STAT 320: Principles of Probability

Unit 5(PART:A)

Introduction to Random Variables & Discrete Random Variables

United Arab Emirates University

Department of Statistics

Outline

Random Variables

In majority of the scenarios, we are often interested some specific numeric attributes of the data rather then overall outcome that often may be recorded in terms of symbols.

Events of major interests to a professional dealing with data are mostly numerical in nature.

- A business person might be interested in aggregate sales in the last few years and wan to predict the amount for the next year.
- An associate in an insurance company may be interested in the funds that should be made available/reserved to compensate the losses of its customers.
- A marketing agent may be interested in determining the additional orders that may take place due to a specific sales-promotion that he plans to launch.

Random Variables

- A television manufacturer might be interested in lifetime of a newly designed led screen that they have designed.
- In a much simple statistical experiment, such as in tossing dices, we may be interested in the number of claims that that are likely to be filed within a specific time period. .

These quantities of interest, or, more formally, these real-valued functions defined on the sample space, are known as random variables.

Definition (Random Variable)

A random variable is a function from a sample space $\mathscr{S}$ into the real numbers.

Example: Random Variable

Experiment	Random Variable
Toss two dice	X= sum of the numbers
Toss a coin 25 times	X number of heads in 25 tosses
Apply different amounts of fertilizer to corn plants	X = yield/acre

Notation: Random variables will always be denoted with uppercase letters and the realized values of the variable (or its range) will be denoted by the corresponding lowercase letters. Thus, the random variable X can take the value x.

Support/Range of a Random Variable

Definition (Support/Range of a Random Variable)

The set containing the all possible values of a random variable is called its **support** or **range**.

Notation: We will use the notation $\mathbb{S}_x$ (or simply $\mathbb{S}$ if there is no ambiguity) to denote the support of a random variable X.

Example: Consider the experiment of tossing a fair coin 3 times from. Define the random variable X to be the number of heads obtained in the 3 tosses. The Support of the random variable is $\mathbb{S}_x = \{0, 1, 2, 3\}$

Example

Example : Suppose that our experiment consists of tossing 3 fair coins. If we let Y denote the number of heads that appear, then Y is a random variable taking on one of the values $0, 1, 2$, and 3 with respective probabilities.

$$p_Y(0) = P(Y = 0) =$$

$$p_Y(1) = P(Y = 1) =$$

$$p_Y(2) = P(Y = 2) =$$

$$p_Y(3) = P(Y = 3) =$$

Outline

Discrete Random Variables & its Probability Mass Function (pmf)

Definition (Discrete Random Variables)

A random variable that can take on at most a countable number of possible values is said to be discrete. *That is, if $\mathbb{S}$, the support of a random variable is finite or countable infinite then the corresponding random variable is discrete.*

Probability Mass Function (pmf) For a discrete random variable X, we define the probability mass function (pmf) $p_x(x)$ of X by

$$p_x(x) = P(X = x) \text{ for all } x \in \mathbb{S}_x$$

Let X be a discrete random variable with probability mass function $p(x)$ defined on the support $\mathbb{S}$. Let $A \subset \mathbb{S}$ be an event, then

$$P\left(A\right) := P(X \in A) = \sum_{\left\{x \in A\right\}} p(x).$$

The pmf of the random variable representing the sum when two dice are rolled can be represented in multiple ways.

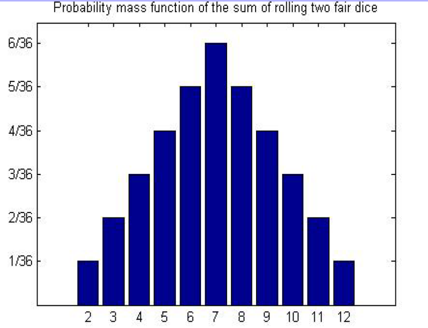
(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

The pmf of the random variable representing the sum when two dice are rolled can be represented in multiple ways.

As a Tabular Format:

x	2	3	4	5	6	7	8	9	10	11	12
$p_x(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

As a Plot/Graph:



As a Function

$$p_x(x) = \begin{cases} \frac{x-1}{36} & \text{if } 2 \leq x \leq 7, \\ \frac{13-x}{36} & \text{if } 8 \leq x \leq 12 \end{cases}$$

Characterization of a pmf

Let $p(x)$ is **probability mass function** of a discrete random variable on the support $\mathbb{S}$, **if and only if** it satisfies the following conditions:

1. *Positivity:* $p(x) > 0$ for all $x \in \mathbb{S}$
2. *Total Probability:* $\sum_{\{x \in \mathbb{S}\}} p(x) = 1$.

