Freebooks.pk

Free Notes • Past Papers • Guides for Students

ICS Part-I Statistics — Punjab Board

CHAPTER 8

Probability Distributions

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

Chapter 8: Probability Distributions

Chapter 7 introduced the random variable as a rule that assigns a number to every outcome of a random experiment. But knowing the possible values isn't enough -- to actually use a random variable for prediction and decision-making, we need to know how likely each value (or range of values) is. That complete description -- every possible value paired with its probability -- is called a probability distribution.

This chapter builds the two core tools for describing probability distributions: the probability mass function (PMF) for discrete random variables and the probability density function (PDF) for continuous random variables. It also introduces the expected value (mean) and variance of a random variable, ways to graph a distribution, and the cumulative distribution function, which answers 'what is the probability that Y is at or below some value?' -- laying the groundwork for the specific distributions (binomial, hypergeometric) covered in Chapter 9.

Learning Objectives

- Define probability mass function (PMF) and probability density function (PDF)
- State and verify the two conditions a function must satisfy to be a valid PMF or PDF
- Construct probability distributions for discrete random variables from real experiments
- Calculate probabilities for continuous random variables as areas under a density curve
- Draw a probability histogram and a bar chart for a discrete probability distribution
- Calculate the expected value (mean) and variance of a discrete random variable
- Apply the properties of expectation to find E(constant), E(bY plus a), E(X+Y), and E(XY)
- Define and construct the cumulative distribution function (CDF) of a discrete random variable

Key Concepts

8.1 Introduction

Whenever we work with random experiments, we need to associate a numerical value with each outcome in order to study it -- this gives rise to two types of random variables: discrete and continuous. A discrete random variable almost always arises from counting, while a continuous random variable arises from measurement. For a discrete random variable, its probability distribution describes how much probability is placed on each possible value, with all these probabilities summing to 1; this is called its probability mass function. For a continuous random variable, we cannot talk about probability at a single point -- instead we talk about probability over an interval, with total probability again equal to 1; this is called its probability density function.

A probability distribution can be described either by a formula (function) or with a two-column table, similar to a frequency distribution -- one column lists the values (or intervals of values, for continuous variables) and the other lists the corresponding probabilities.

8.2 Probability Mass Function

The probability mass function (PMF) of a discrete random variable Y describes the values Y can take and the probability associated with each value, usually presented in a two-column table. For example, if three coins are tossed and Y denotes the number of heads, there are 8 equally likely outcomes: no head occurs once (TTT), one head occurs 3 times (HTT, THT, TTH), two heads occurs 3 times (HHT, HTH, THH), and three heads occurs once (HHH). This gives the probability function $P(Y=0)=1/8$, $P(Y=1)=3/8$, $P(Y=2)=3/8$, $P(Y=3)=1/8$.

8.3 Probability Density Function

The probability density function (PDF) of a continuous random variable Y is specified by a smooth curve such that the total area under the curve equals 1. The probability that Y falls within a particular interval is the area under the curve above that interval -- unlike the discrete case, we never ask for the probability at a single exact point (that probability is always zero for a continuous variable), only for probability over a range.

8.4 Discrete and Continuous Distributions, and Their Properties

A discrete uniform distribution arises when every possible value of a discrete random variable has the same probability -- for example, when a fair die is thrown, each face 1 through 6 has probability 1/6, giving $f(y) = 1/6$ for $y=1,2,\ldots,6$ and 0 otherwise. A continuous uniform distribution is defined by $f(y) = 1/(b-a)$ for $a < y < b$, meaning the probability is spread evenly across the interval from a to b; when $a=0$ and $b=1$, this simplifies to $f(y)=1$ for $0<y<1$.

For a probability mass function $P(y)$ of a discrete random variable Y to be valid, it must satisfy two conditions: first, $0 <= P(y) <= 1$ for every possible value of Y (probabilities are always between 0 and 1); second, the sum of $P(y)$ over all possible values of Y must equal exactly 1. For a probability density function $f(y)$ of a continuous random variable Y to be valid, it must satisfy: first, $f(y) >= 0$ for all y; second, the total area under the curve, P(-infinity < Y < infinity), must equal exactly 1. Not every function defined over the values of a random variable automatically qualifies as a probability distribution -- both conditions must be checked and confirmed.

