CS 107, Probability, Spring 2019
Lecture 38

Michael Poghosyan

AUA

26 April 2019

Content

- The Expected Value of a R.V.

Partial Numerical Characteristics of R.V.s: Expectation of a R.V

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs.

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs. Unfortunately, in real life, these complete characteristics are not given.

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs. Unfortunately, in real life, these complete characteristics are not given. Say, if X is measuring the

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs. Unfortunately, in real life, these complete characteristics are not given. Say, if X is measuring the

- the height of a (random) person;

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs. Unfortunately, in real life, these complete characteristics are not given. Say, if X is measuring the

- the height of a (random) person;
- the insurance claim size for today;

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs. Unfortunately, in real life, these complete characteristics are not given. Say, if X is measuring the

- the height of a (random) person;
- the insurance claim size for today;
- the GPA of a (random) student at AUA;

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs. Unfortunately, in real life, these complete characteristics are not given. Say, if X is measuring the

- the height of a (random) person;
- the insurance claim size for today;
- the GPA of a (random) student at AUA;
- the AMZN stock closing price today,

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs. Unfortunately, in real life, these complete characteristics are not given. Say, if X is measuring the

- the height of a (random) person;
- the insurance claim size for today;
- the GPA of a (random) student at AUA;
- the AMZN stock closing price today,
- the no. of GG Taxi calls or FB page visitors,

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs. Unfortunately, in real life, these complete characteristics are not given. Say, if X is measuring the

- the height of a (random) person;
- the insurance claim size for today;
- the GPA of a (random) student at AUA;
- the AMZN stock closing price today,
- the no. of GG Taxi calls or FB page visitors,

then nobody will give us the CDF/PDF/PMF of X.

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs.
Unfortunately, in real life, these complete characteristics are not given. Say, if X is measuring the

- the height of a (random) person;
- the insurance claim size for today;
- the GPA of a (random) student at AUA;
- the AMZN stock closing price today,
- the no. of GG Taxi calls or FB page visitors,

then nobody will give us the CDF/PDF/PMF of X.
So we want to give, at least, some partial information/characteristics for these r.v.s.

Partial Characteristics of R.V.s: Motivation

So far we have considered Complete Characteristics of r.v.s: we were studying r.v.s given by their CDFs or PMFs/PDFs. Unfortunately, in real life, these complete characteristics are not given. Say, if X is measuring the

- the height of a (random) person;
- the insurance claim size for today;
- the GPA of a (random) student at AUA;
- the AMZN stock closing price today,
- the no. of GG Taxi calls or FB page visitors,

then nobody will give us the CDF/PDF/PMF of X.
So we want to give, at least, some partial information/characteristics for these r.v.s.
The most important 2 characteristics are the **Expected Value (Expectation)** and the **Variance/Standard Deviation** of a r.v.

Expectation of a R.V.

First, let's consider our previous example: X is a r.v. showing

- the height of a (random) person;
- the insurance claim size for today;
- the GPA of a (random) student at AUA;
- the AMZN stock closing price today;
- the no. of GG Taxi calls or FB page visitors;
- the running time of Quicksort for a (random) input;

Expectation of a R.V.

First, let's consider our previous example: X is a r.v. showing

- the height of a (random) person;
- the insurance claim size for today;
- the GPA of a (random) student at AUA;
- the AMZN stock closing price today;
- the no. of GG Taxi calls or FB page visitors;
- the running time of Quicksort for a (random) input;

Then maybe the most important information is the Average Value, or the Expectation of X:

Expectation of a R.V.

First, let's consider our previous example: X is a r.v. showing

- the height of a (random) person;
- the insurance claim size for today;
- the GPA of a (random) student at AUA;
- the AMZN stock closing price today;
- the no. of GG Taxi calls or FB page visitors;
- the running time of Quicksort for a (random) input;

Then maybe the most important information is the Average Value, or the Expectation of X: say, the average Height of a person, the average claim size or the average GPA etc.

