Lesson 20

Probability and Cumulative Distribution Functions

Recall

If $p(x)$ is a density function for some characteristic of a population, then

$$\int_a^b p(x)\,dx = \begin{pmatrix} \text{fraction of the} \\ \text{population for} \\ \text{which } a \le x \le b \end{pmatrix}$$

Recall

If $p(x)$ is a density function for some characteristic of a population, then

$$\int_a^b p(x)\,dx = \begin{pmatrix} \text{fraction of the} \\ \text{population for} \\ \text{which } a \le x \le b \end{pmatrix}$$

We also know that for any density function,

$$\int_{-\infty}^{\infty} p(x)\,dx = 1$$

Recall

We also interpret density functions as probabilities:

If $p(x)$ is a probability density function (pdf), then

$$\int_a^b p(x)\,dx = \begin{pmatrix} \text{probability} \\ \text{that} \\ a \le x \le b \end{pmatrix}$$

Cumulative Distribution Function

Suppose $p(x)$ is a density function for a quantity.

The *cumulative distribution function* (cdf) for the quantity is defined as

$$P(x) = \int_{-\infty}^{x} p(t)\,dt$$

Gives:

- The proportion of population with value less than x
- The probability of having a value less than x.

Example: A Spinner

Last class: A spinner that could take on any value $0^{\circ} \leq x \leq 360^{\circ}$.

Density function: $p(x) = 1/360$ if $0 \leq x \leq 360$, and 0 everywhere else.

Example: A Spinner

Last class: A spinner that could take on any value $0^{o} \leq x \leq 360^{o}$.

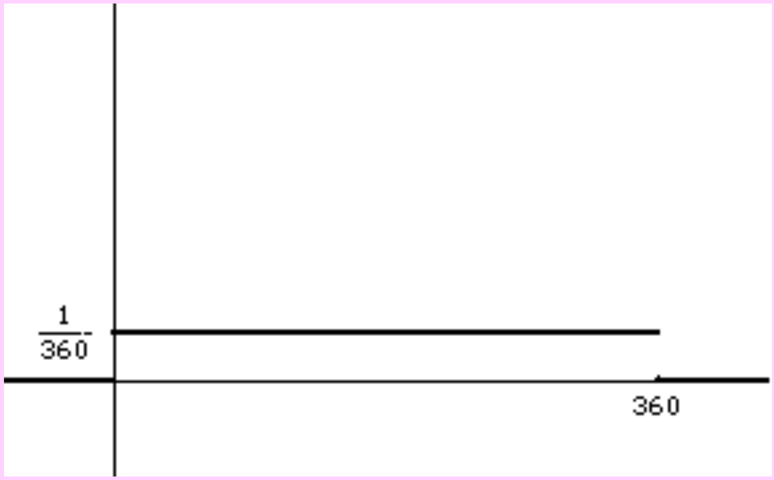
Density function: $p(x) = 1/360$ if $0 \leq x \leq 360$, and 0 everywhere else.

CDF:

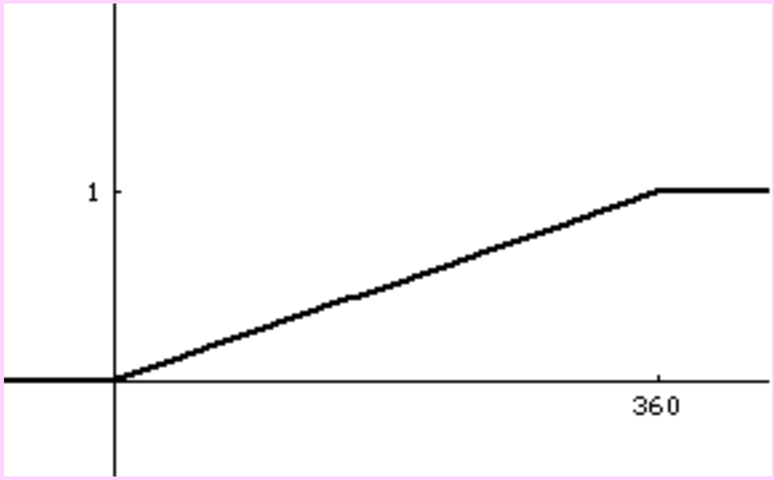
$$P(x) = \int_{-\infty}^{x} p(t)\,dt = \begin{cases} 0, \text{ if } x < 0 \\ \dfrac{x}{360}, \text{ if } 0 \leq x \leq 360 \\ 1, \text{ if } x > 360 \end{cases}$$

