

Freebooks.pk

Free Notes • Past Papers • Guides for Students

ICS Part-I Statistics — Punjab Board

CHAPTER 7

Random Variables

Statistics • 1st Year • ICS • Punjab Board • Free PDF Notes



Chapter 7: Random Variables

When a pair of dice is thrown, we're rarely interested in the raw outcome like (3, 5) -- we care about a number derived from it, such as the total of the two upturned faces, which could be anywhere from 2 to 12. That derived number is a random variable: a rule that assigns a real number to every outcome of a random experiment. This chapter formalizes that idea, introduces how random numbers are generated and used in simulation (Monte Carlo methods), and draws the essential distinction between discrete random variables (which take isolated values, like counts) and continuous random variables (which take any value in an interval, like measurements).

Random variables are the bridge between the abstract sample spaces of Chapter 6 and the probability distributions built in Chapter 8 -- once outcomes are converted into numbers, we can describe their behaviour using tables, formulas, means, and variances, the same tools used throughout the rest of statistics.

Learning Objectives

- Define a random variable and identify examples from real experiments
- Understand what random numbers are and how the pseudo-random number generation formula works
- Use a random number table to simulate outcomes of simple experiments (coin tosses, etc.)
- Explain the role of random numbers in simulation (Monte Carlo methods)
- Construct random variables from experiments in different fields and list their possible values
- Distinguish between discrete and continuous random variables with examples

Key Concepts

7.1 Introduction to Random Variables

Every random experiment produces two or more outcomes, and usually we are interested in some particular numerical aspect of those outcomes rather than the



raw outcome itself. When a pair of dice is thrown, for instance, the interest may lie in the total of the two upturned faces -- a value that could be 2, 3, 4, and so on up to 12. Likewise, if we record the number of children in each of fifty randomly chosen families (assuming no family has more than 5), the values of interest are 0, 1, 2, 3, 4, or 5.

A variable whose values depend upon the outcomes of a random experiment is called a random variable. Random variables are denoted by capital letters such as X , Y , or Z , and the specific values they take are denoted by the corresponding small letters x , y , or z . For example, if a pair of dice is thrown and Y denotes the sum of the two upturned values, then Y assigns the outcome (1,1) the value $1+1=2$, the outcome (2,1) the value $2+1=3$, and so on up to the outcome (6,6), which gets the value $6+6=12$ -- so Y can take any of the values 2 through 12.

7.2 Random Numbers and Their Generation

Random numbers are a sequence of digits from the set $\{0,1,2,\dots,9\}$ such that at every position in the sequence, each digit has the same probability of $1/10$ (0.1) of being selected, regardless of the digits already generated. These are also called random digits. The simplest way to produce them is through games of chance -- dice, coins, cards, or drawing numbered slips from a hat -- but this becomes tedious for long sequences, so printed tables of random digits are widely used instead.

For computer implementation, the most common approach is the pseudo-random technique. These numbers are called 'pseudo' because they are not truly random -- they are generated by a completely deterministic recursive formula, yet they exhibit most of the statistical properties of genuinely random digits. Eventually the sequence repeats itself (a cycle), but for a good generator this cycle can be tens of thousands of digits long. The standard recursive formula is $x(n+1) = [a \cdot x(n) + b] \bmod m$, where a , b , and m are suitably chosen integer constants and $x(1)$ (the seed) is a chosen starting integer. 'mod m ' means: if the result exceeds m , divide by m and keep only the remainder. For the method to be useful, m , a , and b should be chosen reasonably large.



7.3 Application of Random Numbers

Random numbers are widely used in simulation techniques, also called Monte Carlo methods, which are applied across the sciences whenever direct experimentation is impossible, too costly, or too time-consuming. Random number tables group digits (in 2, 3, 4, or 5-digit sequences) purely for ease of reading, and can be read row-wise or column-wise depending on how many digits are needed for a given problem.

To simulate a coin toss using a random number table: assign even digits (0,2,4,6,8) to represent heads and odd digits (1,3,5,7,9) to represent tails, then read single digits from an arbitrarily chosen row and column of the table, converting each digit to H or T according to the assigned rule. To simulate two coins at once, two-digit numbers are read: both digits even means both heads, both odd means both tails, and one of each means exactly one head. When outcomes have unequal probabilities (rather than the equally likely 50-50 split of a fair coin), a related method called probability proportional to size (PPS) is used: cumulative probabilities are computed for each possible outcome, and a corresponding range of random numbers is assigned to each outcome in proportion to its probability, so that a randomly selected number automatically falls into the correct class the right proportion of the time.

