Random Variable

- A Random Variable is also called a *chance variable or* a *probability variable.* These names suggest that the variable has something to do with probabilities.
- The variable which is random is called stochastic variables.

- A random variable *x* takes on a defined set of values with different probabilities.
    - For example, if you roll a die, the outcome is random (not fixed) and there are 6 possible outcomes, each of which occur with probability one-sixth.
    - For example, if you poll people about their voting preferences, the percentage of the sample that responds "Yes on Proposition 100" is a also a random variable (the percentage will be slightly differently every time you poll).

Random variables can be either discrete or continuous

- **Discrete** random variables have a countable number of outcomes
  - Examples: Dead/alive, treatment/placebo, dice, counts, etc.
- **Continuous** random variables have an infinite continuum of possible values.
  - Examples: blood pressure, weight, the speed of a car, the real numbers from 1 to 6.

Probability Functions

- A probability function maps the possible values of x against their respective probabilities of occurrence, $p(x)$
- $p(x)$ is a number from 0 to 1.0.
- The area under a probability function is always 1.

- Every Random Variable has a Probability Distribution.

Discrete Probability Distributions

A **discrete probability distribution** lists each possible value the random variable can assume, together with its probability of occurrence. A probability distribution must satisfy the following conditions.

In Words	*In Symbols*
1. The probability of each value of the discrete random variable is between 0 and 1, inclusive.	$0 <= P(x) <=1$
2. The sum of all the probabilities is 1.	$\Sigma P(x) = 1$

Discrete Probability Distributions

- Discrete Probability Distribution - Listing of outcomes and their corresponding probabilities (y , $P(y)$)

$$0 \leq P(y) \leq 1 \qquad \sum_{all\ y} P(y) = 1$$

DISCRETE PROBABILITY FUNCTIONS

It is represented by a **probability mass function (PMF)**. If a stochastic variable can take **I** different values x_i for (**i=1,2,3,4,5……. I**)and the probability of the value x_i being taken as **P(x_i)** i.e., $x_i \rightarrow$ **P(x_i),** the set of numbers **P(x_i)** is said to be **PMF** and $\sum_{i=1}^{I} P(xi) = 1$.

Example:

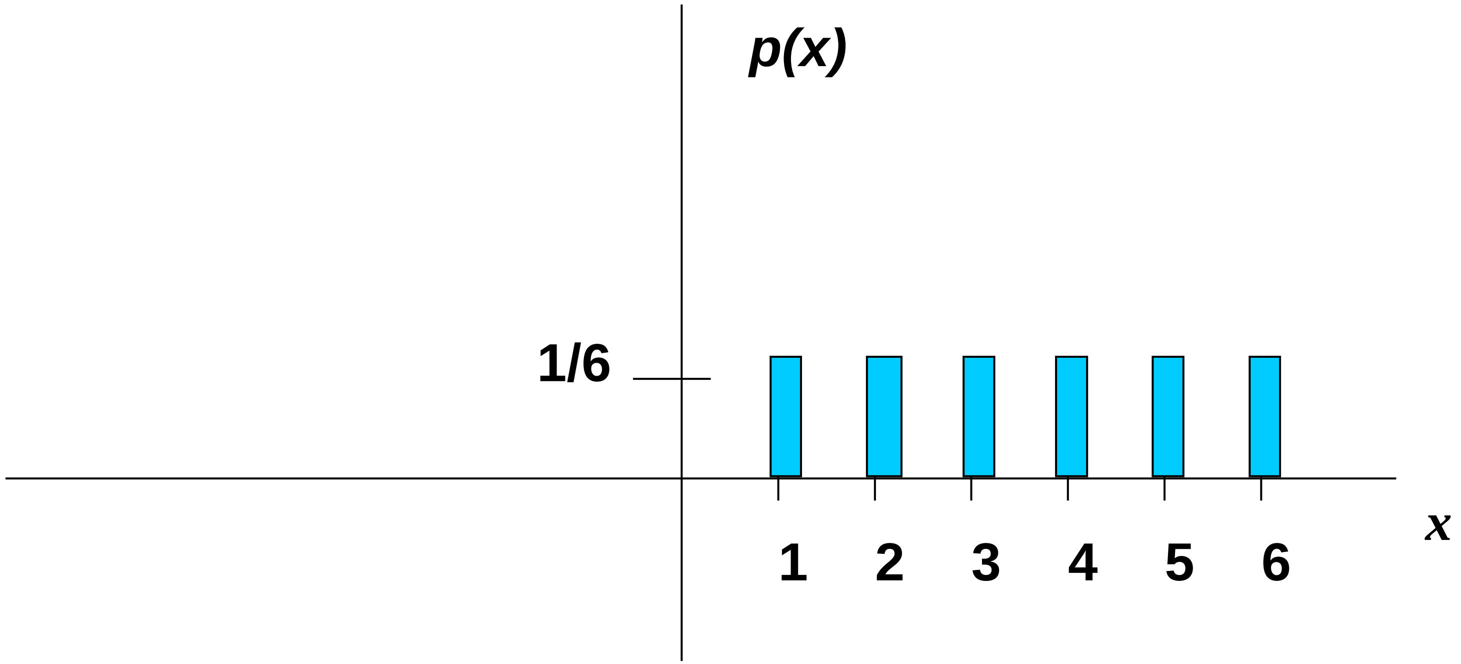
NO OF ITEMS BOUGHT BY CUSTOMERS

No. of items (x_i)	No. of customers (n_i)	Probability distribution	Cumulative Distribution
1	25	0.10	0.10
2	128	0.51	0.61
3	47	0.19	0.80
4	38	0.15	0.95
5	12	0.05	1.00

N=250

A **cumulative distribution function** is defined as a function that gives the probability of a random variable being less than or, equal to a given value. It can be increased to a maximum of 1.

Another example: roll of a die



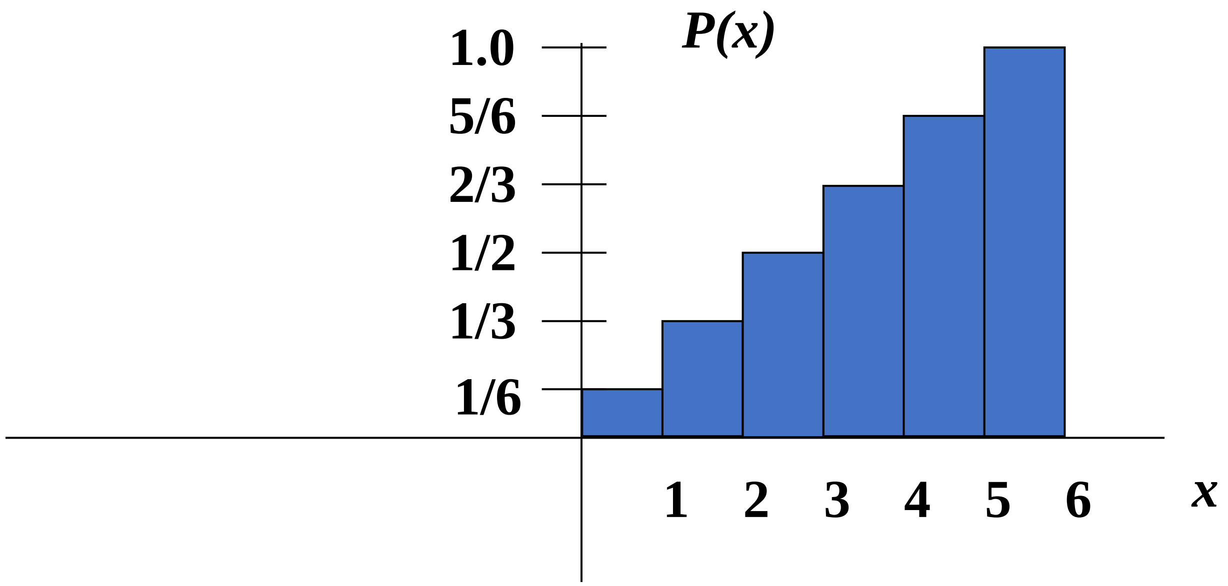
$$\sum_{\text{all } x} P(x) = 1$$

Probability mass function (pmf)

x	$p(x)$
1	$p(x=1)=1/6$
2	$p(x=2)=1/6$
3	$p(x=3)=1/6$
4	$p(x=4)=1/6$
5	$p(x=5)=1/6$
6	$p(x=6)=1/6$

1.0

Cumulative distribution function (CDF)



Cumulative distribution function

x	$P(x \leq A)$
1	$P(x \leq 1)$=1/6
2	$P(x \leq 2)$=2/6
3	$P(x \leq 3)$=3/6
4	$P(x \leq 4)$=4/6
5	$P(x \leq 5)$=5/6
6	$P(x \leq 6)$=6/6

Examples

1. What's the probability that you roll a 3 or less?
$P(x \leq 3) = 1/2$

2. What's the probability that you roll a 5 or higher?
$P(x \geq 5) = 1 - P(x \leq 4) = 1-2/3 = 1/3$

Practice Problem

Which of the following are probability functions?

a. $f(x)=.25$ for x=9,10,11,12

b. $f(x)= (3-x)/2$ for x=1,2,3,4

c. $f(x)= (x^2+x+1)/25$ for x=0,1,2,3

Answer (a)

a. $f(x)=.25$ for x=9,10,11,12

x	***f(x)***
9	**.25**
10	**.25**
11	**.25**
12	**.25**

1.0

Yes, probability function!