An important distinction: for discrete distributions, $P(a <= Y <= b)$ is NOT equal to $P(a < Y < b)$, because the endpoint probabilities $P(Y=a)$ and $P(Y=b)$ are real, nonzero quantities that must be included or excluded depending on the inequality used. For continuous distributions, however, these two probabilities ARE equal, because the probability at any single exact point is always zero (the 'area' at a single point has zero width).

8.5 Drawing the PMF and PDF

A discrete probability distribution can be drawn as a probability histogram: values of the random variable go on the x-axis, probabilities on the y-axis, and adjacent rectangles are drawn with width 1 (extending 0.5 units to each side of the value) and height equal to the probability at that point -- since the width is 1, the area of each rectangle equals its probability. Alternatively, a bar chart can be drawn with thin bars (rather than full-width rectangles) whose height still equals the probability of the corresponding value. When the values of a discrete random variable become very closely spaced, the probability histogram can be approximated by a smooth curve, which is exactly the idea behind a continuous probability density function.

8.6 Expectation and Variance of a Discrete Random Variable

The mathematical expectation, or expected value, of a discrete random variable Y with probability function P(y) is defined as E(Y) = sum of y.P(y) over all values of Y. This expectation is simply the mean of the probability distribution -- E(Y) is an alternative notation for the population mean, mu. Similarly, E(Y-squared) = sum of y-squared.P(y) over all values of Y.

The variance of a random variable Y is defined as sigma-squared = E[(Y - mu)-squared] = sum of (Y-mu)-squared.P(y) over all Y. Through algebraic expansion, this simplifies to the more practical computing formula: Var(Y) = E(Y-squared) - [E(Y)]-squared -- exactly analogous to the shortcut variance formula used for raw data in Chapter 4.

Four key properties of expectation make calculations much easier: (i) if c is a constant, E(c) = c; (ii) if a and b are constants, E(bY plus-or-minus a) = b.E(Y) plus-or-minus a (setting a = -mu and b = 1 gives the useful result E(Y - mu) = 0); (iii) for two random variables X and Y, the expectation of their sum or difference is the sum or difference of their expectations, E(X plus-or-minus Y) = E(X) plus-or-minus E(Y); (iv) if X and Y are independent, the expectation of their product is the product of their expectations, E(XY) = E(X).E(Y).

8.9 Distribution Function (Cumulative Distribution Function)

Very often we want the probability that a random variable takes a value at or below some fixed point -- for example, the chance a student scores no more than 80%, or that 5 coin tosses produce no more than 3 heads. This probability, P(Y <= y), is called the distribution function, or cumulative distribution function (CDF), and is written F(y) = P(Y <= y). For a continuous random variable, F(y) is the integral of the density function from negative infinity up to y.

Every valid CDF satisfies four properties: (i) F(-infinity) = 0 (there's no probability of being below the smallest possible value); (ii) F(+infinity) = 1 (there's certainty of being at or below the largest possible value); (iii) F(y1) <= F(y2) whenever y1 <= y2 (the CDF is non-decreasing); (iv) F(y) is continuous at least from the right at every point. For a discrete random variable, the CDF is a step function: it jumps up by P(Y=y) at each possible value of y and stays flat (constant) in between consecutive possible values.

Important Definitions

What is a probability distribution?: A complete description of a random variable's possible values together with their associated probabilities, presented as a formula or a two-column table.

What is a probability mass function (PMF)?: A function P(y) that gives the probability associated with each possible value of a discrete random variable Y, satisfying 0<=P(y)<=1 and the sum of all P(y) equal to 1.

What is a probability density function (PDF)?: A smooth curve f(y) for a continuous random variable such that f(y)>=0 everywhere and the total area under the curve equals 1; probability corresponds to area over an interval, not a value at a single point.

What is the expected value E(Y) of a discrete random variable?: The mean of its probability distribution, calculated as the sum of y.P(y) over all possible values of Y.

What is the variance of a random variable?: The expected value of the squared deviation from the mean, Var(Y) = E[(Y-mu)^2], which simplifies to E(Y^2) - [E(Y)]^2.

What is a discrete uniform distribution?: A distribution in which every possible value of a discrete random variable has the same probability, such as each face of a fair die having probability 1/6.

What is a cumulative distribution function (CDF)?: The function F(y) = P(Y<=y) giving the probability that a random variable takes a value at or below y; for discrete variables it is a step function.