7.4 Constructing Random Variables from Different Fields

Random variables can be built from experiments across almost any field, and the values they take depend entirely on how the variable is defined. Consider three students each choosing between Biology (B) and Computer Science (C): if Y is defined as the number of students taking computer science, the eight equally likely outcomes (BBB, CBB, BCB, BBC, CCB, CBC, BCC, CCC) get mapped to the values 0, 1, 1, 1, 2, 2, 2, and 3 respectively -- so Y takes only the isolated values 0, 1, 2, and 3.

By contrast, consider recording the time (in minutes) a customer waits in a queue at a utility store: this random variable can take any value in an interval -- 5.0 minutes, 5.3 minutes, 12.0 minutes, or anything in between -- rather than a fixed set of separate values. This contrast between a variable that jumps



between isolated points and one that can take any value in a range is exactly the distinction formalized in the next section.

7.5 Discrete and Continuous Random Variables

A random variable is called discrete if the set of values it can take is a collection of isolated points on the real number line -- that is, its sample space is discrete. The outcomes of the experiment are noted, and a number is assigned to each outcome by some rule. For example, if three coins are tossed and Y is the number of heads, then Y takes only the whole-number values 0, 1, 2, or 3 -- a discrete random variable. The 'number of students taking computer science' example above is also discrete.

A random variable is called continuous if the set of values it can take is an entire interval on the number line -- its sample space is continuous. The outcomes are represented by points on a line, and a number is assigned to each point by some rule. For example, if the heights of students in a statistics class range from a minimum of 5.0 feet to a maximum of 5.8 feet, the height variable Y can take any value in the interval 5.0 to 5.8 feet -- a continuous random variable. The customer-waiting-time example above is also continuous. In short: discrete random variables almost always arise from counting, while continuous random variables arise from measuring.

Important Definitions

What is a random variable?: A variable whose values depend on the outcomes of a random experiment; denoted by capital letters (X , Y , Z) with specific values denoted by the corresponding small letters.

What are random numbers?: A sequence of digits from $\{0,1,\dots,9\}$ such that each digit at each position has the same probability, $1/10$, of being selected, regardless of prior digits.

What is a pseudo-random number?: A number generated by a deterministic recursive formula (such as $x(n+1) = [a \cdot x(n) + b] \bmod m$) that is not truly random but exhibits the statistical properties of random digits, eventually repeating in a long cycle.



What is a Monte Carlo method?: A simulation technique that uses random numbers to study problems where direct experimentation is impossible, too costly, or too time-consuming.

What is a discrete random variable?: A random variable whose possible values form a collection of isolated points on the number line, typically arising from counting.

What is a continuous random variable?: A random variable whose possible values form an entire interval on the number line, typically arising from measurement.

What is probability proportional to size (PPS)?: A simulation method that assigns a range of random numbers to each possible outcome in proportion to its probability, using cumulative probabilities to define the ranges.

