Ferhat Abbas-Sétif 1 University
Faculty of Sciences
Department Informatics

Chapter 2 : Random Variables

A random variable is a numerical description of the outcome of a statistical experiment. A random variable that may assume only a finite number or an infinite sequence of values is said to be discrete ; one that may assume any value in some interval on the real number line is said to be continuous. For instance, a random variable representing the number of auto-mobiles sold at a particular dealership on one day would be discrete, while a random variable representing the weight of a person in kilograms would be continuous.

Définition 1. A random variable is a function which maps from sample space of an experiment S to the real numbers. Mathematically, Random Variable is expressed as,

$$X : S \to \mathbb{R}$$

Exemple 1. The sample space of this experiments is $S = \{TT, HT, TH, HH\}$. If Y is the random variable that refers to the number of heads from tossing two coins, then Y takes the values $\{0, 1, 2\}$. This means that we could have no heads, one head, or both heads on a two-coin toss.

Exemple 2. A non-biased coin is tossed three times. The sample space is

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Let X be the random variable which assigns to each point in S the largest number of successive heads which occurs. Then R.V. X takes the values $0, 1, 2$ and 3. Indeed,

$$X(TTT) = 0, X(HTH) = 1, X(HTT) = 1, X(THT) = 1,$$

$$X(TTH) = 1, X(HHT) = 2, X(THH) = 2, X(HHH) = 3.$$

1. Discrete random variables.

Définition 2. A random variable X is discrete if its support R_X is countable and there exists a function $p_X : \mathbb{R} \to [0, 1]$, called Probability Mass Function (PMF) of X, such that

$$p_X(x) = P(X = x) = P(X^{-1}(x))$$

where $P(X = x)$ is the probability that X that takes the value x.

Exemple 3. 1- The PMF that corresponds to the random variable Y of Example 1 is

$$p_Y(x) = 1/4 \text{ if } x \in \{0, 1, 2\}, \text{ and } p_Y(x) = 0 \text{ elswere.}$$

$$p_Y(x) = \begin{cases} 1/4, & \text{if } x \in \{0, 2\} \\ 1/2, & \text{if } x = 1 \\ 0, & \text{otherwise.} \end{cases}$$

2- The PMF that corresponds to the random variable X of Example 2 is

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

1.1. Cumulative distribution function

In probability, the Cumulative Distribution Function (CDF) of a random variable X, is the probability that X takes a value less than or equal to x.

Définition 3. Let X be a random variable. The cumulative distribution function (CDF) or the distribution function of X is defined as

$$F_X(x) = P(X \leq x)$$

for any $x \in \mathbb{R}$. If X is discrete , then $F_X(x_0) = P(X \leq x_0) = \sum_{x \leq x_0} p_X(x)$.

Exemple 4. We toss a coin three times successively, and let X the number of obtained heads. The PMF that corresponds to the random variable X is

$$p_X(x) = \begin{cases} 1/8, \text{ if } x \in \{0, 3\} \\ 3/8, \text{ if } x \in \{1, 2\} \\ 0, \text{ otherwise.} \end{cases}$$

Therefore, by summing up the probabilities up to the value of x we get the following CDF,

$$F_X(x) = \begin{cases} 0 & -\infty < x < 0 \\ \frac{1}{8} & 0 \leq x < 1 \\ \frac{4}{8} & 1 \leq x < 2 \\ \frac{7}{8} & 2 \leq x < 3 \\ 1 & 3 \leq x < +\infty \end{cases}$$

Proposition 1. The CDF of a random variable X has the following properties :

1- F_X is right continuous and increasing, that is, if $x < y$, then $F(x) \leq F(y)$,

2- $\lim\limits_{x \to +\infty} F_X(x) = 1$,

3- $\lim\limits_{x \to -\infty} F_X(x) = 0$.

Independent random variables

The concept of independent random variables is very similar to independent events. Remember, two events A and B are independent if we have $P(A, B) = P(A)P(B)$(comma means "and", i.e., $P(A, B) = P(A \cap B)$). Similarly, we have the following definition for independent discrete random variables.