Example: Check whether the examples are valid pmf

Example : Suppose a random variable X has the following support $\mathbb{S}_x = \{1, 2, 3, 4, 5\}$.

x	1	2	3	4	5
$p_x(x)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{10}$	c	$\frac{1}{10}$

1. What is the value of c so that $p_x(x)$ become a valid probability mass function on the support $\mathbb{S}_x$?
2. What is the probability that $X \in A$, where the event $A = \{2, 4\}$.

Reminder from Unit1: Geometric Series

Let $p \in \mathbb{R}$ be such that $|p| < 1$, then

$$\boxed{\sum_{i=0}^{\infty} p^i} = 1 + p + p^2 + p^3 + \cdots = \boxed{\frac{1}{1-p}}.$$

1. What is the value of $1 + 0.7 + (0.7)^2 + (0.7)^3 + \cdots =$
2. What is the value of $1 + \frac{1}{7} + \frac{1}{7^2} + \frac{1}{7^3} + \cdots =$

Example : Suppose a random variable X has the following support $\mathbb{S}_X = \{0, 1, 2, 3, \dots\}$. The corresponding probability mass function is

$$p(x) := \frac{c}{5^x} \quad \text{for } x = 0, 1, 2, 3, \dots$$

where c is an appropriately chosen constant.

1. What is the value of c so that $p(x)$ become a valid probability mass function on the support $\mathbb{S}_X$?
2. What is the probability that $X \in A$, where the event $A = \{1, 3, 5\}$.
3. What is the probability that $X \leq 3$?
4. What is the probability that $X > 3$?

Example

Example : A system consists of 2 components connected in parallel, then at least one must work correctly for the system to work correctly. Each component operates correctly with probability 0.8 and independent of the other. Let X be the number of components that work correctly. Find the probability distribution of X.

Example

Example : A system consists of 2 components connected in parallel, then at least one must work correctly for the system to work correctly. Each component operates correctly with probability 0.8 and independent of the other. Let X be the number of components that work correctly. Find the probability distribution of X.

Solution: X can take on only three possible values; 0, 1, or 2. Let E_i denote the event that component i works correctly. Then $P(E_i) = 0.8$. Thus, we have

- $p_X(0) = P(\overline{E_1} \cap \overline{E_2}) = P(\overline{E_1})P(\overline{E_2}) = (0.2)(0.2) = 0.04.$
- $p_X(1) = P(\overline{E_1} \cap E_2) + P(E_1 \cap \overline{E_2}) = (0.2)(0.8) + (0.8)(0.2) = 0.32.$
- $p_X(2) = P(E_1 \cap E_2) = P(E_1)P(E_2) = (0.8)(0.8) = 0.64.$

x	0	1	2
$p_X(x)$	0.04	0.32	0.64

Outline

Cumulative Distribution Function (CDF) of a discrete Random Variable

cumulative distribution function

Definition (cumulative distribution function)

Let X be a discrete random variable on the support $\mathbb{S}_X$ with the corresponding probability mass function

$$P(X = x) = p_X(x) \text{ for } x \in \mathbb{S}_X .$$

Then for any $a \in \mathbb{R}$, the cumulative distribution function (cdf), denoted by $F_X(\cdot)$ is the following quantity

$$F_X(a) = P(X \leq a) = \sum_{\{x \leq a \,:\, x \in \mathbb{S}_X\}} p_X(x)$$

A **pmf** of a discrete random variable is only positive/ revelant on the support of the random variable $\mathbb{S}$, However the **CDF** is defined for any real number.

Example

Example : If X be a discrete random variable on the support $\mathbb{S}_x = \{1, 2, 3, 4\}$ with the corresponding pmf specified as $p_x(1) = \frac{1}{4}, p_x(2) = \frac{1}{2}, p_x(3) = \frac{1}{8}$, and $p_x(4) = \frac{1}{8}$. Calculate the CDF function of X.

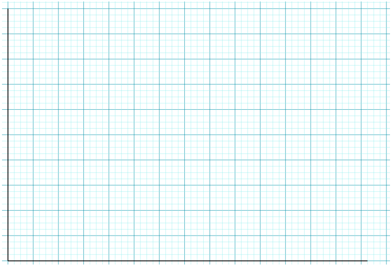
Example

Example : If X be a discrete random variable on the support $\mathbb{S}_X = \{1, 2, 3, 4\}$ with the corresponding pmf specified as $p_X(1) = \frac{1}{4}, p_X(2) = \frac{1}{2}, p_X(3) = \frac{1}{8}$, and $p_X(4) = \frac{1}{8}$. Calculate the CDF function of X.