Example: A Spinner

Density Function:



Cumulative Distribution Function:



Properties of CDFs

- $P(x)$ is the probability of values less than x
  - If $P(x)$ is the cdf for the age in months of fish in a lake, then $P(10)$ is the probability a random fish is 10 months or younger.
- $P(x)$ goes to 0 as x gets smaller:

$$\lim_{x \to -\infty} P(x) = 0$$

(In many cases, we may reach 0.)

Properties of CDFs

- Conversely,

$$\lim_{x \to \infty} P(x) = 1$$

- $P(x)$ is non-decreasing.
  - The derivative is a density function, which cannot be negative.
  - Also, $P(4)$ *can't* be less than $P(3)$, for example.

Practice

Life expectancy (in days) of electronic component has density function $p(x) = \frac{1}{x^2}$, for $x \geq 1$, and $p(x) = 0$ for $x < 1$.

a) Probability component lasts between 0 and 1 day?

b) Probability it lasts between 0 and 10 days?

c) More than 10 days?

d) Find the CDF for the life expectancy.

Practice

Life expectancy (in days) of electronic component has density function $p(x) = \frac{1}{x^2}$, for $x \geq 1$, and $p(x) = 0$ for $x < 1$.

a) Probability component lasts between 0 and 1 day?

Zero!

$$\int_0^1 p(x)\,dx = \int_0^1 0\,dx = 0$$

Practice

Life expectancy (in days) of electronic component has density function $p(x) = \frac{1}{x^2}$, for $x \geq 1$, and $p(x) = 0$ for $x < 1$.

b) Probability it lasts between 0 and 10 days?

$$\int_0^{10} p(x)\,dx = \int_1^{10} \frac{1}{x^2}\,dx = -\frac{1}{x}\Big|_1^{10} = -\frac{1}{10} + 1 = 0.9$$

Practice

Life expectancy (in days) of electronic component has density function $p(x) = \frac{1}{x^2}$, for $x \geq 1$, and $p(x) = 0$ for $x < 1$.

c) More than 10 days?

$$\int_{10}^{\infty} p(x)\,dx = \lim_{b\to\infty}\int_{10}^{b} \frac{1}{x^2}\,dx$$

$$= \lim_{b\to\infty}\left[-\frac{1}{x}\right]_{10}^{b}$$

$$= \lim_{b\to\infty}\left(-\frac{1}{b} + \frac{1}{10}\right)$$

Practice

Life expectancy (in days) of electronic component has density function $p(x) = \frac{1}{x^2}$, for $x \geq 1$, and $p(x) = 0$ for $x < 1$.

c) More than 10 days?

$$\begin{aligned}\int_{10}^{\infty} p(x)\,dx &= \lim_{b\to\infty}\int_{10}^{b} \tfrac{1}{x^2}\,dx \\ &= \lim_{b\to\infty}\left[-\tfrac{1}{x}\right]_{10}^{b} \\ &= \lim_{b\to\infty}\left(-\tfrac{1}{b} + \tfrac{1}{10}\right) = 0.1\end{aligned}$$

Practice

Life expectancy (in days) of electronic component has density function $p(x) = \frac{1}{x^2}$, for $x \geq 1$, and $p(x) = 0$ for $x < 1$.

d) Find the CDF for life expectancy.