Expectation of a R.V.

First, let's consider our previous example: X is a r.v. showing

- the height of a (random) person;
- the insurance claim size for today;
- the GPA of a (random) student at AUA;
- the AMZN stock closing price today;
- the no. of GG Taxi calls or FB page visitors;
- the running time of Quicksort for a (random) input;

Then maybe the most important information is the Average Value, or the Expectation of X: say, the average Height of a person, the average claim size or the average GPA etc.
So the first of our Partial Characteristics will be the **Expectation**, or the Average, Mean Value of a r.v.

Expectation of a R.V.

We give the Definition of the Expected Value separately for the Discrete and Continuous r.v.s.

Expectation of a R.V.

We give the Definition of the Expected Value separately for the Discrete and Continuous r.v.s.

Expected Value of a Discrete R.V.

Assume X is a Discrete r.v. with the PMF

Values of X	x_1	x_2	...
$\mathbb{P}(X=x)$	p_1	p_2	...

Expectation of a R.V.

We give the Definition of the Expected Value separately for the Discrete and Continuous r.v.s.

Expected Value of a Discrete R.V.

Assume X is a Discrete r.v. with the PMF

Values of X	x_1	x_2	...
$\mathbb{P}(X = x)$	p_1	p_2	...

Then the **Expected Value** or the **Expectation** of X is the following number:

$$\mathbb{E}(X) = \sum_i x_i \cdot p_i = \sum_i x_i \cdot \mathbb{P}(X = x_i),$$

Expectation of a R.V.

We give the Definition of the Expected Value separately for the Discrete and Continuous r.v.s.

Expected Value of a Discrete R.V.

Assume X is a Discrete r.v. with the PMF

Values of X	x_1	x_2	...
$\mathbb{P}(X=x)$	p_1	p_2	...

Then the **Expected Value** or the **Expectation** of X is the following number:

$$\mathbb{E}(X) = \sum_i x_i \cdot p_i = \sum_i x_i \cdot \mathbb{P}(X = x_i),$$

if $\sum_i |x_i| \cdot p_i < +\infty$

Expectation of a R.V.

We give the Definition of the Expected Value separately for the Discrete and Continuous r.v.s.

Expected Value of a Discrete R.V.

Assume X is a Discrete r.v. with the PMF

Values of X	x_1	x_2	$\dots$
$\mathbb{P}(X=x)$	p_1	p_2	$\dots$

Then the **Expected Value** or the **Expectation** of X is the following number:

$$\mathbb{E}(X) = \sum_i x_i \cdot p_i = \sum_i x_i \cdot \mathbb{P}(X = x_i),$$

if $\sum_i |x_i| \cdot p_i < +\infty$ (this is important if the sum is infinite).

Example:

Example: Assume we are rolling a fair die. We get 0 points, if the number shown is less than 3, and get otherwise

$$10 \times (\text{the number shown - 4}) \quad \text{points.}$$

Example:

Example: Assume we are rolling a fair die. We get 0 points, if the number shown is less than 3, and get otherwise

$$10 \times (\text{the number shown - 4}) \quad \text{points.}$$

Let X be the r.v. showing the points we will get.

- Find the PMF of X;
- Calculate $\mathbb{E}(X)$.

Expectation of a R.V.

Note: The Expectation of a Discrete Variable X is the weighted average of its values, where the weights are the corresponding probabilities.

Expectation of a R.V.

Note: The Expectation of a Discrete Variable X is the weighted average of its values, where the weights are the corresponding probabilities.
In other words, if we think about our distribution as a mass distribution, then the Expected Value is the Center of Mass.

Expectation of a R.V.

Note: The Expectation of a Discrete Variable X is the weighted average of its values, where the weights are the corresponding probabilities.
In other words, if we think about our distribution as a mass distribution, then the Expected Value is the Center of Mass.
For example, if $X \sim DiscreteUnif(x_1, x_2, ..., x_n)$, then

Expectation of a R.V.