Solution:

$$F_X\left(a\right) = \begin{cases} 0 & \text{if } a < 1 \\ \frac{1}{4} & \text{if } 1 \leq a < 2 \\ \frac{3}{4} & \text{if } 2 \leq a < 3 \\ \frac{7}{8} & \text{if } 3 \leq a < 4 \\ \frac{7}{8} & \text{if } 4 \leq a \end{cases}$$

Graph of CDF of X



Let the pmf of a discrete random varibale X is given as

x	0	1	2
$p_X(x)$	0.04	0.32	0.64

Find the corresponding CDF.

Let the pmf of a discrete random varibale X is given as

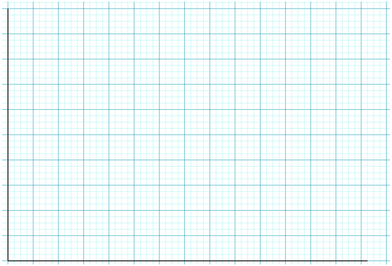
x	0	1	2
$p_X(x)$	0.04	0.32	0.64

Find the corresponding CDF.

Solution:

$$F_X\left(a\right) = \begin{cases} 0 & \text{if } a < 0 \\ 0.04 & \text{if } 0 \le a < 1 \\ 0.36 & \text{if } 1 \le a < 2 \\ 1 & \text{if } 2 \le a \end{cases}$$

Graph of CDF of X



Outline

Expected Value & Variance of a Discrete Random Variable

The "Expected Value" or "Mean" of a Discrete Random Variable

Definition (The "**Expected Value**" or "**Mean**" of a Discrete Random Variable)

If X is a random variable with pmf $p_X(x)$ on the support $\mathbb{S}_X$, then the expected value (the mean) of X denoted by $E(X)$ (or μ_X) is given by

$$\boxed{E\left(X\right) = \sum_{\left\{x \in \mathbb{S}_X\right\}} x \, p_X(x),}$$

assuming the above summation/series exists /well-defined.

Definition (The Expected Value of a Function of a Discrete Random Variable)

Let the random variable X has the probability mass funciton $p_X(x)$ for all $x \in \mathbb{S}_X$, the support of X. Let $h(x)$ be any* function, then the expected value of $h(X)$ is defined as

$$\boxed{E\Big(h(X)\Big) = \sum_{\{x \in \mathbb{S}_X\}} h(x)\, p_X(x),}$$

assuming the above summation/series exists /well-defined.

Variance

Variance The variance of X, denoted by Var(X) is deifined as

$$\boxed{\mathrm{Var}(X) := E\left(X - \mu_X\right)^2},$$

where $\mu_X = E(X)$, the mean of the random variable.

Definition (Variance)

The variance of X, denoted by Var(X) is deifined as

$$\boxed{\mathrm{Var}(X) := E(X^2) - \left(E(X)\right)^2}$$

Standard Deviation

Definition (Variance)

The variance of X, denoted by Var(X) is deifined as

$$\boxed{\sigma_X = \mathrm{SD}(X) := \sqrt{\mathrm{Var}(X)}}$$

Var(X) is often denoted by σ^2.

Properties of Expected Value and Variance

Let $a, b \in \mathbb{R}$ are two constants where as X is a random variable , then

1. $E(aX + b) = aE(X) + b$
2. $\text{Var}(aX + b) = a^2\text{Var}(X)$
3. $\text{SD}(aX + b) = |a|\,\text{SD}(X)$

Outline

Moment Generating Function (mgf) of a Discrete Random Variable

Moment Generating Function (mgf)

Definition (Moment Generating Function)

The Moment Generating Function (mgf) of X, denoted by $\mathrm{M}_x(t)$ is deifined as

$$\boxed{\mathrm{M}_x(t) := E\left(e^{tX}\right)},$$

whenever it exists.

If the random variable X has the probability mass funciton $p_x(x)$ for all $x \in \mathbb{S}_x$, the support of X, then assuming it exists

$$M_x(t) := E\left(e^{tX}\right) = \sum_{\{x \in \mathbb{S}_x\}} e^{tx}\, p_x(x).$$

Properties of a Moment Generating Function

- If it exists, the moment generating function is unique for a random variable. It means, no two random variable/distribution can have same moment geneting function. Therefore, a distribution can be identified by the form of its moment generating function.