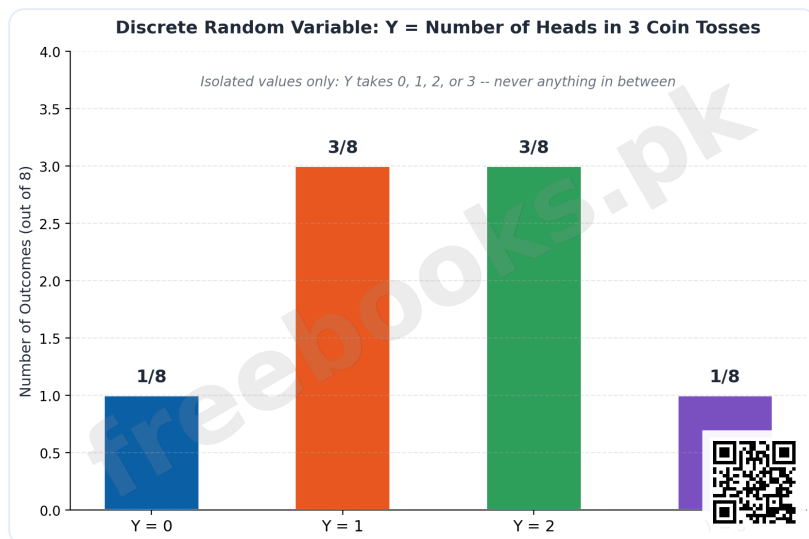
Key Facts and Relations

Topic	Key Fact / Relation
Pseudo-random number generator	$x(n+1) = [a.x(n) + b] \text{ mod } m$
Random digit probability	$P(\text{any single digit } 0-9) = 1/10 = 0.1$
Discrete random variable	Values are isolated points (arise from counting)
Continuous random variable	Values fill an entire interval (arise from measuring)
PPS cumulative range rule	Random number range assigned to a class is proportional to its cumulative probability

Diagrams

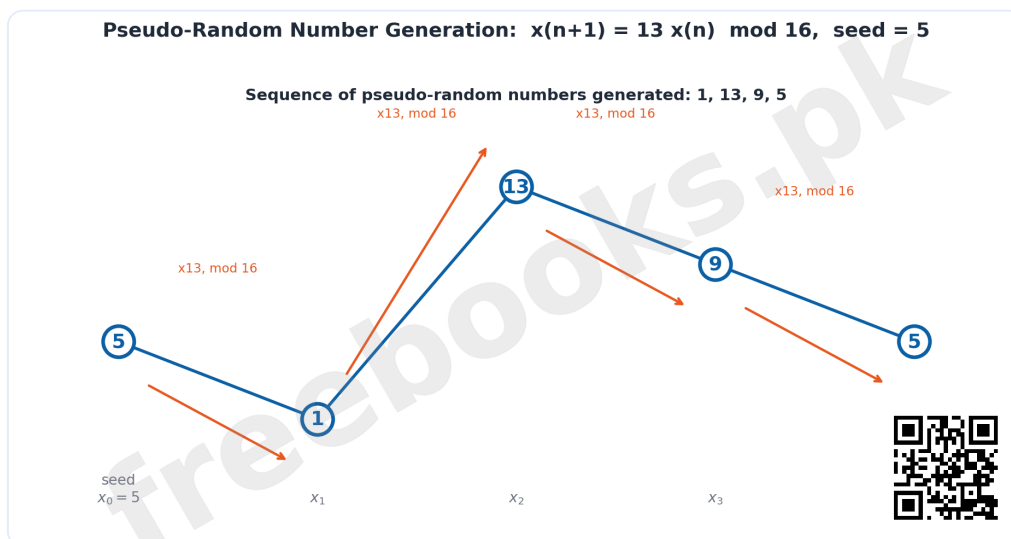
Discrete Random Variable: Number of Heads in 3 Coin Tosses: A probability mass bar chart for $Y = \text{number of heads when 3 coins are tossed}$, showing the 8 equally likely outcomes mapping to isolated values 0, 1, 2, 3 with their frequencies, illustrating a discrete random variable





Discrete Random Variable: Number of Heads in 3 Coin Tosses

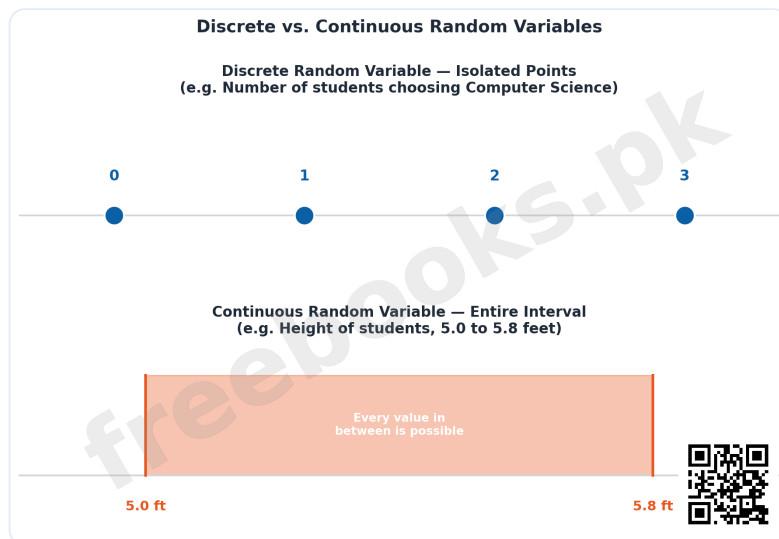
Pseudo-Random Number Generation (LCG Sequence): A step-by-step visualization of the linear congruential generator $x(n+1) = 13x(n) \bmod 16$ starting from seed 5, tracing the sequence 5 → 1 → 13 → 9 → 5, matching Example 7.2 from the textbook