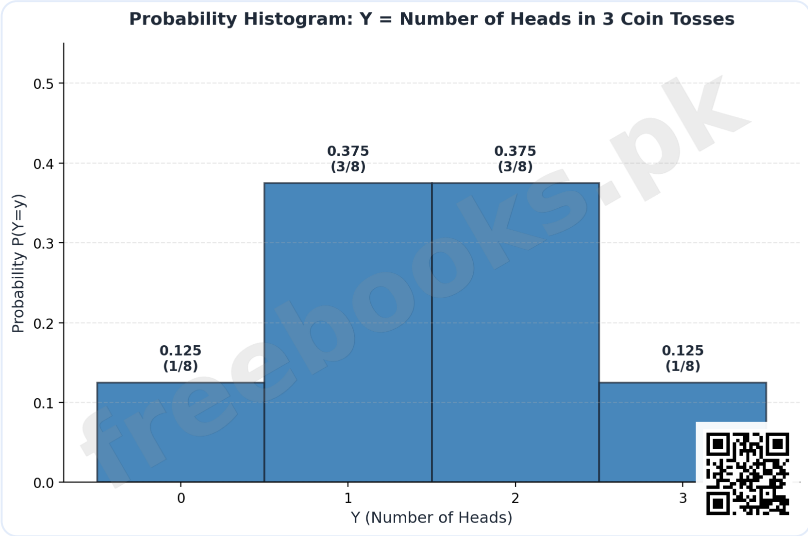
Why is P(Y=a) always zero for a continuous random variable?: Because probability for a continuous variable corresponds to area under the density curve, and the area over a single point (zero width) is always zero.

Key Facts and Relations

Topic	Key Fact / Relation
PMF validity conditions	$0 <= P(y) <= 1$ for each y, and sum of $P(y)$ over all $y = 1$
PDF validity conditions	$f(y) >= 0$ for all y, and total area under curve $= 1$
Discrete uniform distribution (die)	$f(y) = 1/6$ for $y = 1,2,...,6$
Continuous uniform distribution	$f(y) = 1/(b-a)$ for $a < y < b$
Expected value (discrete)	$E(Y) = \text{sum } y.P(y)$
Variance (discrete, computing formula)	$Var(Y) = E(Y^2) - [E(Y)]^2$
Expectation of a constant	$E(c) = c$
Linear transformation of expectation	$E(bY + a) = b.E(Y) + a$
Expectation of sum/difference	$E(X +/- Y) = E(X) +/- E(Y)$
Expectation of product (independent X, Y)	$E(XY) = E(X).E(Y)$
Cumulative distribution function	$F(y) = P(Y <= y)$
Trapezoidal area rule (continuous probability)	$P(a<Y<b) = [f(a)+f(b)]/2 \times (b-a)$ (for linear density functions)

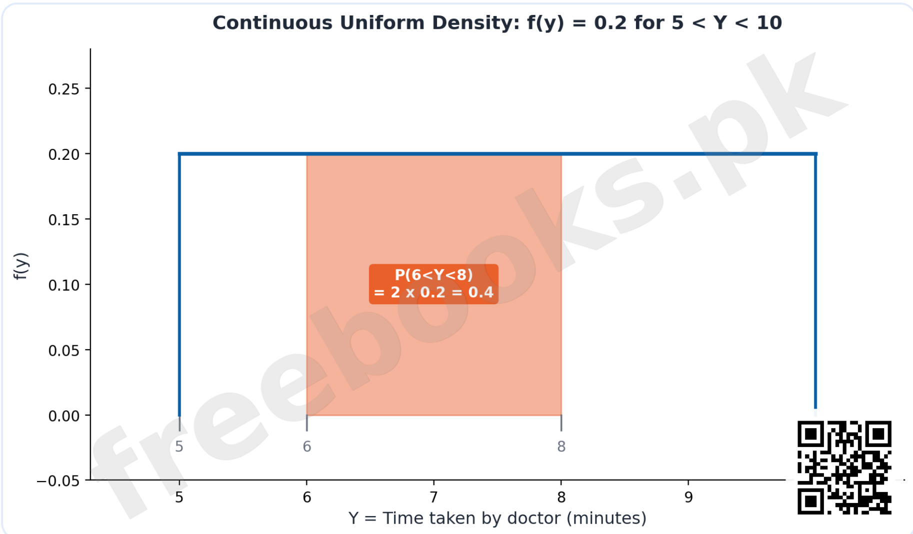
Diagrams

Probability Histogram: Number of Heads in 3 Coin Tosses: A probability histogram for the discrete random variable Y = number of heads in 3 coin tosses, with rectangles of width 1 at Y=0,1,2,3 and heights 1/8, 3/8, 3/8, 1/8 respectively, illustrating a valid probability mass function