Answer (b)

b. $f(x) = (3-x)/2$ for x=1,2,3,4

x	***f(x)***
1	**(3-1)/2=1.0**
2	**(3-2)/2=.5**
3	**(3-3)/2=0**
4	**(3-4)/2=-.5**

Though this sums to 1, you can't have a negative probability; therefore, it's not a probability function.

Answer (c)

c. $f(x) = (x^2+x+1)/25$ for x=0,1,2,3

x	**f(x)**
0	**1/25**
1	**3/25**
2	**7/25**
3	**<u>13/25</u>**

24/25

Doesn't sum to 1. Thus, it's not a probability function.

Practice Problem:

- The number of ships to arrive at a harbor on any given day is a random variable represented by *x*. The probability distribution for *x* is:

x	**10**	**11**	**12**	**13**	**14**
P(x)	**.4**	**.2**	**.2**	**.1**	**.1**

Find the probability that on a given day:

a. exactly 14 ships arrive $\quad p(x=14) = .1$

b. At least 12 ships arrive $\quad p(x \geq 12) = (.2 + .1 + .1) = .4$

c. At most 11 ships arrive $\quad p(x \leq 11) = (.4 + .2) = .6$

Practice Problem:

You are lecturing to a group of 1000 students. You ask them to each randomly pick an integer between 1 and 10. Assuming, their picks are truly random:

- What's your best guess for how many students picked the number 9?

  Since $p(x=9)$ = 1/10, we'd expect about 1/10th of the 1000 students to pick 9. 100 students.

- What percentage of the students would you expect picked a number less than or equal to 6?

  Since $p(x \leq 6)$ = 1/10 + 1/10 + 1/10 + 1/10 + 1/10 + 1/10 =.6 60%

Mean

The **mean** of a discrete random variable is given by

$$\mu = \Sigma x P(x).$$

Each value of x is multiplied by its corresponding probability and the products are added.

Example:

Find the mean of the probability distribution for the sum of the two spins.

x	$P(x)$	$xP(x)$
2	0.0625	2(0.0625) = 0.125
3	0.375	3(0.375) = 1.125
4	0.5625	4(0.5625) = 2.25

$\Sigma x P(x) = 3.5$

The mean for the two spins is 3.5.

Variance

The **variance** of a discrete random variable is given by

$$\sigma^2 = \Sigma(x - \mu)^2 P(x).$$

Example:

Find the variance of the probability distribution for the sum of the two spins. The mean is 3.5.

x	$P(x)$	$x - \mu$	$(x - \mu)^2$	$P(x)(x - \mu)^2$
2	0.0625	–1.5	2.25	≈ 0.141
3	0.375	–0.5	0.25	≈ 0.094
4	0.5625	0.5	0.25	≈ 0.141

$\Sigma P(x)(x - 2)^2$

≈ 0.376

The variance for the two spins is approximately 0.376

Standard Deviation

The **standard deviation** of a discrete random variable is given by

$$\sigma = \sqrt{\sigma^2}.$$

Example:

Find the standard deviation of the probability distribution for the sum of the two spins. The variance is 0.376.

x	$P(x)$	$x - \mu$	$(x-\mu)^2$	$P(x)(x-\mu)^2$
2	0.0625	–1.5	2.25	0.141
3	0.375	–0.5	0.25	0.094
4	0.5625	0.5	0.25	0.141

$$\sigma = \sqrt{\sigma^2}$$
$$= \sqrt{0.376} \approx 0.613$$

Most of the sums differ from the mean by no more than 0.6 points.

Expected Value

The **expected value** of a discrete random variable is equal to the mean of the random variable.

$$\text{Expected Value} = E(x) = \mu = \Sigma x P(x).$$

Example:

At a raffle, 500 tickets are sold for $1 each for two prizes of $100 and $50. What is the expected value of your gain?

Your gain for the $100 prize is $100 – $1 = $99.

