviation from mean.

$$\mathbb{V}\text{ar}(X) = \mathbb{E}\left(\left(X - \mathbb{E}(X)\right)^2\right) = \sum_{\text{all values } x} (x - \mathbb{E}(X))^2 f(x) \quad \text{[when } X \text{ is discrete]}$$

$$\mathbb{V}\text{ar}(X) = \mathbb{E}\left(\left(X - \mathbb{E}(X)\right)^2\right) = \int_{-\infty}^{\infty} (x - \mathbb{E}(X))^2 f(x)dx \quad \text{[when } X \text{ is continuous]}$$

- Always remember! The expected value of a random variable is a constant, so $\mathbb{E}(X) =$ constant number . Sometimes we write this constant number with μ regardless of the fact that X follows normal distribution or not. This is not a good notation, but people generally use it.

- Again 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.

Rules of Expectation and Variance

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

Expected Value or Mean is a constant

*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.*

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

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

- Note that X^2, X^3 or $2X+3$ are all functions of random variable X.

Rules of Expectation and Variance

- So our question is for a function $g(X)$, where $g(X)$ can be X^2, X^3 or $2X+3$, what is $\mathbb{E}(g(X))$?
- First of all note that $g(X)$ is also a random variable. So a function of a random variable is also a random variable.
- It turns out that (we are avoiding the formal proof) in this case we can calculate $\mathbb{E}(g(X))$ by using the distribution of X.
- So this means ....

$$\mathbb{E}(g(X)) = \sum_{\text{all values } x} g(x)f(x) \quad \text{[when } X \text{ is discrete]}$$

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx \quad \text{[when } X \text{ is continuous]}$$

- Now we can apply this rule more or less for any function $g()$. This rule has an interesting name, it is called - *Law of the unconscious Statistician* or in short *LOTUS* (why this name?)
- We have already applied LOTUS previously,
  - Calculating variance is an application of this rule, because we are doing $\mathbb{E}\left((X-\mu)^2\right)$. Here $g(X) = (X-\mu)^2$.
  - Also in PS-3 when we calculated $\mathbb{E}\left(X^2\right)$, we have applied LOTUS, here $g(X) = X^2$.
  - You can create more examples, but application of LOTUS is easy!
- So LOTUS is one rule for expectation, there are two other rules for expectation, and we can use LOTUS to prove the following rules,
  - $\mathbb{E}(a) = a$, where a is any constant.

Rules of Expectation and Variance

- $\mathbb{E}(bX) = b\mathbb{E}(X)$, where b is any constant.
- So together we have $\mathbb{E}(a + bX) = a + b\mathbb{E}(X)$

- The last rule is sometimes called the *linearity of expectation*.
- We will see the proof for a discrete random variable X with values $x_1, x_2, x_3, \ldots, x_n$.

$$\begin{aligned}
\mathbb{E}[a + bX] &= (a + bx_1)\, f(x_1) + (a + bx_2)\, f(x_2) + \ldots, + (a + bx_n)\, f(x_n) && (3)\\
&= \sum_{j=1}^{n} (a + bx_i)\, f(x_i) && (4)\\
&= \sum_{j=1}^{n} af(x_i) + \sum_{j=1}^{n} (bx_i)\, f(x_i) && (5)\\
&= a \sum_{j=1}^{n} f(x_i) + b \sum_{j=1}^{n} x_i f(x_i) && (6)\\
&= a + b\mathbb{E}[X] && (7)
\end{aligned}$$

- At (4) we applied the formula for expectation
- At (5) we just wrote with summation
- At (6) summation can be separated
- At (7) take the constant out from summation (this is a rule for summation, see below)
- At (8) $\sum_{j=1}^{n} f(x_i) = 1$ because this is a PMF, and $\sum_{j=1}^{n} x_i f(x_i) = \mathbb{E}(X)$

- When it comes to summation you can always apply following rules,

Rules of Expectation and Variance

- 1. $\sum_{i=1}^{n} a = na$, where a is constant. For example, $\sum_{i=1}^{4} 3 = 4 \cdot 3 = 12$.
- 2. $\sum_{i=1}^{n} bx_i = b \sum_{i=1}^{n} x_i$, where b is a constant.
- 3. $\sum_{i=1}^{n} (a + bx_i) = na + b \sum_{i=1}^{n} x_i$, where a and b are constants and where used property 1 and 2 above.
- 4. $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$.
- Even if you don't understand the proof, it is ok, you need to understand what does *"linearity of Expectation mean"*
- You can think about linearity this way - *if we add first then take the expectation, the expected value will be equal to taking expectation first and then adding the expectations*.
- So applying linearity we get,

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

- Where we see that Expectation of a constant is always constant.
- if constant is multiplied we can 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, so if we have 2 (or more) random variables (*even if they are independent or not*), applying linearity

$$\mathbb{E}(X_1 + X_2) = \mathbb{E}(X_1) + \mathbb{E}(X_2)$$

Rules of Expectation and Variance

- Using the linearity of Expectation and the rules above we can show

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

- This is an alternative formula to calculate the variance. Let's see this on board! (this formula should be familiar)
- Because Variance is also an expectation, we can also get different rules for Variance.
- Here we will learn only two rules for variance,
  - $\mathbb{V}\mathrm{ar}(a) = 0$, where a is any constant.
  - $\mathbb{V}\mathrm{ar}(bX) = b^2\mathbb{V}\mathrm{ar}(X)$
  - Then together for the functions like $a + bX$, we have $\mathbb{V}\mathrm{ar}(a + bX) = b^2\mathbb{V}\mathrm{ar}(X)$
- Note that from the last calculation you might conclude that - like expectation we also have *linearity of variance*. But this is wrong in general. Here we have a special case, so it looks like linearity of variance, but variance calculation in general is not linear (we will talk about this in detail in the next chapter!).
- So in general for any two random variables, we have

$$\mathbb{V}\mathrm{ar}(X_1 + X_2) \neq \mathbb{V}\mathrm{ar}(X_1) + \mathbb{V}\mathrm{ar}(X_2)$$

- Later we will see that this holds only for independent random variables. So if X_1 and X_2 are independent random variables, only then we can apply linearity of Variance, but if we don't know this information, we cannot.

Rules of Expectation and Variance

- Here is the summary of what we have learned so far regarding the rules of Expectation and Variance.

Rules of Expectation and Variance

Let X be a random variable, then

- 1. [LOTUS] We can calculate $\mathbb{E}(g(X))$ using

$$\mathbb{E}(g(X)) = \sum_{\text{all values } x} g(x)f(x) \quad \text{[when } X \text{ is discrete]}$$

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx \quad \text{[when } X \text{ is continuous]}$$

- 2. For a function $g(X) = a + bX$ (we call this a linear function of X), we have

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

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

- Be careful: Linear function of X and linearity of expectation are two different things, we call the function $a + bX$ linear function because the power of X is 1.

† Recall we call it also mean or expectation.

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.

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