Pseudo-Random Number Generation (LCG Sequence)

Discrete vs Continuous Random Variables: A side-by-side comparison: isolated dots at 0,1,2,3 representing a discrete random variable (number of computer-science students) versus a shaded continuous interval from 5.0 to 5.8 representing a continuous random variable (student heights)





Discrete vs Continuous Random Variables

Short Questions & Answers

Why are pseudo-random numbers called 'pseudo'?

Because they are produced by a completely deterministic recursive formula rather than a truly random physical process, even though they display most of the statistical properties of genuine random digits.

Give an example of a discrete random variable and a continuous random variable.

Discrete: the number of heads in three coin tosses (values 0,1,2,3). Continuous: the height of students in a class, which can take any value in an interval such as 5.0 to 5.8 feet.

Why is the number of children in a family a discrete random variable?

Because its possible values (0,1,2,3,...) are isolated whole numbers arising from counting, not from measurement over a continuous interval.

What role do random numbers play in Monte Carlo simulation?

They allow researchers to imitate the outcomes of a random experiment on paper or by computer, which is useful when direct experimentation is impossible, too expensive, or too time-consuming.

In the pseudo-random formula $x(n+1) = [a \cdot x(n) + b] \bmod m$, what does 'mod m' mean?

It means that if $a \cdot x(n) + b$ exceeds m , you divide the result by m and keep only the remainder as the next random number in the sequence.



Why must a discrete random variable's sample space consist of isolated points?

Because the variable arises from counting outcomes, which can only take specific whole-number (or otherwise separated) values, with no possible values existing between them.

Long Questions & Answers

Explain what a random variable is, and describe with examples how random variables can be constructed from experiments in different fields.

A random variable is best understood as a bridge between the raw outcomes of a random experiment and the numerical values we actually want to analyze, since most of the time we are not directly interested in an outcome itself but in some measurable quantity that outcome produces. Formally, a random variable is defined as a variable whose values depend upon the outcomes of a random experiment, and by convention it is represented using a capital letter such as X , Y , or Z , while its specific realized values are written using the corresponding lowercase letter x , y , or z . The reason this concept matters so much in statistics is that random experiments on their own -- rolling dice, tossing coins, or interviewing families -- don't come with built-in numbers; it is the random variable that supplies the rule for converting each outcome into a number we can then summarize, average, or build a probability distribution around.

Consider, for example, an experiment in which three students in a class are each asked to choose between two courses, Biology or Computer Science. On its own, an outcome of this experiment is simply a string like 'first student picks Biology, second picks Computer Science, third picks Biology' -- not yet a number. To turn this into a random variable, we might define Y as the number of students who choose computer science. Once this rule is fixed, every one of the eight equally likely outcomes of the experiment can be mapped onto a specific value of Y : the all-Biology outcome maps to $Y=0$, each of the three outcomes with exactly one computer-science student maps to $Y=1$, each of the three outcomes with exactly two computer-science students maps to $Y=2$, and the all-computer-science outcome maps to $Y=3$. Notice that the random variable has compressed eight distinct outcomes down into just four possible numerical values, which is



precisely what makes it useful: instead of tracking every possible combination of student choices, we can now describe the situation purely in terms of the values 0, 1, 2, and 3 and how likely each one is. A completely different field can produce an equally valid but very differently shaped random variable. Consider an experiment where we record, for each customer arriving at a utility store, the number of minutes they wait in a queue before being served. Here, we might define Y as the waiting time of a given customer. Unlike the course-selection example, this random variable cannot be reduced to a small fixed list of possible values -- a customer might wait 5.0 minutes, or 5.3 minutes, or 12.0 minutes, or any other value that a stopwatch could register, and there is no natural gap between one possible waiting time and the next. In both cases, notice that the underlying logic of constructing a random variable is identical regardless of the field: first, identify the random experiment; second, decide precisely what numerical aspect of the outcome is of interest; and third, define the rule that assigns exactly one number to every possible outcome of that experiment. What differs entirely from field to field, and indeed from definition to definition even within the same experiment, is the resulting set of possible values the random variable can take -- sometimes a short list of isolated whole numbers, as with counting how many students chose a particular course, and sometimes an entire continuous range of values, as with measuring how long a customer waits. This very difference in the character of the resulting values is what leads directly into the distinction between discrete and continuous random variables, since the same three-step construction process applied to a counting-based question versus a measurement-based question naturally produces these two fundamentally different types of variables.