Note: The Expectation of a Discrete Variable X is the weighted average of its values, where the weights are the corresponding probabilities.
In other words, if we think about our distribution as a mass distribution, then the Expected Value is the Center of Mass.
For example, if $X \sim DiscreteUnif(x_1, x_2, ..., x_n)$, then

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

Expectation of a R.V.

Note: The Expectation of a Discrete Variable X is the weighted average of its values, where the weights are the corresponding probabilities.
In other words, if we think about our distribution as a mass distribution, then the Expected Value is the Center of Mass.
For example, if $X \sim DiscreteUnif(x_1, x_2, ..., x_n)$, then

$$\mathbb{E}(X) = x_1 \cdot \frac{1}{n} + x_2 \cdot \frac{1}{n} + ... + x_n \cdot \frac{1}{n} =$$

$$= \frac{x_1 + x_2 + ... + x_n}{n}.$$

Expectation of a R.V.

Note: The Expectation of a Discrete Variable X is the weighted average of its values, where the weights are the corresponding probabilities.
In other words, if we think about our distribution as a mass distribution, then the Expected Value is the Center of Mass.
For example, if $X \sim DiscreteUnif(x_1, x_2, ..., x_n)$, then

$$\mathbb{E}(X) = x_1 \cdot \frac{1}{n} + x_2 \cdot \frac{1}{n} + ... + x_n \cdot \frac{1}{n} =$$

$$= \frac{x_1 + x_2 + ... + x_n}{n}.$$

And if $Y \sim Bernoulli(0.99)$, then $\mathbb{E}(Y)$ is close to

Expectation of a R.V.

Note: The Expectation of a Discrete Variable X is the weighted average of its values, where the weights are the corresponding probabilities.
In other words, if we think about our distribution as a mass distribution, then the Expected Value is the Center of Mass.
For example, if $X \sim DiscreteUnif(x_1, x_2, ..., x_n)$, then

$$\mathbb{E}(X) = x_1 \cdot \frac{1}{n} + x_2 \cdot \frac{1}{n} + ... + x_n \cdot \frac{1}{n} =$$

$$= \frac{x_1 + x_2 + ... + x_n}{n}.$$

And if $Y \sim Bernoulli(0.99)$, then $\mathbb{E}(Y)$ is close to 1:

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

Expectation of a R.V.

Note: The Expectation of a Discrete Variable X is the weighted average of its values, where the weights are the corresponding probabilities.

In other words, if we think about our distribution as a mass distribution, then the Expected Value is the Center of Mass.

For example, if $X \sim DiscreteUnif(x_1, x_2, ..., x_n)$, then

$$\mathbb{E}(X) = x_1 \cdot \frac{1}{n} + x_2 \cdot \frac{1}{n} + ... + x_n \cdot \frac{1}{n} =$$

$$= \frac{x_1 + x_2 + ... + x_n}{n}.$$

And if $Y \sim Bernoulli(0.99)$, then $\mathbb{E}(Y)$ is close to 1:

$$\mathbb{E}(Y) = 0. \cdot 0.01 + 1 \cdot 0.99 = 0.99.$$

Expectation of a R.V.

Note: The Expected Value of X is not necessary a value of X.

Expectation of a R.V.

Note: The Expected Value of X is not necessary a value of X. Say, the Expected Value of points X we will get when rolling a fair die is

Expectation of a R.V.

Note: The Expected Value of X is not necessary a value of X. Say, the Expected Value of points X we will get when rolling a fair die is 3.5,

Expectation of a R.V.

Note: The Expected Value of X is not necessary a value of X. Say, the Expected Value of points X we will get when rolling a fair die is 3.5, but we can't roll 3.5 $\ddot\smile$.

Expectation of a R.V.

Note: The Expected Value of X is not necessary a value of X. Say, the Expected Value of points X we will get when rolling a fair die is 3.5, but we can't roll 3.5 $\ddot\smile$. Then what the Expected Value in this case shows?