Your gain for the $50 prize is $50 – $1 = $49.

Write a probability distribution for the possible gains (or outcomes).

Continued.

Expected Value

Example continued:

At a raffle, 500 tickets are sold for $1 each for two prizes of $100 and $50. What is the expected value of your gain?

Gain, x	$P(x)$
$\$99$	$\frac{1}{500}$
$\$49$	$\frac{1}{500}$
$-\$1$	$\frac{498}{500}$

Winning no prize

$E(x) = \Sigma x P(x).$

$$= \$99 \cdot \frac{1}{500} + \$49 \cdot \frac{1}{500} + (-\$1) \cdot \frac{498}{500}$$

$$= -\$0.70$$

Because the expected value is negative, you can expect to lose $0.70 for each ticket you buy.

Important Discrete Distributions

- Binomial
  - Yes/no outcomes (dead/alive, treated/untreated, smoker/non-smoker, sick/well, etc.)
- Poisson
  - Counts (e.g., how many cases of disease in a given area)

Binomial Distributions

Binomial Experiments

A **binomial experiment** is a probability experiment that satisfies the following conditions.

1. The experiment is repeated for a fixed number of trials, where each trial is independent of other trials.
2. There are only two possible outcomes of interest for each trial. The outcomes can be classified as a success (S) or as a failure (F).
3. The probability of a success $P(S)$ is the same for each trial.
4. The random variable x counts the number of successful trials.

Notation for Binomial Experiments

Symbol	*Description*
n	The number of times a trial is repeated.
$p = P(S)$	The probability of success in a single trial.
$q = P(F)$	The probability of failure in a single trial. $(q = 1 - p)$
x	The random variable represents a count of the number of successes in n trials: $x = 0, 1, 2, 3, \ldots, n$.

Binomial Experiments

Example:

Decide whether the experiment is a binomial experiment. If it is, specify the values of n, p, and q, and list the possible values of the random variable x. If it is not a binomial experiment, explain why.

- You randomly select a card from a deck of cards, and note if the card is an Ace. You then put the card back and repeat this process 8 times.

This is a binomial experiment. Each of the 8 selections represent an independent trial because the card is replaced before the next one is drawn. There are only two possible outcomes: either the card is an Ace or not.

$n = 8 \qquad p = \frac{4}{52} = \frac{1}{13} \qquad q = 1 - \frac{1}{13} = \frac{12}{13} \qquad x = 0,1,2,3,4,5,6,7,8$

Binomial Experiments

Example:
Decide whether the experiment is a binomial experiment. If it is, specify the values of n, p, and q, and list the possible values of the random variable x. If it is not a binomial experiment, explain why.

- You roll a die 10 times and note the number the die lands on.

This is not a binomial experiment. While each trial (roll) is independent, there are more than two possible outcomes: 1, 2, 3, 4, 5, and 6.

Binomial Probability Formula

In a binomial experiment, the probability of exactly x successes in n trials is

$$P(x) = {}_nC_x p^x q^{n-x} = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}.$$

Example:

A bag contains 10 chips. 3 of the chips are red, 5 of the chips are white, and 2 of the chips are blue. Three chips are selected, with replacement. Find the probability that you select exactly one red chip.

$p =$ the probability of selecting a red chip $= \frac{3}{10} = 0.3$

$q = 1 - p = 0.7$

$n = 3$

$x = 1$

$$P(1) = {}_3C_1(0.3)^1(0.7)^2$$

$$= 3(0.3)(0.49)$$

$$= 0.441$$

Binomial Probability Distribution

Example:

A bag contains 10 chips. 3 of the chips are red, 5 of the chips are white, and 2 of the chips are blue. Four chips are selected, with replacement. Create a probability distribution for the number of red chips selected.

$p =$ the probability of selecting a red chip $= \frac{3}{10} = 0.3$

$q = 1 - p = 0.7$

$n = 4$

$x = 0, 1, 2, 3, 4$

x	$P(x)$
0	0.240
1	0.412
2	0.265
3	0.076
4	0.008

The binomial probability formula is used to find each probability.