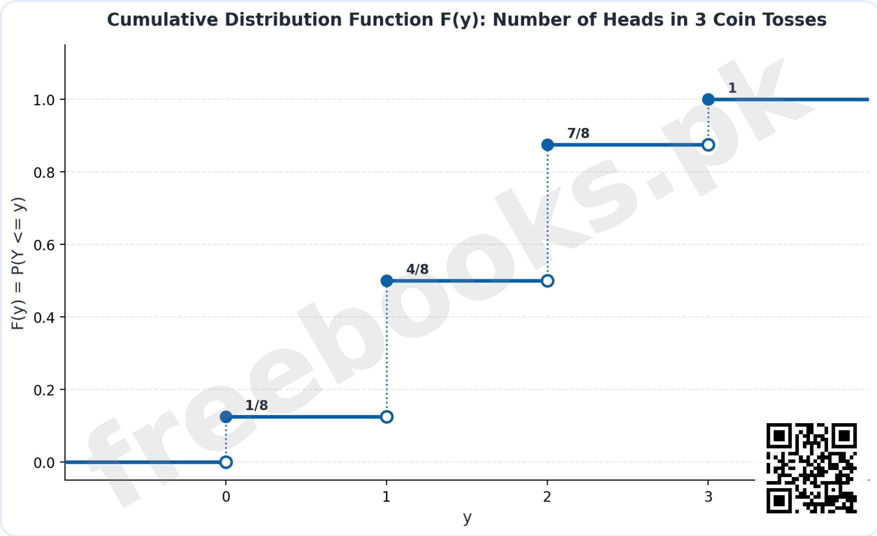
Probability Histogram: Number of Heads in 3 Coin Tosses

Continuous Uniform Density: Doctor's Waiting Time: The continuous uniform density function f(y)=0.2 for 5<y<10 minutes (doctor's time with a patient), with the area between y=6 and y=8 shaded to show P(6<Y<8)=0.4, illustrating that continuous probabilities are areas under the curve



Continuous Uniform Density: Doctor's Waiting Time

Cumulative Distribution Function: 3 Coin Tosses: The step-function cumulative distribution function F(y)=P(Y<=y) for the number of heads in 3 coin tosses, jumping from 0 to 1/8 to 4/8 to 7/8 to 1 at Y=0,1,2,3, illustrating the CDF of a discrete random variable



Cumulative Distribution Function: 3 Coin Tosses

Short Questions & Answers

What two conditions must a function satisfy to be a valid probability mass function?

It must give a probability between 0 and 1 for each possible value of the random variable, and the sum of all these probabilities over every possible value must equal exactly 1.

Why can't we ask for the exact probability that a continuous random variable equals a specific value?

Because probability for continuous variables is measured as area under the density curve, and the area above a single point has zero width, so that probability is always exactly zero.

What is the difference between a probability histogram and a bar chart for a discrete distribution?

In a probability histogram, adjacent rectangles of width 1 are drawn so that area equals probability; in a bar chart, thin bars are drawn whose height alone (not area) represents the probability.

State the formula that relates variance to E(Y) and E(Y^2).

Var(Y) = E(Y^2) - [E(Y)]^2, the computing (shortcut) formula for variance.

If E(X) = 5, what is E(3X + 2)?

Using E(bY+a) = b.E(Y)+a: E(3X+2) = 3(5)+2 = 17.

What does the cumulative distribution function F(y) represent?

F(y) = P(Y<=y), the probability that the random variable takes a value at or below y.

Why is a discrete random variable's CDF a step function rather than a smooth curve?

Because probability only accumulates at the isolated points the discrete variable can take, so F(y) jumps up at each possible value and stays flat in between.

Long Questions & Answers

Explain the concept of a probability mass function and a probability density function, including the conditions each must satisfy, and explain why the treatment of probability differs between discrete and continuous random variables.