Définition 4. Consider n discrete random variables $X_1, X_2, X_3, ..., X_n$. We say that $X_1, X_2, X_3, ..., X_n$ are independent if

$$P\Big(X_1 = x_1, X_2 = x_2, ..., X_n = x_n\Big) = P(X_1 = x_1)P(X_2 = x_2)...P(X_n = x_n).$$

In general, if n random variables are independent, then

$$P\Big(X_1 \in A_1, X_2 \in A_2 \ldots X_n \in A_n\Big) = P(X_1 \in A_1)P(X_2 \in A_2) \ldots P(X_n \in A_n).$$

Exemple 5. I toss a coin twice and define X to be the number of heads I observe. Then, I toss the coin two more times and let Y be the number of heads that I observe in last two tosses. Find

$$P\Big(X \leq 1, Y = 2\Big).$$

Since X and Y are the result of different independent coin tosses, the two random variables X and Y are independent. We can write

$$\begin{aligned} P\Big(X \leq 1, Y = 2\Big) &= P(X \leq 1)P(Y = 2) \qquad \text{(since } X \text{ and } Y \text{ are independent)} \\ &= \big(p_X(0) + p_X(1)\big)p_Y(2) \\ &= \left(\frac{1}{4} + \frac{1}{2}\right)\frac{1}{4} = 3/16 \end{aligned}$$

1.2. Mean and Standard Deviation of Discrete Random Variables

The most important characteristics of any probability distribution are the mean, expected value, (or even average value) and the standard deviation.

1.2.1. The expected value of discrete random variables

The mean or expected value of a random variable X is written as $\mathbb{E}(x)$ or μ_x. If we observe n random values of X, then the mean of the n values will be approximately equal to $\mathbb{E}(x)$ for large n. The expected value is defined differently for continuous and discrete random variables.

Définition 5. Let X be a discrete random variable with PMF $p_X(x)$. The expected value of X is

$$\mathbb{E}(X) = \sum_{x \in R_X} x p_X(x)$$

The following example illustrates how to calculate the mean of a discrete random variable.

Exemple 6. Find the mean number of heads obtained in 3 flips of a balanced coin.
Solution
The PMF that corresponds to the r.v. X is

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

Therefore,

$$\mathbb{E}(X) = \sum_{x \in \{0,1,2,3\}} x p_X(x) = \frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8} = \frac{3}{2}$$

Remarque 1. If g is a function of the random variable X then the expectation of $g(X)$ is given by

$$\mathbb{E}(g(X)) = \sum_{x \in R_X} g(X) p_X(x)$$

1.2.2. Properties of the expected value

Théorème 1. *Let a be a constant, then $\mathbb{E}(a) = a$*

Démonstration. From the mathematical definition of expectation we know that

$$\begin{aligned} \mathbb{E}(c) =& \sum_x x p_c(x) \\ =& c \times 1 = c \end{aligned}$$

□

Théorème 2. *Let X and Y be random variables on the same sample space S. Then $E(X+Y) = E(X)+E(Y)$. More generally, if $X_1, X_2, \ldots, X_n$ be n random variables on the same sample space S. Then, $\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i)$.*

Théorème 3. .
1- Let X be a random variable on the same sample space S and c is a constant. Then , $\mathbb{E}(cX) = c\mathbb{E}(X)$.
2- Let X be a random variables on the same sample space S, and a and c are two constants, then $\mathbb{E}(aX + c) = a\mathbb{E}(X) + c$.

Démonstration. 1-

$$\begin{aligned}\mathbb{E}(cX) &= \sum_{x\in R_X} cxp_X(x)\\ &= c\sum_{x\in R_X} xp_X(x) = c\mathbb{E}(X)\end{aligned}$$

2-

$$\begin{aligned}\mathbb{E}(aX+c) &= \mathbb{E}(aX)+\mathbb{E}(c)\\ &= a\mathbb{E}(X)+c\end{aligned}$$

□

As a consequence, the mean value of the difference of two random variables are equal to the difference of their expectations i.e. $\mathbb{E}(X-Y) = \mathbb{E}(X) - \mathbb{E}(Y)$.

Théorème 4. *Let X and Y be random variables on the same sample space S. If X and Y are independent random variables then,* $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

Remarque 2. If X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. However, the converse is not generally true : it is possible for $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ even though X and Y are dependent.