If $x < 1$, then $\int_{-\infty}^{x} p(t)\,dt = 0$

If $x \geq 1$, then
$$\int_{-\infty}^{x} p(t)\,dt = \int_{1}^{x} t^{-2}\,dt$$
$$= -\frac{1}{t}\Big|_{1}^{x}$$
$$= -\frac{1}{x} + 1$$

Practice

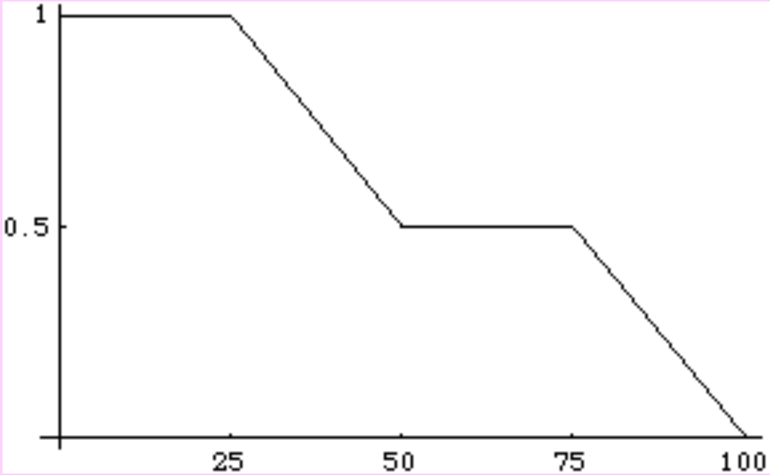
Life expectancy (in days) of electronic component has density function $p(x) = \frac{1}{x^2}$, for $x \geq 1$, and $p(x) = 0$ for $x < 1$.

d) Find the CDF for life expectancy.

$$P(x) = \begin{cases} 0, \text{ if } x < 1 \\ 1 - \frac{1}{x}, \text{ if } x \geq 1 \end{cases}$$

Example

Someone claims this is the CDF for grades on the 2015 final exam.



Probability a random student scored….

- 25 or lower?

    1, or 100%.

- 50 or lower?

    0.5, or 50%.

Conclusion?

They're lying! This *cannot be* a cumulative distribution function! (It *decreases*.)

Relating the CDF and DF

According the FTC version 2, since

$$P(x) = \int_{-\infty}^{x} p(t)\,dt,$$

then $P'(x) = p(x)$.

Relating the CDF and DF

According the FTC version 2, since

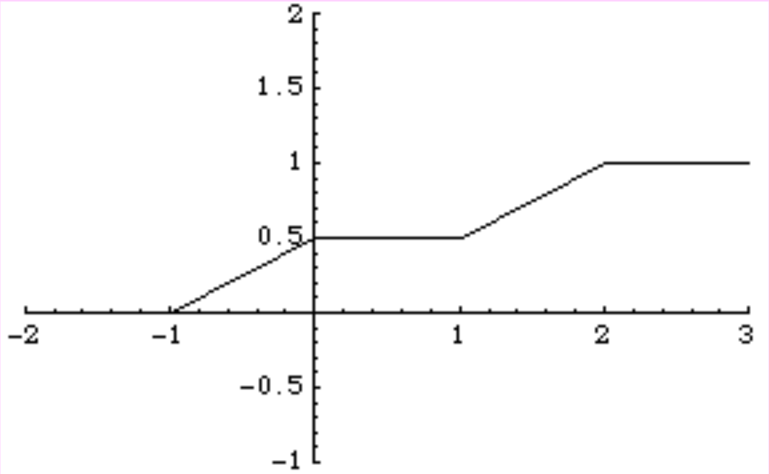
$$P(x) = \int_{-\infty}^{x} p(t)\,dt,$$

then $P'(x) = p(x)$.

So the density function is the derivative, or rate of change, of the cumulative distribution function.