Finding Probabilities

Example:

The following probability distribution represents the probability of selecting 0, 1, 2, 3, or 4 red chips when 4 chips are selected.

x	$P(x)$
0	0.24
1	0.412
2	0.265
3	0.076
4	0.008

a.) Find the probability of selecting no more than 3 red chips.

b.) Find the probability of selecting at least 1 red chip.

a.) $P(\text{no more than } 3) = P(x \leq 3) = P(0) + P(1) + P(2) + P(3)$

$= 0.24 + 0.412 + 0.265 + 0.076 = 0.993$

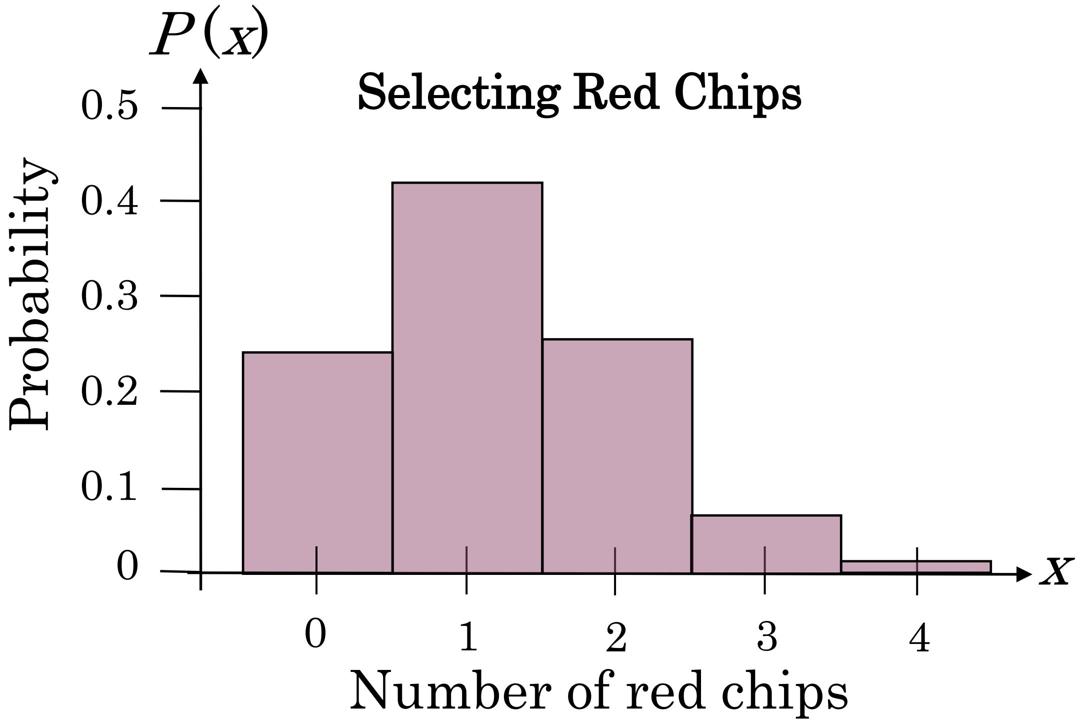
b.) $P(\text{at least } 1) = P(x \geq 1) = 1 - \underbrace{P(0)}_{\text{Complement}} = 1 - 0.24 = 0.76$

Graphing Binomial Probabilities

Example:

The following probability distribution represents the probability of selecting 0, 1, 2, 3, or 4 red chips when 4 chips are selected. Graph the distribution using a histogram.

x	$P(x)$
0	0.24
1	0.412
2	0.265
3	0.076
4	0.008



Mean, Variance and Standard Deviation

Population Parameters of a Binomial Distribution

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

Standard deviation: $\sigma = \sqrt{npq}$

Example:

One out of 5 students at a local college say that they skip breakfast in the morning. Find the mean, variance and standard deviation if 10 students are randomly selected.

$$n = 10 \qquad p = \frac{1}{5} = 0.2 \qquad q = 0.8$$

$$\mu = np = 10(0.2) = 2$$

$$\sigma^2 = npq = (10)(0.2)(0.8) = 1.6$$

$$\sigma = \sqrt{npq} = \sqrt{1.6} \approx 1.3$$

Geometric Distribution

A **geometric distribution** is a discrete probability distribution of a random variable x that satisfies the following conditions.

1. A trial is repeated until a success occurs.
2. The repeated trials are independent of each other.
3. The probability of a success p is constant for each trial.

The **probability that the first success will occur on trial $\boldsymbol{x}$** is

$$P(x) = p(q)^{x-1}, \text{ where } q = 1 - p.$$

Geometric Distribution

Example:

A fast food chain puts a winning game piece on every fifth package of French fries. Find the probability that you will win a prize,

a.) with your third purchase of French fries,

b.) with your third or fourth purchase of French fries.