Expectation of a R.V.

Note: The Expected Value of X is not necessary a value of X. Say, the Expected Value of points X we will get when rolling a fair die is 3.5, but we can't roll 3.5 $\ddot\smile$. Then what the Expected Value in this case shows? We will see later that it shows the long range average of the values of X:

Note: The Expected Value of X is not necessary a value of X. Say, the Expected Value of points X we will get when rolling a fair die is 3.5, but we can't roll 3.5 $\ddot\smile$. Then what the Expected Value in this case shows? We will see later that it shows the long range average of the values of X: if we will keep rolling our die, fixing the outcomes and calculating the average of all outcomes received so far, then that averages will approach 3.5 when we will roll our die again and again and again.

Expectation of a R.V.

Now, let us define the Expected Value for a Continuous r.v.

Expectation of a R.V.

Now, let us define the Expected Value for a Continuous r.v.

Expected Value of a Continuous R.V.

Assume X is a Continuous r.v. with the PDF $f(x)$.

Expectation of a R.V.

Now, let us define the Expected Value for a Continuous r.v.

Expected Value of a Continuous R.V.

Assume X is a Continuous r.v. with the PDF $f(x)$. Then the **Expected Value** or the **Expectation** of X is the following number:

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

Expectation of a R.V.

Now, let us define the Expected Value for a Continuous r.v.

Expected Value of a Continuous R.V.

Assume X is a Continuous r.v. with the PDF $f(x)$. Then the **Expected Value** or the **Expectation** of X is the following number:

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

assuming that $\int_{-\infty}^{\infty} |x| \cdot f(x)\, dx < +\infty$.

Example:

Example: Assume $X \sim Exp(2)$. Calculate $\mathbb{E}(X)$.

Expectation of a R.V., Properties

Note: Again, the Expectation $\mathbb{E}(X)$ shows the average value of X, the mean value of X.

Expectation of a R.V., Properties

Note: Again, the Expectation $\mathbb{E}(X)$ shows the average value of X, the mean value of X.
Another interpretation is that the Expectation is the Center of Mass of the Mass (Probability) Distribution given by its PDF.

Expectation of a R.V., Properties

Note: Again, the Expectation $\mathbb{E}(X)$ shows the average value of X, the mean value of X.

Another interpretation is that the Expectation is the Center of Mass of the Mass (Probability) Distribution given by its PDF.

Properties of the Expectation

Assume X, Y are r.v. defined on the same Experiment, with finite Expectations. Then

Expectation of a R.V., Properties

Note: Again, the Expectation $\mathbb{E}(X)$ shows the average value of X, the mean value of X.
Another interpretation is that the Expectation is the Center of Mass of the Mass (Probability) Distribution given by its PDF.

Properties of the Expectation

Assume X, Y are r.v. defined on the same Experiment, with finite Expectations. Then

- $\mathbb{E}(X)$ is a (deterministic) number, is not random;

Expectation of a R.V., Properties

Note: Again, the Expectation $\mathbb{E}(X)$ shows the average value of X, the mean value of X.
Another interpretation is that the Expectation is the Center of Mass of the Mass (Probability) Distribution given by its PDF.

Properties of the Expectation

Assume X, Y are r.v. defined on the same Experiment, with finite Expectations. Then

- $\mathbb{E}(X)$ is a (deterministic) number, is not random;
- $\mathbb{E}(C) = C$ for any constant C;

Expectation of a R.V., Properties

Note: Again, the Expectation $\mathbb{E}(X)$ shows the average value of X, the mean value of X.
Another interpretation is that the Expectation is the Center of Mass of the Mass (Probability) Distribution given by its PDF.