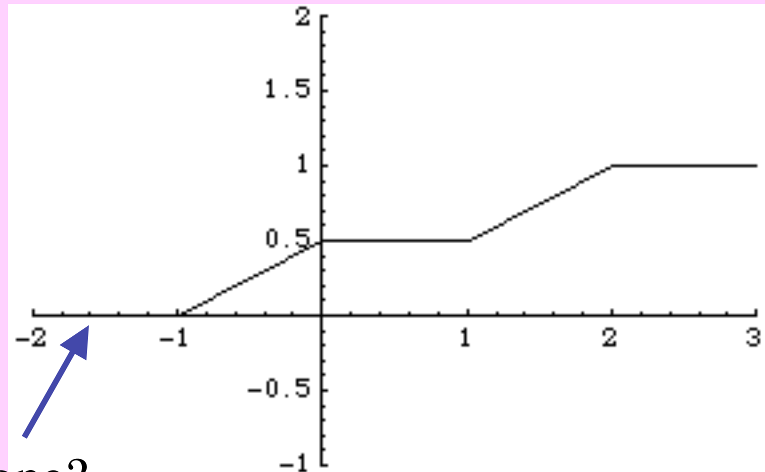
Example

Sketch the density function for the cdf shown.



Example

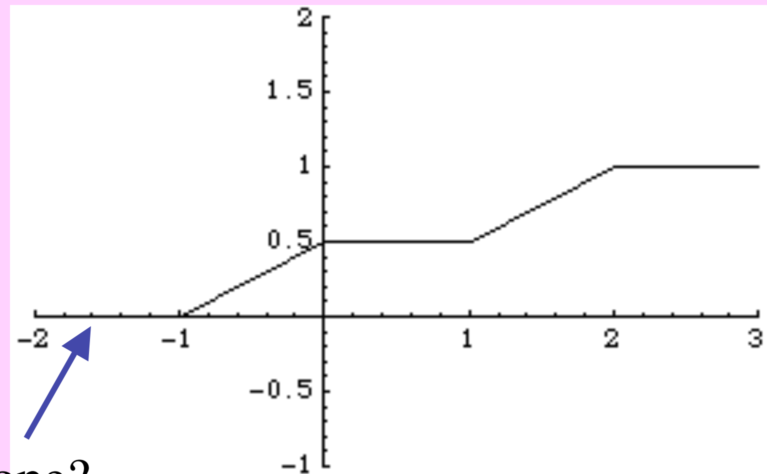
Sketch the density function for the cdf shown.



Slope?

Example

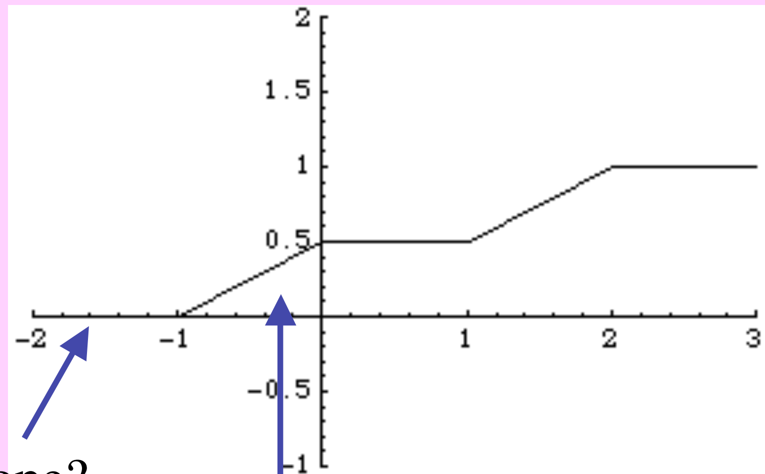
Sketch the density function for the cdf shown.



Slope?
0

Example

Sketch the density function for the cdf shown.