1.2.3. ***Variance and standard deviation of discrete random variables***

The variance is an important characteristic of a random variable. It is a measure of dispersion of the random variable, it measures how much the probability mass is spread out around the mean value.

Définition 6. If X is a random variable with mean $\mathbb{E}(X) = \mu_X$, then the variance of X is defined by

$$Var(x) = \mathbb{E}\left((X-\mu_X)^2\right) = \mathbb{E}(X^2) - \mu_X^2.$$

The standard deviation σ of X is defined by

$$\sigma = \sqrt{Var(X)}$$

1.2.4. Properties of variance

The following properties are most useful for computing variance

1- If X and Y are independent then $Var(X+Y) = Var(X) + Var(Y)$.
2- For a constant c, $Var(cX) = c^2 Var(X)$
3- For a constant c, $Var(X+c) = Var(X)$.

1.3. ***Some discrete distributions***

This section presents some common probability distributions for discrete random variables. In each case, we will define a probability mass function by specifying an explicit formula and we define the characteristics that will help to determine when to use a certain distribution in a given context.

1.3.1. ***Bernoulli distribution***

Bernoulli distribution the simplest model that can be used in many real-life applications. A random experiment that can have only two outcomes 1 or 0, is known as a Bernoulli trial. In other words, the random variable can be 1 with a probability p or it can be 0 with a probability $q = (1-p)$. Typically, a value 1 is assigned to "success" and the value 0 indicates a "failure". The parameter p in the Bernoulli distribution refers to the probability of "success".

Définition 7. A random variable X is governed by the Bernoulli random distribution with parameter p, noted as $X \sim Bernoulli(p)$, if its PMF is as follows

$$p_X(x) = \begin{cases} p & \text{if } x = 1\\ q & \text{if } x = 0\\ 0 & \text{otherwise.}\end{cases}$$

Proposition 2. Suppose X is a Bernoulli random variable with parameter p, then

$$\begin{aligned}\mathbb{E}(X) &= p\\ Var(X) &= pq\end{aligned}$$

1.3.2. ***Binomial distribution***

In general, we can connect binomial random variables to Bernoulli random variables. If X is a binomial random variable, with parameters n and p, then it can be written as the sum of n independent Bernoulli random variables, $X_1, \ldots, X_n$. If we define the random variable X_i, for $i = 1, \ldots, n$, to be 1 when the $i-th$ trial is a "success", and 0 when it is a "failure", then the sum

$$X = X_1 + \cdots + X_n$$

gives the total number of success in n trials.

Définition 8. A random variable X is governed by the binomial distribution with parameters n and p, noted as $X \sim Binomial(n, p)$, if its PMF is as follows

$$p_X(k) = \begin{cases} C_n^k p^k q^{n-k} & \text{for } k = 0, 1, 2, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

where $q = 1 - p$.

Proposition 3. Suppose X is a binomial random variable with parameters n and p, then

$$\mathbb{E}(X) = np$$
$$Var(X) = npq$$

Exemple 7. A basketball player takes 4 independent free throws with a probability of 0.7 of getting a basket on each shot, and let X be the r.v. that counts the number of successes in all trials. This is an example of a Bernoulli experiment with 4 trials, and we write $X \sim Binomial(4, 0.7)$.

Exemple 8. Flip a coin 12 times, and let X be the r.v. that counts the number of heads. This is an example of a Bernoulli Experiment with 12 trials, and we write $X \sim Binomial(12, 0.5)$.

1.3.3. ***Geometric and Pascal distribution***

The geometric and Pascal distributions are related to the binomial distribution in that the underlying probability experiment is the same, i.e., independent trials with two possible outcomes. However, the random variable defined in the geometric and negative binomial case highlights a different aspect of the experiment, namely the number of trials needed to obtain a specific number of "successes". We start with the geometric distribution.