Properties of the Expectation

Assume X, Y are r.v. defined on the same Experiment, with finite Expectations. Then

- $\mathbb{E}(X)$ is a (deterministic) number, is not random;
- $\mathbb{E}(C) = C$ for any constant C;
- $\mathbb{E}(\mathbb{E}(X)) = \mathbb{E}(X)$;

Expectation of a R.V., Properties

Note: Again, the Expectation $\mathbb{E}(X)$ shows the average value of X, the mean value of X.
Another interpretation is that the Expectation is the Center of Mass of the Mass (Probability) Distribution given by its PDF.

Properties of the Expectation

Assume X, Y are r.v. defined on the same Experiment, with finite Expectations. Then

- $\mathbb{E}(X)$ is a (deterministic) number, is not random;
- $\mathbb{E}(C) = C$ for any constant C;
- $\mathbb{E}(\mathbb{E}(X)) = \mathbb{E}(X)$;
- If $X \geq 0$, then $\mathbb{E}(X) \geq 0$;

Expectation of a R.V., Properties

Note: Again, the Expectation $\mathbb{E}(X)$ shows the average value of X, the mean value of X.
Another interpretation is that the Expectation is the Center of Mass of the Mass (Probability) Distribution given by its PDF.

Properties of the Expectation

Assume X, Y are r.v. defined on the same Experiment, with finite Expectations. Then

- $\mathbb{E}(X)$ is a (deterministic) number, is not random;
- $\mathbb{E}(C) = C$ for any constant C;
- $\mathbb{E}(\mathbb{E}(X)) = \mathbb{E}(X)$;
- If $X \geq 0$, then $\mathbb{E}(X) \geq 0$;
- If $X \geq 0$ and $\mathbb{E}(X) = 0$, then $X = 0$ a.s., i.e., $\mathbb{P}(X = 0) = 1$;

Expectation of a R.V., Properties

Properties of the Expectation

Expectation of a R.V., Properties

Properties of the Expectation

- If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$;

Expectation of a R.V., Properties

Properties of the Expectation

- If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
- $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$;

Expectation of a R.V., Properties

Properties of the Expectation

- If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
- $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$;
- $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;

Expectation of a R.V., Properties

Properties of the Expectation

- If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
- $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$;
- $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
- $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$;

Expectation of a R.V., Properties

Properties of the Expectation

- If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
- $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$;
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
- $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$;
- $\mathbb{E}(\sum_{k=1}^{n} \alpha_k X_k) = \sum_{k=1}^{n} \alpha_k \mathbb{E}(X_k)$, for any r.v. X_k with finite Expectation and for any $\alpha_k \in \mathbb{R}$;

Expectation of a R.V., Properties

Properties of the Expectation

- If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
- $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$;
- $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
- $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$;
- $\mathbb{E}(\sum_{k=1}^n \alpha_k X_k) = \sum_{k=1}^n \alpha_k \mathbb{E}(X_k)$, for any r.v. X_k with finite Expectation and for any $\alpha_k \in \mathbb{R}$;
- If $\mathbb{1}_A$ is the **characteristic (indicator) function** of the Event A,

Expectation of a R.V., Properties

Properties of the Expectation

- If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
- $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$;
- $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
- $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$;
- $\mathbb{E}(\sum_{k=1}^{n} \alpha_k X_k) = \sum_{k=1}^{n} \alpha_k \mathbb{E}(X_k)$, for any r.v. X_k with finite Expectation and for any $\alpha_k \in \mathbb{R}$;
- If $\mathbb{1}_A$ is the **characteristic (indicator) function** of the Event A, i.e., $\mathbb{1}_A$ shows 1 as A occurs,

Expectation of a R.V., Properties

Properties of the Expectation

- If $X \geq Y$, then $\mathbb{E}(X) \geq \mathbb{E}(Y)$;
- $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$;
- $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
- $\mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$;
- $\mathbb{E}(\sum_{k=1}^{n} \alpha_k X_k) = \sum_{k=1}^{n} \alpha_k \mathbb{E}(X_k)$, for any r.v. X_k with finite Expectation and for any $\alpha_k \in \mathbb{R}$;
- If $\mathbb{1}_A$ is the **characteristic (indicator) function** of the Event A, i.e., $\mathbb{1}_A$ shows 1 as A occurs,