Explain the difference between discrete and continuous random variables, and describe how random numbers are generated and used in simulation.

Once a random variable has been defined for a given experiment, the next essential step is to understand what kind of values it can actually take, and this is where the distinction between discrete and continuous random variables becomes essential to the rest of the subject. A random variable is called discrete if the complete set of values it can take forms a collection of isolated points on the real number line -- in other words, its underlying sample space is itself



discrete, and the outcomes of the experiment are noted and then converted into a number through some assignment rule. A classic illustration is tossing three coins and letting Y represent the number of heads observed: working through all eight possible outcomes (HHH, HHT, HTH, HTT, THH, THT, TTH, TTT) and counting the heads in each shows that Y can only ever equal 0, 1, 2, or 3 -- four separated whole-number values with nothing meaningful existing in between them, such as 1.5 heads. Because these values arise from a simple counting process, discrete random variables almost always show up whenever a problem asks 'how many' of something occurred. In sharp contrast, a random variable is called continuous if the complete set of values it can take forms an entire interval on the number line rather than a set of separated points -- here, the outcomes of the experiment correspond to points along a continuous line, and a number is assigned to each point by some appropriate rule. A clear illustration is measuring the heights of students in a statistics class: if the shortest student measures 5.0 feet and the tallest measures 5.8 feet, then the height variable can, in principle, take on any value at all within that interval -- 5.1, 5.23, 5.457, or any other real number the measuring instrument is capable of registering -- with no natural gaps between one achievable value and the next. Because continuous random variables arise from measurement rather than counting, they show up whenever a problem asks 'how much' or 'how long' rather than 'how many.' This distinction matters immensely for everything that follows in statistics, because discrete and continuous random variables are eventually described using entirely different mathematical tools -- probability mass functions for the discrete case and probability density functions for the continuous case. Turning to how experiments involving randomness are actually simulated on paper or by computer, this relies on the concept of random numbers: a sequence of digits drawn from the set 0 through 9, chosen so that at every single position in the sequence, each of the ten digits has exactly the same probability, one-tenth, of appearing, completely independent of whatever digits came immediately before it. While physical devices such as dice, coins, or numbered slips drawn from a hat can generate short bursts of genuinely random digits, this quickly becomes impractical for producing the long sequences that real simulation work requires, which is why computers instead rely on what are called pseudo-random numbers. These numbers are termed 'pseudo' precisely because they are not, in the strictest sense, actually random at all -- they are produced by a completely deterministic



mathematical formula that will, given the exact same starting point, always reproduce the exact same sequence of numbers every single time it is run. Despite this underlying determinism, however, a well-designed pseudo-random number generator produces sequences that convincingly display all of the important statistical properties expected of true randomness, and the sequence will only begin visibly repeating itself after an extremely long stretch, or cycle, of generated digits -- for a properly designed generator this cycle can run into the tens of thousands of digits before any repetition becomes apparent. The specific recursive formula almost universally used for this purpose is $x(n+1)$ equals the quantity a times $x(n)$ plus b , all taken modulo m , where a , b , and m are constants chosen carefully in advance by whoever designs the generator, and $x(1)$, referred to as the seed, is simply the starting integer that kicks the entire sequence into motion. The 'modulo m ' operation appearing in this formula means that whenever the raw calculated result of $a \cdot x(n) + b$ happens to exceed m , that result is divided by m and only the remainder of that division is retained and carried forward as the next number in the generated sequence. Once a long sequence of these pseudo-random digits has been generated, it becomes the essential raw material for Monte Carlo simulation methods -- techniques that use random numbers to imitate, and thereby study, the behaviour of random experiments in situations where actually carrying out the real physical experiment would be genuinely impossible, prohibitively expensive, or simply take far too long to be practical, such as simulating decades of customer arrivals at a business in a matter of seconds on a computer rather than genuinely waiting for that much real time to pass.