Définition 9. Suppose that a sequence of independent Bernoulli trials is performed, with $p = P(\text{"}success\text{"})$ for each trial. Let X be the random variable that refers to the number of trial at which the first success occurs. Then X is governed by the geometric distribution with parameter p, noted as $X \sim Geometric(p)$ and its PMF is as follows

$$\begin{aligned} p_X(k) &= P(1^{st} \text{ success on } k^{th} \text{ trial}) \\ &= P(1^{st}\ (k-1) \text{ trials are failures } \& \ k^{th} \text{ trial is success}) \\ &= q^{k-1}p, \quad \text{for } k = 1, 2, 3, \ldots \end{aligned}$$

where $q = 1 - p$.

Proposition 4. If X is a Geometric random variable with parameter p, then

$$\mathbb{E}(X) = \frac{1}{p}$$
$$Var(X) = \frac{q}{p^2}$$

Exemple 9. Both of the following examples are random variables governed by the geometric distribution.

1- Toss a fair coin until the first heads occurs. In this case, a "success" is getting a heads and so the parameter $p = 0.5$.

2- Buy lottery tickets until getting the first win. In this case, a "success" is getting a lottery ticket that wins money, and a "failure" is not winning. The parameter p will depend on the odds of wining for a specific lottery.

1.3.4. ***Pascal distribution***

Pascal distribution (negative binomial) generalizes the geometric distribution. It is a repetition of independent trials of random experiments until getting r successes.

Définition 10. Suppose that a sequence of independent Bernoulli trials is performed, with $p = P(\text{"}success\text{"})$ for each trial. Let $r \geq 2$ integer and X be the random variable that refers to the r-th success occurs. Then X is governed by the Pascal distribution with parameters r and p, noted as $X \sim Pascal(r, p)$, if its PMF is as follows

$$\begin{aligned}
P(X = k) &= P(r^{th} \text{ success is on } k^{th} \text{ trial}) \\
&= \underbrace{P(1^{st}\ (r-1) \text{ successes in } 1^{st}\ (k-1) \text{ trials})}_{\text{bniomial with } n=k-1} \times P(r^{th} \text{ success on } k^{th} \text{ trial}) \\
&= C_{r-1}^{k-1} p^{r-1} q^{(k-1)-(r-1)} \times p \\
&= C_{r-1}^{k-1} p^r q^{k-r}, \quad \text{for } k = r, r+1, r+2, \ldots
\end{aligned}$$

Proposition 5. If X is a Pascal random variable with parameters r and p, then

$$\mathbb{E}(X) = \frac{rp}{q}$$
$$Var(X) = \frac{rp}{q^2}$$

Exemple 10. For examples of the Pascal distribution, we can alter the geometric examples given in 9

– Toss a fair coin until get 8 heads. In this case, the parameter p is still given by $p = 0.5$, but now we also have the parameter $r = 8$, the number of desired "successes", i.e., heads.
– Buy lottery tickets until win 5 times. In this case, the parameter p is still given by the odds of winning the lottery, but now we also have the parameter $r = 5$, the number of desired wins.

1.3.5. ***Poison distribution***

The Poisson distribution is a probability distribution that describes the number of events that occur in a fixed interval of time or space, given a known average rate of occurrence and assuming that events happen independently of each other. It's often used in situations where events happen randomly and at a constant average rate, such as the number of phone calls received by a call center in an hour, the number of accidents at a particular intersection in a day, or the number of emails received in an hour. Poisson distribution is used under certain conditions. They are :

- The number of trials $n$tends to infinity
- Probability of success p tends to zero
- the parameter $\lambda = np$ is finite

Définition 11. A random variable X is governed by the Poisson distribution with parameters λ (the average rate of events per interval), noted as $X \sim Poisson(\lambda)$, if its PMF is as follows

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Proposition 6. Suppose X is a Poisson random variable with parameter λ, then

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

Exemple 11. Suppose on average there are 5 homicides per month in a given city. What is the probability there will be at most 1 in a certain month ?
Solution If X is the number of homicides, we are given that $\lambda = 5$. Therefore

$$P(X \leq 1) = P(X = 0) + P(X = 1) = e^{-5} + 5e^{-5}.$$

Exemple 12. Assume on average there is one large earthquake per year in a country. What is the probability that next year there will be exactly 2 large earthquakes ?
Solution :
We have $\lambda = 1$, then $P(X = 2) = e^{-1}\frac{1}{2}$.