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A, \end{cases}$$

  then $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$;

Comments about the Indicator function

We can describe the Indicator R.V. by giving its Distribution:

Values of $\mathbb{1}_A$	0	1
$\mathbb{P}(\mathbb{1}_A = x)$	$1 - \mathbb{P}(A)$	$\mathbb{P}(A)$

Comments about the Indicator function

We can describe the Indicator R.V. by giving its Distribution:

Values of $\mathbb{1}_A$	0	1
$\mathbb{P}(\mathbb{1}_A = x)$	$1 - \mathbb{P}(A)$	$\mathbb{P}(A)$

And in this case it is clear that

$$\mathbb{E}(\mathbb{1}_A) = 0 \cdot (1 - \mathbb{P}(A)) + 1 \cdot \mathbb{P}(A) = \mathbb{P}(A).$$

Comments about the Indicator function

We can describe the Indicator R.V. by giving its Distribution:

Values of $\mathbb{1}_A$	0	1
$\mathbb{P}(\mathbb{1}_A = x)$	$1 - \mathbb{P}(A)$	$\mathbb{P}(A)$

And in this case it is clear that

$$\mathbb{E}(\mathbb{1}_A) = 0 \cdot (1 - \mathbb{P}(A)) + 1 \cdot \mathbb{P}(A) = \mathbb{P}(A).$$

Note: Note that the PMF of $\mathbb{1}_A$ contains less information than the Definition of $\mathbb{1}_A$: PMF says which values can take $\mathbb{1}_A$, and with which probabilities (proportions), but the definition will give us *in which cases* $\mathbb{1}_A$ *will take the value* 0, *and in which case - the value* 1.

Expectation of Transformed R.V.

Now, assume X is a r.v., $g : \mathbb{R} \to \mathbb{R}$ is a given function, and $Y = g(X)$.

Expectation of Transformed R.V.

Now, assume X is a r.v., $g : \mathbb{R} \to \mathbb{R}$ is a given function, and $Y = g(X)$. We want to calculate the Expected Value $\mathbb{E}(Y) = \mathbb{E}(g(X))$.

Expectation of Transformed R.V.

Now, assume X is a r.v., $g : \mathbb{R} \to \mathbb{R}$ is a given function, and $Y = g(X)$. We want to calculate the Expected Value $\mathbb{E}(Y) = \mathbb{E}(g(X))$.

- **Method 1:**

Expectation of Transformed R.V.

Now, assume X is a r.v., $g : \mathbb{R} \to \mathbb{R}$ is a given function, and $Y = g(X)$. We want to calculate the Expected Value $\mathbb{E}(Y) = \mathbb{E}(g(X))$.

- **Method 1:**
  - *Discrete Case* We first calculate the PMF of Y, $\mathbb{P}(Y = x)$ (we know how to do it!!), and use

$$\mathbb{E}(Y) = \sum_x x \cdot \mathbb{P}(Y = x);$$

Expectation of Transformed R.V.

Now, assume X is a r.v., $g : \mathbb{R} \to \mathbb{R}$ is a given function, and $Y = g(X)$. We want to calculate the Expected Value $\mathbb{E}(Y) = \mathbb{E}(g(X))$.

- **Method 1:**
  - *Discrete Case* We first calculate the PMF of Y, $\mathbb{P}(Y = x)$ (we know how to do it!!), and use
  $$\mathbb{E}(Y) = \sum_x x \cdot \mathbb{P}(Y = x);$$
  - *Continuous Case* We first calculate the PDF of Y, $f_Y(x)$ (we know how to do it!!), and use
  $$\mathbb{E}(Y) = \int_{-\infty}^{\infty} x \cdot f_Y(x)\, dx.$$

Expectation of Transformed R.V.