$p = 0.20$ $\quad$ $q = 0.80$

a.) $x = 3$

$$P(3) = (0.2)(0.8)^{3-1}$$
$$= (0.2)(0.8)^2$$
$$= (0.2)(0.64)$$
$$= 0.128$$

b.) $x = 3, 4$

$$P(3 \text{ or } 4) = P(3) + P(4)$$
$$\approx 0.128 + 0.102$$
$$\approx 0.230$$

Poisson Distribution

The **Poisson distribution** is a discrete probability distribution of a random variable x that satisfies the following conditions.

1. The experiment consists of counting the number of times an event, x, occurs in a given interval. The interval can be an interval of time, area, or volume.
2. The probability of the event occurring is the same for each interval.
3. The number of occurrences in one interval is independent of the number of occurrences in other intervals.

The probability of exactly x occurrences in an interval is

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

where $e \approx 2.71818$ and μ is the mean number of occurrences.

Poisson Distribution

Example:

The mean number of power outages in the city of Brunswick is 4 per year. Find the probability that in a given year,

a.) there are exactly 3 outages,

b.) there are more than 3 outages.

a.) $\mu = 4, \quad x = 3$

$$P(3) = \frac{4^3(2.71828)^{-4}}{3!}$$

$$\approx 0.195$$

b.) $P(\text{more than } 3)$

$$= 1 - P(x \le 3)$$

$$= 1 - [P(3) + P(2) + P(1) + P(0)]$$

$$= 1 - (0.195 + 0.147 + 0.073 + 0.018)$$

$$\approx 0.567$$

At a theme park, there is a roller coaster that sends an average of three cars through its circuit every minute between 6pm and 7pm. A random variable, X, represents the number of roller coaster cars to pass through the circuit between 6pm and 6:10pm.

What is the probability that 35 cars will pass through the circuit between 6pm and 6:10pm?

Mean:

$$\mu_x = \lambda t$$

Standard Deviation:

$$\sigma_x = \sqrt{\mu_x}$$

x = number of successes = 35

t = a length of time = 10 minutes

λ = average number of successes in an interval of (t = 1) = 3

e = constant = 2.718

$$\mu_x = (3)(10) = 30$$

$$\sigma_x = \sqrt{30} = 5.477$$

Continuous case

- The probability function that accompanies a continuous random variable is a continuous mathematical function that integrates to 1.
- The probabilities associated with continuous functions are just areas under the curve (integrals!).
- Probabilities are given for a range of values, rather than a particular value (e.g., the probability of getting a math SAT score between 700 and 800 is 2%).

Continuous case

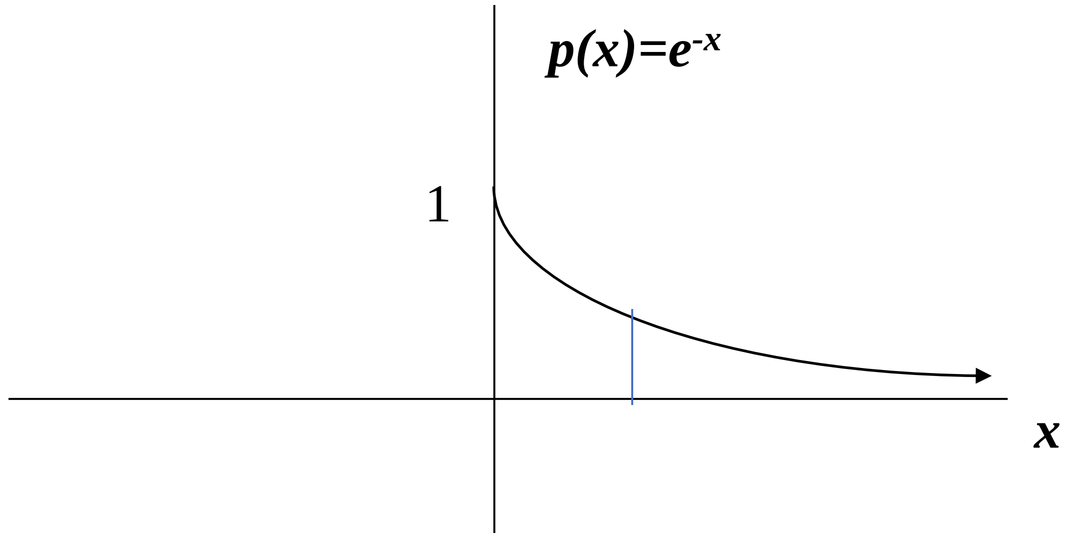
- For example, recall the negative exponential function (in probability, this is called an “exponential distribution”):

$$f(x) = e^{-x}$$

- This function integrates to 1:

