6.3: Probability Distribution Function (PDF) for Discrete Random Variables

All random variables have the value assigned in accordance with a probability model. For discrete variables, this assigning of probabilities to each possible value of the random variable is called a **probability distribution function**, or PDF for short.

This probability distribution function is written as $P(X = x)$ or $P(x)$ for short. This PDF can be read as "The probability the random variable X equals the value x."

Additionally, probability statements can be written as inequalities.

$P(X < x)$ means the probability the value of the random variable is less than x.

$P(X \leq x)$ means the probability the value of the random variable is at most x.

$P(X > x)$ means the probability the value of the random variable is more than x.

$P(X \geq x)$ means the probability the value of the random variable is at least x.

Like any function in Mathematics, a probability distribution function can be defined by a description, a table, a graph or a formula. The general method of assigning probabilities to values follows this procedure.

Procedure for creating a discrete probability distribution function

1. Define the random Variable X
2. List out all possible values
3. Assign probabilities to each value. You can use counting methods or relative frequencies.
4. This assignment must follow these two rules: $P(x) \geq 0$ and $\sum P(x) = 1$

Example: Flip two coins

Two coins are flipped and the number of heads are counted.

X = the number of heads when two coins are flipped

Possible Values = {0, 1, 2}

Here are 5 possible probability distribution functions:

A		B		C		D		E	
x	$P(x)$	x	$P(x)$	x	$P(x)$	$1x$	$P(x)$	x	$P(x)$
0	1/3	0	0.25	0	0	0	0.3	0	0.6
1	1/3	1	0.50	1	0	1	0.3	1	-0.1
2	1/3	2	0.25	2	1	2	0.3	2	0.5

Models A, B and C are valid because each probability assignment is non-negative and all probabilities total to 1.

Model B is the correct model for flipping fair coins as there are two ways to get one head.

Model C (a coin that only comes up head) is valid since zero probability is allowed.

Model D is invalid since the probabilities do not total to 1.

Model E is invalid because negative probabilities are not allowed.

Example: Multiple choice test

Students are given a multiple choice exam with 4 questions.

The random variable X = the number answers correct. Possible values = {0, 1, 2, 3, 4}

From past data, 10% of students get zero correct answers, 10% get exactly one correct answer, 20% get two correct, and 40% get three correct. Since the probabilities must add to 1, it can be determined that 20% of students got all correct, and the PDF can be finished.

x	$P(x)$
0	0.1
1	0.1

x	$P(x)$
2	0.2
3	0..4
4	0.2

Solution

We can use the table to answer any type of probability question:

The probability of exactly 2 questions correct: $P(X = 2) = P(2) = 0.2$

The probability of fewer than 2 questions correct: $P(X < 2) = P(0) + P(1) = 0.1 + 0.1 = 0.2$

The probability of more than 2 questions correct: $P(X > 2) = P(3) + P(4) = 0.4 + 0.2 = 0.6$

The probability of at least 2 questions correct: $P(X \geq 2) = P(2) + P(3) + P(4) = 0.2 + 0.4 + 0.2 = 0.8$

The probability of at most 2 questions correct: $P(X \leq 2) = P(0) + P(1) + P(2) = 0.1 + 0.1 + 0.2 = 0.4$

The probability at least 1 question correct: $P(X > 1) = 1 - P(0) = 1 - 0.1 = 0.9$

The last example was done using the Rule of Complement. The complement of “at least one correct answer” is “zero correct answers”.