Expectation of Transformed R.V.

- **Method 2:** We can omit the calculation of the PD(M)F of the transformed variable Y, and use only the Distribution of X:

Expectation of Transformed R.V.

- **Method 2:** We can omit the calculation of the PD(M)F of the transformed variable Y, and use only the Distribution of X:
  - *Discrete Case* We can use

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \sum_{x} g(x) \cdot \mathbb{P}(X = x);$$

Expectation of Transformed R.V.

- **Method 2:** We can omit the calculation of the PD(M)F of the transformed variable Y, and use only the Distribution of X:
  - *Discrete Case* We can use
$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \sum_x g(x) \cdot \mathbb{P}(X = x);$$
  - *Continuous Case* We can use
$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx.$$

Expectation of Transformed R.V.

- **Method 2:** We can omit the calculation of the PD(M)F of the transformed variable Y, and use only the Distribution of X:
  - *Discrete Case* We can use
  $$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \sum_x g(x) \cdot \mathbb{P}(X = x);$$
  - *Continuous Case* We can use
  $$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx.$$

Note: Note that here we are using the PDF/PMF of X only!

Expectation of Transformed R.V.

- **Method 2:** We can omit the calculation of the PD(M)F of the transformed variable Y, and use only the Distribution of X:
  - *Discrete Case* We can use
$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \sum_x g(x) \cdot \mathbb{P}(X = x);$$
  - *Continuous Case* We can use
$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)\, dx.$$

Note: Note that here we are using the PDF/PMF of X only!
Note: This method is called the **law of the unconscious statistician (LOTUS)**, see, e.g. `https://en.wikipedia.org/wiki/Law_of_the_unconscious_statistician`

Example

Example: Assume $X \sim Binom(3, 0.2)$. Calculate $\mathbb{E}(\sin\left(\frac{\pi}{2} \cdot X\right))$.

Example

Example: Assume $X \sim Binom(3, 0.2)$. Calculate $\mathbb{E}(\sin\left(\frac{\pi}{2} \cdot X\right))$.

Example: Assume $X \sim Unif[-2, 1]$. Calculate $\mathbb{E}(X^2)$.

Expectation for the r.v. obtained from 2 r.v.s

Assume X and Y are Jointly Distributed r.v.s, and $Z = g(X, Y)$ for some known function g. We want to calculate $\mathbb{E}(Z)$.

Expectation for the r.v. obtained from 2 r.v.s

Assume X and Y are Jointly Distributed r.v.s, and $Z = g(X, Y)$ for some known function g. We want to calculate $\mathbb{E}(Z)$.

- **Discrete Case:** If the values of X are $x_1, x_2, ...$ and the values of Y are $y_1, y_2, ...,$ then

$$\mathbb{E}(Z) = \mathbb{E}(g(X, Y)) = \sum_{i,j} g(x_i, y_j) \cdot \mathbb{P}(X = x_i, Y = y_j)$$

Expectation for the r.v. obtained from 2 r.v.s

Assume X and Y are Jointly Distributed r.v.s, and $Z = g(X, Y)$ for some known function g. We want to calculate $\mathbb{E}(Z)$.

- **Discrete Case:** If the values of X are $x_1, x_2, \ldots$ and the values of Y are $y_1, y_2, \ldots$, then

$$\mathbb{E}(Z) = \mathbb{E}(g(X, Y)) = \sum_{i,j} g(x_i, y_j) \cdot \mathbb{P}(X = x_i, Y = y_j)$$

- **Continuous Case:** If $f(x, y)$ is the Joint PDF of X and Y, then

$$\mathbb{E}(Z) = \mathbb{E}(g(X, Y)) = \iint_{\mathbb{R}^2} g(x, y) \cdot f(x, y)\, dxdy.$$

Example:

Assume we know the Joint Distribution of X and Y, say, we work in the Continuous case, and their Joint PDF is $f(x, y)$. We want to calculate the Expected Value of X.