Once a random variable has been defined for a given experiment, the natural next question is not just what values it can take, but how likely each of those values actually is -- and answering that question completely, for every possible value at once, is exactly what a probability distribution provides. For a discrete random variable, this complete description is called the probability mass function, commonly abbreviated as PMF, and it works by pairing up each possible value of the random variable with its own specific probability, usually organized into a simple two-column table with the values in one column and their corresponding probabilities in the other. A textbook illustration of building a PMF from scratch involves tossing three fair coins and letting Y represent the number of heads observed. Working carefully through the eight equally likely outcomes of this experiment -- TTT, HTT, THT, TTH, HHT, HTH, THH, and HHH -- shows that exactly one outcome produces zero heads, exactly three outcomes produce one head, exactly three outcomes produce two heads, and exactly one outcome produces three heads. Applying the classical definition of probability, this translates directly into P(Y=0)=1/8, P(Y=1)=3/8, P(Y=2)=3/8, and P(Y=3)=1/8, and together these four probability statements constitute the complete probability mass function for this experiment. Not every mathematical function that happens to be defined over a random variable's possible values, however, is automatically entitled to be called a valid probability mass function -- two strict conditions must always hold before a function earns that label. First, every single probability the function produces must fall somewhere between 0 and 1 inclusive, since a negative probability or a probability exceeding 1 has no meaningful interpretation in the real world. Second, and equally essential, when all of the probabilities across every single possible value of the random variable are added together, that total sum must come out to exactly 1, reflecting the simple logical fact that the random variable is guaranteed to take on some value among all of its listed possibilities. For continuous random variables, the situation requires a fundamentally different mathematical tool, called the probability density function, or PDF, precisely because a continuous random variable can take on infinitely many possible values within any given interval, no matter how small that interval is drawn. Because of this, it no longer makes any sense to ask for the probability that the random variable equals one single, exact value -- that probability is mathematically always zero, since a single point on a continuous interval has literally zero width and therefore zero area beneath the curve directly above it. Instead, a probability density function is visualized as a smooth curve drawn over the entire range of values the random variable can take, and probability is redefined entirely in terms of area: specifically, the probability that the random variable falls somewhere within a given interval is precisely equal to the area enclosed underneath the density curve and directly above that particular interval on the horizontal axis. Just as with the discrete case, an arbitrary mathematical function cannot automatically be treated as a legitimate probability density function without first satisfying its own pair of validity conditions, which closely mirror the discrete conditions in spirit even though they are phrased somewhat differently. First, the function must never dip below zero at any point across its entire domain, since a curve with any negative height would imply an impossible negative area, and therefore an impossible negative probability, over whatever interval contained that dip. Second, when the total area beneath the entire curve is calculated across the complete range of values the random variable is capable of taking, that grand total area must work out to exactly 1, which is simply the continuous-case restatement of the same underlying certainty that the random variable must take on some value from within its full range of possibilities. This fundamental difference in how probability is defined and measured -- discrete probabilities read directly off a table of individual values, versus continuous probabilities computed as areas beneath a curve over a range -- also explains a subtle but important distinction that trips up many students encountering this material for the first time: for a discrete random variable, the probability $P(a <= Y <= b)$ is genuinely and numerically different from the probability $P(a < Y < b)$, because the individual endpoint probabilities $P(Y=a)$ and $P(Y=b)$ are real, calculable, nonzero quantities that must be deliberately included when the inequality is written with 'or equal to' and just as deliberately excluded when it is written with strict inequality symbols alone. For a continuous random variable, by beautiful contrast, these two probabilities always turn out to be exactly equal to one another, and the reason traces directly back to everything already established above: since the probability of a continuous random variable landing on any single exact point is always precisely zero regardless of which particular point is chosen, including or excluding those two single endpoints from the calculation literally changes nothing about the total area being measured, and therefore changes nothing about the final probability that gets computed.

Explain the concept of expectation (mean) and variance for a discrete random variable, state the properties of expectation, and explain the purpose and construction of the cumulative distribution function.