Multiple Choice Questions (MCQs)

A random variable is best described as: (A) A fixed constant used in every experiment (B) A variable whose values depend on the outcomes of a random experiment (C) Only the outcomes of a coin toss (D) A type of sample space

Correct answer: (B) A variable whose values depend on the outcomes of a random experiment. A random variable assigns a numerical value to each outcome of a random experiment.



Random numbers are digits from the set $\{0, \dots, 9\}$ where each digit has probability: (A) 1 (B) 0.5 (C) 0.1 (D) 0

Correct answer: (C) 0.1. Each of the 10 digits has an equal probability of $1/10 = 0.1$ of being selected at any position.

Pseudo-random numbers are called 'pseudo' because they are: (A) Truly random (B) Generated by a deterministic formula, not a truly random process (C) Always irrational numbers (D) Only usable once

Correct answer: (B) Generated by a deterministic formula, not a truly random process. Pseudo-random numbers come from a deterministic recursive formula, yet they mimic the statistical behaviour of true random digits.

The general pseudo-random number generator formula is: (A) $x(n+1) = a + b \cdot x(n)$ (B) $x(n+1) = [a \cdot x(n) + b] \bmod m$ (C) $x(n+1) = x(n)/m$ (D) $x(n+1) = a \cdot m + b$

Correct answer: (B) $x(n+1) = [a \cdot x(n) + b] \bmod m$. The standard linear congruential formula is $x(n+1) = [a \cdot x(n) + b] \bmod m$.

Simulation techniques that use random numbers are also known as: (A) Linear regression methods (B) Monte Carlo methods (C) Least squares methods (D) Index number methods

Correct answer: (B) Monte Carlo methods. Simulation using random numbers is widely known as the Monte Carlo method.

The number of heads obtained in 3 coin tosses is an example of a: (A) Continuous random variable (B) Discrete random variable (C) Constant (D) Sample space

Correct answer: (B) Discrete random variable. It takes only isolated whole-number values (0,1,2,3), making it a discrete random variable.

The height of students in a class, measured in feet, is an example of a: (A) Discrete random variable (B) Continuous random variable (C) Random number (D) Pseudo-random variable

Correct answer: (B) Continuous random variable. Height can take any value within an interval, making it a continuous random variable.

Discrete random variables typically arise from: (A) Measuring (B) Counting (C) Rounding (D) Sampling error



Correct answer: (B) Counting. Discrete random variables almost always arise in connection with counting.

Continuous random variables typically arise from: (A) Counting (B) Measurement (C) Guessing (D) Random digit tables only

Correct answer: (B) Measurement. Continuous random variables are typically obtained by measurement, taking any value in an interval.

In probability proportional to size (PPS), random numbers are assigned to classes based on: (A) Alphabetical order (B) Cumulative probabilities of each class (C) Random guessing (D) The size of the sample space only

Correct answer: (B) Cumulative probabilities of each class. PPS assigns a range of random numbers to each class proportional to its cumulative probability.

Quick Revision Summary

- Random variable: assigns a real number to each outcome of a random experiment; denoted X, Y, Z (values: x, y, z)
- Random numbers: digits 0-9, each with probability 0.1 at every position, independent of prior digits
- Pseudo-random generator: $x(n+1) = [a \cdot x(n) + b] \bmod m$, deterministic but statistically random-like
- Random number tables read row-wise or column-wise; digit grouping (1,2,3...digit) depends on the problem
- Monte Carlo methods: simulate random experiments using random numbers when real experimentation is costly/impossible
- PPS method: cumulative probabilities define a range of random numbers assigned to each outcome class
- Discrete random variable: isolated values, arises from counting (e.g., number of heads)
- Continuous random variable: any value in an interval, arises from measuring (e.g., height, waiting time)



Exam Tips

- To classify a random variable quickly, ask: 'Am I counting or measuring?'
Counting = discrete, measuring = continuous
- In LCG problems, always compute step by step: multiply, add b , then take mod m (divide and keep remainder)
- When reading random number tables, note whether the problem needs 1-digit, 2-digit, or 3-digit groups before you start reading
- For PPS problems, build the cumulative probability column first, then convert cumulative probabilities into number ranges (0 to 999 for 3-decimal probabilities)
- Remember: a random variable is a function/rule, not the outcome itself — always double-check what rule assigns values to outcomes