Example:

Assume we know the Joint Distribution of X and Y, say, we work in the Continuous case, and their Joint PDF is $f(x, y)$. We want to calculate the Expected Value of X.

- **Method 1:** First calculate the Marginal PDF of X, $f_X(x)$, and then use

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

Example:

Assume we know the Joint Distribution of X and Y, say, we work in the Continuous case, and their Joint PDF is $f(x, y)$. We want to calculate the Expected Value of X.

- **Method 1:** First calculate the Marginal PDF of X, $f_X(x)$, and then use
$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx.$$
- **Method 2:** Just use Double Integration:
$$\mathbb{E}(X) = \iint_{\mathbb{R}^2} x \cdot f(x, y)\, dxdy.$$

Example:

Question: Is it true that

$$\mathbb{E}(X^2) = \Big(\mathbb{E}(X)\Big)^2 ?$$

Expectation and Independence

Assume X and Y are Independent r.v. Then

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Expectation and Independence

Assume X and Y are Independent r.v. Then

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Moreover, if g and h are any (nice!) functions, then

$$\mathbb{E}(g(X) \cdot h(Y)) = \mathbb{E}(g(X)) \cdot \mathbb{E}(h(Y)).$$

Expectation and Independence

Assume X and Y are Independent r.v. Then

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Moreover, if g and h are any (nice!) functions, then

$$\mathbb{E}(g(X) \cdot h(Y)) = \mathbb{E}(g(X)) \cdot \mathbb{E}(h(Y)).$$

Example: Say, if $X \perp\!\!\!\perp Y$, then

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

Expectation and Independence

Assume X and Y are Independent r.v. Then

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Moreover, if g and h are any (nice!) functions, then

$$\mathbb{E}(g(X) \cdot h(Y)) = \mathbb{E}(g(X)) \cdot \mathbb{E}(h(Y)).$$

Example: Say, if $X \perp\!\!\!\perp Y$, then

$$\mathbb{E}((X + Y)^2) = \mathbb{E}(X^2 + 2XY + Y^2) =$$

Jensen's Inequality

Recall the Jensen's Inequality for convex functions: if $g : [a, b] \to \mathbb{R}$ is a convex function, $x_1, ..., x_n \in [a, b]$ and $\alpha_1, ..., \alpha_n \in [0, 1]$ with $\alpha_1 + \alpha_2 + ... + \alpha_n = 1$, then

$$g\left(\sum_{k=1}^{n} \alpha_k x_k\right) \leq \sum_{k=1}^{n} \alpha_k \cdot g(x_k).$$

Jensen's Inequality

Recall the Jensen's Inequality for convex functions: if $g : [a, b] \to \mathbb{R}$ is a convex function, $x_1, \dots, x_n \in [a, b]$ and $\alpha_1, \dots, \alpha_n \in [0, 1]$ with $\alpha_1 + \alpha_2 + \dots + \alpha_n = 1$, then

$$g\left(\sum_{k=1}^{n} \alpha_k x_k\right) \leq \sum_{k=1}^{n} \alpha_k \cdot g(x_k).$$

Now, the generalization, and Probabilistic Interpretation is:

Jensen's Inequality

Recall the Jensen's Inequality for convex functions: if $g : [a, b] \to \mathbb{R}$ is a convex function, $x_1, ..., x_n \in [a, b]$ and $\alpha_1, ..., \alpha_n \in [0, 1]$ with $\alpha_1 + \alpha_2 + ... + \alpha_n = 1$, then

$$g\left(\sum_{k=1}^{n} \alpha_k x_k\right) \leq \sum_{k=1}^{n} \alpha_k \cdot g(x_k).$$

Now, the generalization, and Probabilistic Interpretation is:

Jensen's Inequality

If X is a r.v. and $g : \mathbb{R} \to \mathbb{R}$ is a **convex function**, then

$$g(\mathbb{E}(X)) \leq \mathbb{E}(g(X)).$$