# Random Variables

A Random Variable is a set of **possible values** from a random experiment.

A ***random variable***, usually written *X*, is a variable whose possible values are numerical outcomes of a random phenomenon. There are two types of random variables, ***discrete*** and ***continuous***.

# Discrete Random Variables

- A ***discrete random variable*** is one which may take on only a countable number of distinct values such as 0,1,2,3,4,........ Discrete random variables are usually (but not necessarily) counts. If a random variable can take only a finite number of distinct values, then it must be discrete. Examples of discrete random variables include the number of children in a family, the Friday night attendance at a cinema, the number of patients in a doctor's surgery, the number of defective light bulbs in a b
- The ***probability distribution*** of a discrete random variable is a list of probabilities associated with each of its possible values. It is also sometimes called the probability function or the probability mass function.
-

# Discrete random variable

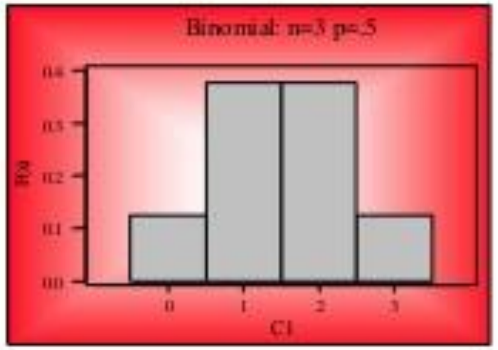
## Discrete and Continuous Random Variables - Revisited



- A **discrete random variable:**
  - counts occurrences
  - has a countable number of possible values
  - has discrete jumps between successive values
  - has measurable probability associated with individual values
  - probability is **height**

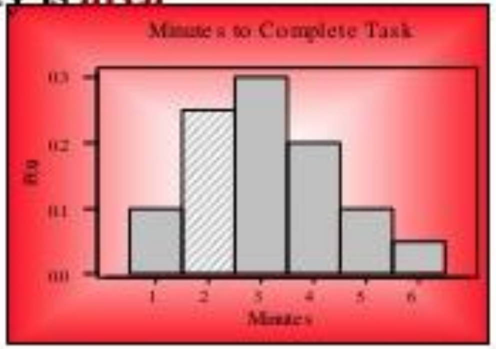
For example:
Binomial
n=3 p=.5

| x | P(x) |
|---|---|
| 0 | 0.125 |
| 1 | 0.375 |
| 2 | 0.375 |
| 3 | 0.125 |
| | 1.000 |




- A **continuous random variable:**
  - measures (e.g.: height, weight, speed, value, duration, length)
  - has an uncountably infinite number of possible values
  - moves continuously from value to value
  - has **no** measurable probability associated with individual values
  - probability is **area**

For example:
In this case, the shaded area epresents the probability that the task takes between 2 and 3 minutes.