Once a complete probability distribution has been established for a discrete random variable, one of the most immediately useful things we can do with it is summarize the entire distribution using just two single numbers, in exactly the same spirit as computing a mean and a variance for an ordinary set of raw data back in earlier chapters -- except that here, instead of averaging over data points that were actually observed, we are averaging over every theoretically possible value the random variable could produce, weighted according to how likely each value actually is. This theoretical average is called the mathematical expectation, or expected value, of the random variable, conventionally written $E(Y)$, and it is calculated by taking each possible value y that the random variable Y can assume, multiplying it by its own corresponding probability $P(y)$, and then summing all of these individual products together across every single possible value in the distribution. What makes this expectation genuinely important, rather than merely an abstract computational exercise, is that $E(Y)$ turns out to be nothing more than an alternative, more formal notation for exactly the same population mean, conventionally denoted by the Greek letter mu, that appeared constantly throughout the earlier chapters covering measures of location -- the crucial difference is only that here the mean is being computed directly from a theoretical probability distribution rather than estimated from a finite batch of collected sample data. Building directly on this same underlying logic, the variance of a discrete random variable is formally defined as the expected value of the squared deviation of the random variable away from its own mean, written out fully as sigma-squared equals E of the quantity (Y minus mu) squared, which unpacks computationally into the sum, taken over every possible value of Y, of (Y minus mu) squared multiplied by P(y). While this defining formula is conceptually the clearest way to understand exactly what variance measures -- namely, the average squared spread of the distribution around its own center -- it is rarely the most convenient version of the formula to actually use for computation by hand. Through straightforward algebraic expansion and simplification of the squared term, this defining formula reduces neatly down to the far more practical computing formula Var(Y) equals E(Y-squared) minus the quantity E(Y) squared, which is structurally identical in spirit to the computational shortcut formula for variance that was already introduced for raw, ungrouped data all the way back in the dispersion chapter, just now expressed in terms of expectations rather than raw sums. Working with expectations directly, rather than being forced to recompute an entire distribution completely from scratch every single time some new random variable is introduced, becomes dramatically easier once four essential and very general properties of expectation are firmly understood and readily available for use. The first and most basic property states that the expectation of any constant is simply that same constant itself, written E(c) equals c, which makes intuitive sense since a constant, by its very definition, never actually varies at all and therefore trivially averages out to itself no matter how many times it might hypothetically be observed. The second property, arguably the single most useful and most frequently applied of the four in practice, states that for any two fixed constants a and b, the expectation of the linear transformation bY plus or minus a works out to exactly b times E(Y), plus or minus that same constant a -- in other words, constants can always be pulled directly outside of the expectation operation according to entirely ordinary and predictable algebraic rules, without needing to redo the underlying summation from scratch. The third property extends this same convenient logic to combinations of two entirely separate random variables X and Y, stating that the expectation of their sum, or equally of their difference, is simply equal to the sum, or difference, of each variable's own individual expectation calculated completely separately -- a property that, notably, holds true universally regardless of whether X and Y happen to be independent of one another or not. The fourth and final property, however, does specifically require independence between the two variables involved: when X and Y are genuinely independent random variables, the expectation of their product, E(XY), can be calculated simply as the ordinary product of each variable's own separate individual expectation, E(X) multiplied by E(Y), which importantly does not hold true in general whenever the two variables in question happen to be dependent on one another in any way. Turning now to an entirely different but closely related question that a completed probability distribution is perfectly positioned to answer, we frequently find ourselves needing to know not the probability that a random variable takes on one single specific value, but rather the accumulated probability that it takes on some value at or below a given fixed threshold -- questions phrased exactly like 'what are the chances a particular student scores no more than 80 percent' or 'what is the probability that five tosses of a coin produce no more than three heads total.' This particular type of accumulated, running-total probability is formally captured by what is called the distribution function, more commonly referred to as the cumulative distribution function or simply the CDF for short, written using the notation F(y) and formally defined as F(y) equals P(Y is less than or equal to y). For a discrete random variable specifically, this cumulative distribution function takes on the very distinctive visual shape of a step function when it is graphed out: the function value remains perfectly flat and completely unchanged across any stretch of the number line lying strictly between two consecutive possible values of the random variable, and then it abruptly jumps straight upward by an amount precisely equal to that particular value's own individual probability P(y) at the exact instant the graph passes over each new possible value the random variable is capable of taking. Every legitimate cumulative distribution function, without exception, must satisfy four defining properties that together fully characterize its overall shape and behavior: it must approach exactly 0 as y approaches negative infinity, since there is certainly zero accumulated probability of landing at or below any value lying below the very smallest value the random variable could ever possibly take;

it must likewise approach exactly 1 as y approaches positive infinity, since by that point every single possible value the random variable could ever take has already been fully accounted for and included in the running cumulative total; the function must be non-decreasing throughout its entire domain, meaning F(y1) is always less than or equal to F(y2) whenever y1 is less than or equal to y2, since accumulated probability can only ever increase or stay flat as more possible values get progressively included, and can certainly never decrease as we move further along the number line; and finally, the function must remain continuous at the very least from the right-hand side at every single point along its domain, which is simply the more technical and mathematically rigorous way of describing that characteristic staircase-like step shape already described above.