- If it exists, the mgf can be used to obtain the moments of a random variable in the following way:

Assuming it exists

$$\frac{d}{dt}\{M_X(t)\}\Big|_{t=0} = E(X)$$

Assuming it exists

$$\frac{d^k}{dt^k}\{M_X(t)\}\Big|_{t=0} = E(X^k) \text{ for } k = 1, 2, \ldots$$

Outline

A Few Examples

Example

The probability distribution of X, the number of daily network black-outs is given by

x	0	1	2
$p_X(x)$	0.7	0.2	0.1

Find the Expected value and variance of the random variable **X**.

Example

The probability distribution of X, the number of daily network black-outs is given by

x	0	1	2
$p_X(x)$	0.7	0.2	0.1

Find the Expected value and variance of the random variable **X.**

Solution:

$$\begin{aligned}\mu_X = E(X) &= \sum_{x\in\{0,1,2\}} x p_X(x)\\ &= 0\times p_X(0)+1\times p_X(1)+2\times p_X(2)\\ &= 0\times 0.7+1\times 0.2+2\times 0.1\\ &= 0.4\end{aligned}$$

$$\begin{aligned}E(X^2) &= \sum_{x\in\{0,1,2\}} x^2 p_X(x)\\ &= 0^2\times p_X(0)+1^2\times p_X(1)+2^2\times p_X(2)\\ &= 0\times 0.7+1\times 0.2+4\times 0.1\\ &= 0.6\end{aligned}$$

Hence $\text{Var}(X) := E(X^2)-(E(X))^2 = 0.6-(0.4)^2 = 0.6-0.16 = 0.44$

The Standard Deviation $\sigma = \sqrt{\sigma^2} = \sqrt{\text{Var}(X)} = \sqrt{0.44} = 0.6633$

Example

The probability distribution of X, the number of daily network blackouts is given by

x	0	1	2
$p_X(x)$	0.7	0.2	0.1

A small internet trading company estimates that each network blackout results in a $500 loss. Compute expectation and variance of this company's daily loss due to blackouts.

Example

The probability distribution of X, the number of daily network blackouts is given by

x	0	1	2
$p_X(x)$	0.7	0.2	0.1

A small internet trading company estimates that each network blackout results in a \$500 loss. Compute expectation and variance of this company's daily loss due to blackouts.

The daily loss due to blackouts is given by $h(X) = 500X$. We need to find $E(h(X))$ and Varinace of $Var(h(X))$.

Solution:

$$\begin{aligned}
\mu_X = E(X) &= \sum_{x\in\{0,1,2\}} x p_X(x) \\
&= 0 \times p_X(0) + 1 \times p_X(1) + 2 \times p_X(2) \\
&= 0 \times 0.7 + 1 \times 0.2 + 2 \times 0.1 \\
&= 0.4
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= \sum_{x\in\{0,1,2\}} x^2 p_X(x) \\
&= 0^2 \times p_X(0) + 1^2 \times p_X(1) + 2^2 \times p_X(2) \\
&= 0 \times 0.7 + 1 \times 0.2 + 4 \times 0.1 \\
&= 0.6
\end{aligned}$$

Let the pmf of a discrete is given as

$$p_X(x) := \frac{1}{2^x} \text{ for } x = 1, 2, 3, \ldots$$

Exercises on Computing $E(X)$, $Var(X)$, MGF

Find $E(X)$ and $Var(X)$, where X is the outcome when we roll a fair die.

Reminder from Unit1

Definition (Indicator Function)

Let A be a set. The **indicator function for the set** A, denoted by $\mathbb{I}_A(x)$, is defined to be

$$\mathbb{I}_A(x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

- $\mathbb{I}_{[0,5]}(1.33) =$
- $\mathbb{I}_{[0,5]}(-9.12) =$
- $\mathbb{I}_{\{HH,TT,HT\}}(HH) =$
- $\mathbb{I}_{\{HH,TT,HT\}}(HHHH) =$
- $\mathbb{I}_{\mathbb{Z}}(1.87) =$
- $\mathbb{I}_{\mathbb{R}_+}(-4.87) =$
- $\mathbb{I}_{\mathbb{R}_+}(14) =$

Exercises on Computing $E(X)$, $Var(X)$, $MGF(X)$

Question : : Let X be a discrete random variable with the probability mass function $p_X(x)$ on the support $\mathbb{S}_X$. Let $Y = \mathbb{I}_A(X)$, obtain $E(Y)$, $\mathrm{Var}(Y)$, and $\mathrm{MGF}(Y)$.

Questions?