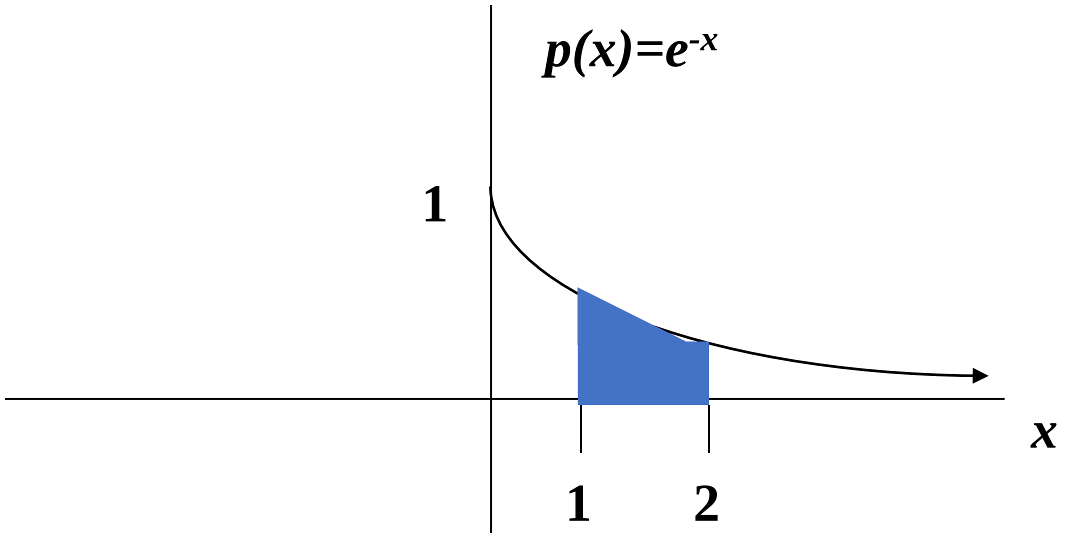
$$\int_0^{+\infty} e^{-x} = -e^{-x} \Big|_0^{+\infty} = 0 + 1 = 1$$

Continuous case: “probability density function” (pdf)



The probability that *x* is any exact particular value (such as 1.9976) is 0; we can only assign probabilities to possible ranges of x.

For example, the probability of *x* falling within 1 to 2:



$$P(1 \le \mathrm{x} \le 2) = \int_1^2 e^{-x} = -e^{-x} \Big|_1^2 = -e^{-2} - -e^{-1} = -.135 + .368 = .23$$

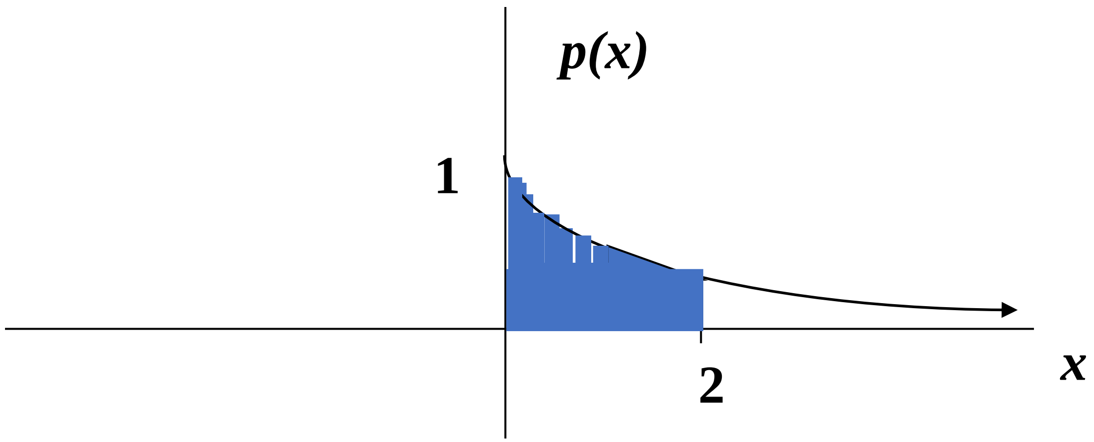
Cumulative Distribution Function

As in the discrete case, we can specify the "cumulative distribution function" (CDF):

The CDF here = P(x≤A)=

$$\int_0^A e^{-x} = -e^{-x} \Big|_0^A = -e^{-A} - -e^{0} = -e^{-A} + 1 = 1 - e^{-A}$$