# Continuous Random Variables

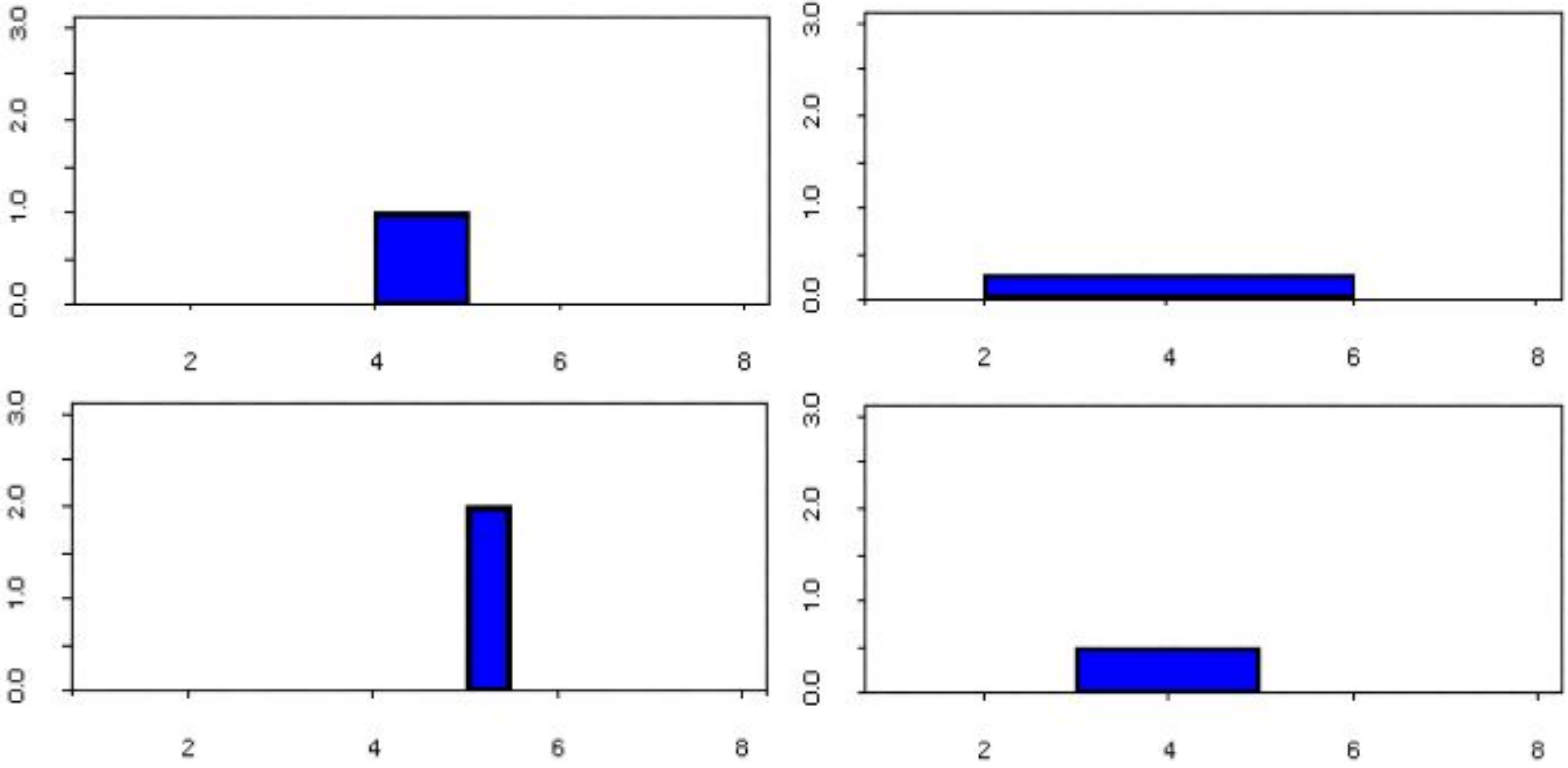
- A ***continuous random variable*** is one which takes an infinite number of possible values. Continuous random variables are usually measurements. Examples include height, weight, the amount of sugar in an orange, the time required to run a mile.
- 
- A continuous random variable is not defined at specific values. Instead, it is defined over an *interval* of values, and is represented by the ***area under a curve*** (in advanced mathematics, this is known as an *integral*). The probability of observing any single value is equal to 0, since the number of values which may be assumed by the random variable is infinite.
- 

- Suppose a random variable *X* may take all values over an interval of real numbers. Then the probability that *X* is in the set of outcomes *A, P(A)*, is defined to be the area above *A* and under a curve. The curve, which represents a function *p(x)*, must satisfy the following:
- ***1:*** *The curve has no negative values ($p(x) \geq 0$ for all x)*
- ***2:*** *The total area under the curve is equal to 1.*
- A curve meeting these requirements is known as a ***density curve***.

# The Uniform Distribution

- A random number generator acting over an interval of numbers *(a,b)* has a continuous distribution. Since any interval of numbers of equal width has an equal probability of being observed, the curve describing the distribution is a rectangle, with constant height across the interval and 0 height elsewhere. Since the area under the curve must be equal to 1, the length of the interval determines the height of the curve.
- The following graphs plot the density curves for random number generators over the intervals (4,5) (top left), (2,6) (top right), (5,5.5) (lower left), and (3,5) (lower right). The distributions corresponding to these curves are known as ***uniform distributions***.
-

# The Uniform Distribution



- Consider the uniform random variable *X* defined on the interval (2,6). Since the interval has width = 4, the curve has height = 0.25 over the interval and 0 elsewhere. The probability that *X* is less than or equal to 5 is the area between 2 and 5, or (5-2)*0.25 = 0.75. The probability that *X* is greater than 3 but less than 4 is the area between 3 and 4, (4-3)*0.25 = 0.25. To find that probability that *X* is less than 3 *or*greater than 5, add the two probabilities: $P(X \leq 3 \text{ and } X \geq 5) = P(X \leq 3) + P(X \geq 5) = (3-2)*0.25 + (6-5)*0.25 = 0.25 + 0.25 = 0.5$.
- The uniform distribution is often used to simulate data. Suppose you would like to simulate data for 10 rolls of a regular 6-sided die. Using the MINITAB "RAND" command with the "UNIF" subcommand generates 10 numbers in the interval (0,6):

- ```
  MTB > RAND 10 c2;
  SUBC> unif 0 6.
  ```
  Assign the discrete random variable X to the values 1, 2, 3, 4, 5, or 6 as follows:
  if $0<X<1$, $X=1$
  if $1<X<2$, $X=2$
  if $2<X<3$, $X=3$
  if $3<X<4$, $X=4$
  if $4<X<5$, $X=5$
  if $X>5$, $X=6$.
  ```

- Use the generated MINITAB data to assign X to a value for each roll of the die:

| Uniform Data | X Value |
|---|---|
| 4.53786 | 5 |
| 5.77474 | 6 |
| 3.69518 | 4 |
| 1.03929 | 2 |
| 4.23835 | 5 |
| 0.37096 | 1 |
| 0.75272 | 1 |
| 5.56563 | 6 |
| 0.89045 | 1 |
| 3.18086 | 4 |