Multiple Choice Questions (MCQs)

The probability mass function (PMF) applies to: (A) Continuous random variables only (B) Discrete random variables (C) Constants only (D) Sample spaces only

Correct answer: (B) Discrete random variables. The PMF describes the probability of each specific value of a discrete random variable.

For a valid PMF, the sum of all probabilities P(y) must equal: (A) 0 (B) 0.5 (C) 1 (D) Any positive number

Correct answer: (C) 1. One of the two defining conditions of a valid PMF is that all probabilities sum to exactly 1.

For a continuous random variable, P(Y = a) for any specific value a is: (A) 1 (B) 0.5 (C) Undefined (D) Always 0

Correct answer: (D) Always 0. Probability at a single point has zero width under the density curve, so it is always exactly 0 for continuous variables.

The area under a valid probability density function (PDF) must equal: (A) 0 (B) 1 (C) Infinity (D) The mean

Correct answer: (B) 1. A defining condition of a valid PDF is that the total area under the curve equals 1.

The expected value E(Y) of a discrete random variable is calculated as: (A) Sum of Y only (B) Sum of y.P(y) over all values (C) Sum of P(y) only (D) Maximum value of Y

Correct answer: (B) Sum of y.P(y) over all values. E(Y) = sum of y.P(y), the probability-weighted average of all possible values.

The computing (shortcut) formula for variance of a discrete random variable is: (A) Var(Y) = E(Y) - E(Y^2) (B) Var(Y) = E(Y^2) - [E(Y)]^2 (C) Var(Y) = [E(Y)]^2 (D) Var(Y) = E(Y)^2 + E(Y^2)

Correct answer: (B) Var(Y) = E(Y^2) - [E(Y)]^2. Var(Y) = E(Y^2) - [E(Y)]^2 is the standard computing formula, derived by expanding E[(Y-mu)^2].

If E(X) = 4, then E(2X + 3) equals: (A) 7 (B) 8 (C) 11 (D) 14

Correct answer: (C) 11. Using E(bY+a) = b.E(Y)+a: E(2X+3) = 2(4)+3 = 11.

If X and Y are independent random variables, E(XY) equals: (A) E(X) + E(Y) (B) E(X) - E(Y) (C) E(X).E(Y) (D) E(X)/E(Y)

Correct answer: (C) E(X).E(Y). For independent random variables, the expectation of the product equals the product of the expectations.

The cumulative distribution function F(y) is defined as: (A) P(Y = y) (B) P(Y >= y) (C) P(Y <= y) (D) P(Y < 0)

Correct answer: (C) P(Y <= y). F(y) = P(Y <= y), the probability of the random variable being at or below y.

The graph of the CDF of a discrete random variable looks like a: (A) Smooth curve (B) Straight line (C) Step function (D) Circle

Correct answer: (C) Step function. Since probability accumulates only at isolated points for a discrete random variable, its CDF forms a step function.

Quick Revision Summary

- PMF (discrete): P(y), with 0<=P(y)<=1 for each y and sum of P(y)=1
- PDF (continuous): f(y)>=0 everywhere, total area under curve = 1; probability = area over an interval
- Discrete: P(a<=Y<=b) != P(a<Y<b). Continuous: these two ARE equal (P(Y=a)=0 always)
- Probability histogram: rectangles of width 1, height = probability, area = probability

- E(Y) = sum y.P(y) (the mean); Var(Y) = E(Y^2) - [E(Y)]^2
- Properties: E(c)=c; E(bY+a)=bE(Y)+a; E(X+/-Y)=E(X)+/-E(Y); E(XY)=E(X)E(Y) if independent
- CDF: F(y) = P(Y<=y); F(-inf)=0, F(+inf)=1, non-decreasing, step function for discrete variables

Exam Tips

- Before treating any function as a probability distribution, always check BOTH validity conditions: non-negativity and total probability/area = 1
- For continuous PDF area problems with a linear f(x), use the trapezoid formula: [f(a)+f(b)]/2 x base, rather than full integration when possible
- Remember discrete P(a<=Y<=b) includes both endpoints separately — don't confuse it with the strict-inequality version
- When computing variance, build a table with columns y, P(y), yP(y), $y^2P(y)$ to avoid arithmetic slips
- To sketch a discrete CDF, list cumulative sums of P(y) at each value, then draw a step that jumps at each value and stays flat in between