Example

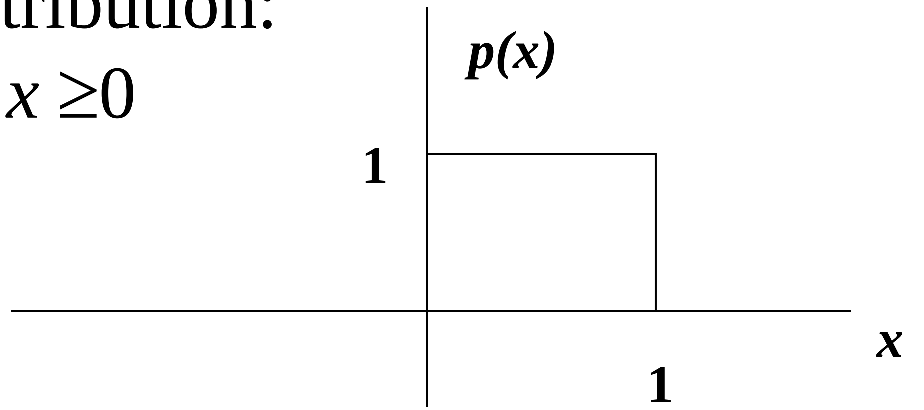


$$P(x \leq 2) = 1 - e^{-2} = 1 - .135 = .865$$

Example 2: Uniform distribution

The uniform distribution: all values are equally likely

The uniform distribution:
$f(x) = 1$, for $1 \geq x \geq 0$

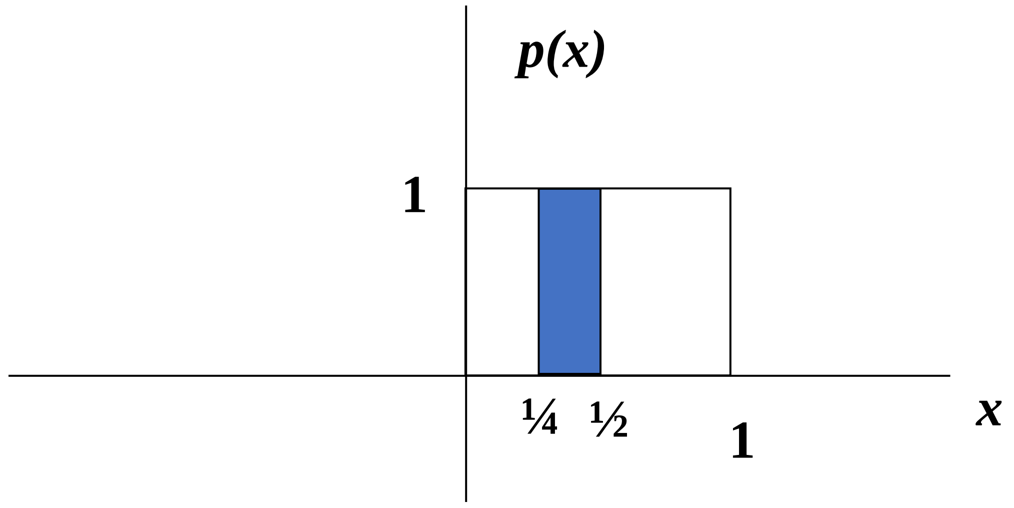


We can see it's a probability distribution because it integrates to 1 (the area under the curve is 1):

$$\int_0^1 1 = x \Big|_0^1 = 1 - 0 = 1$$

Example: Uniform distribution

What's the probability that x is between ¼ and ½?



P(½ ≥x≥ ¼)= ¼

Practice Problem

Suppose that survival drops off rapidly in the year following diagnosis of a certain type of advanced cancer. Suppose that the length of survival (or time-to-death) is a random variable that approximately follows an exponential distribution with parameter 2 (makes it a steeper drop off):

$$\text{probability function}: p(x = T) = 2e^{-2T}$$

$$[note: \int_0^{+\infty} 2e^{-2x} = -e^{-2x} \Big|_0^{+\infty} = 0 + 1 = 1]$$

What's the probability that a person who is diagnosed with this illness survives a year?

Answer

The probability of dying within 1 year can be calculated using the cumulative distribution function:

Cumulative distribution function is:

$$P(x \le T) = -e^{-2x} \Big|_0^T = 1 - e^{-2(T)}$$

The chance of surviving past 1 year is: *P(x≥1) = 1 – P(x≤1)*

$$1-(1-e^{-2(1)}) = .135$$