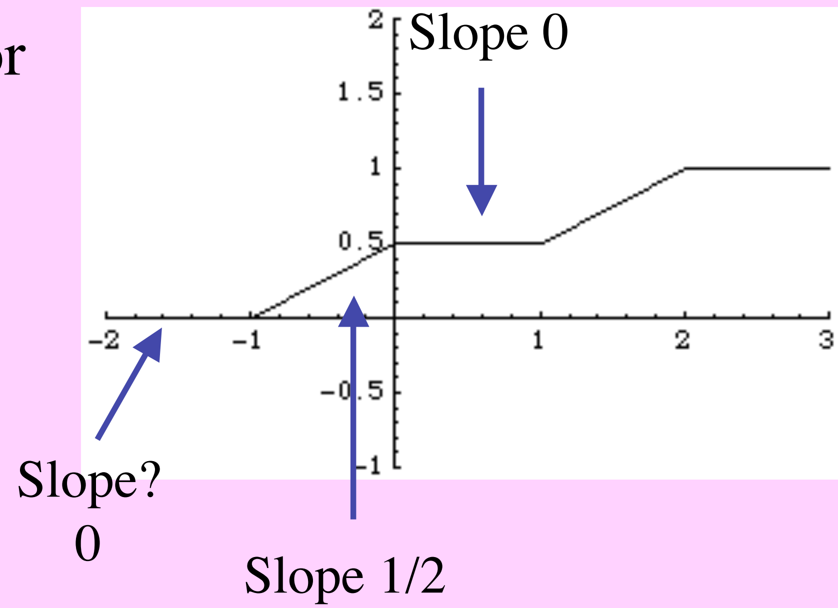


Slope?
0

Slope 1/2

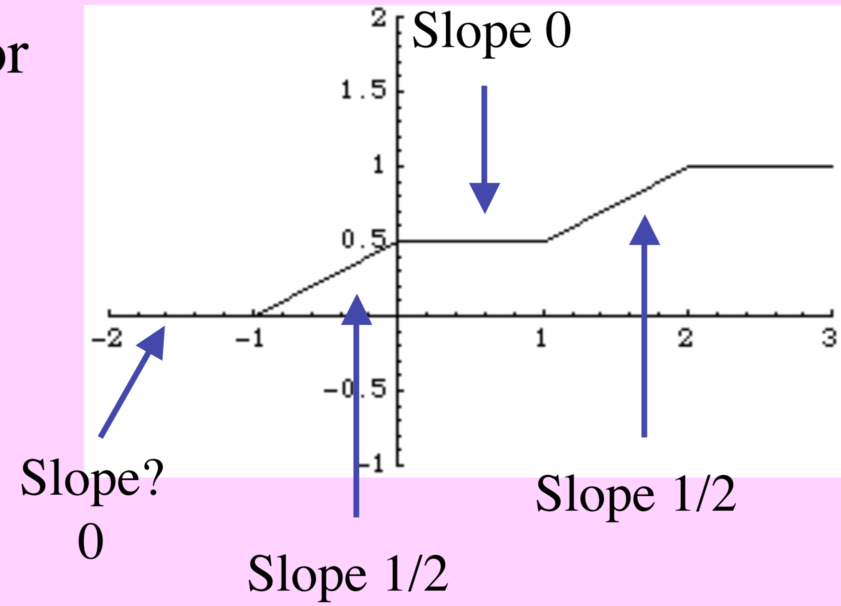
Example

Sketch the density function for the cdf shown.



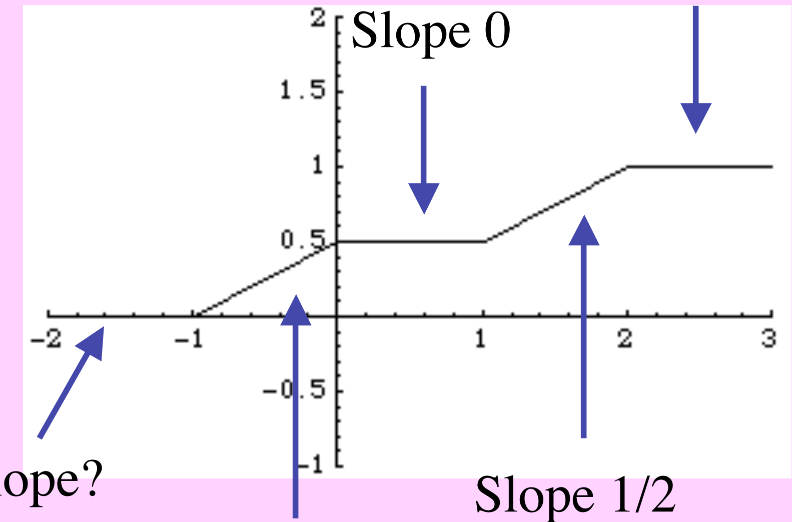
Example

Sketch the density function for the cdf shown.



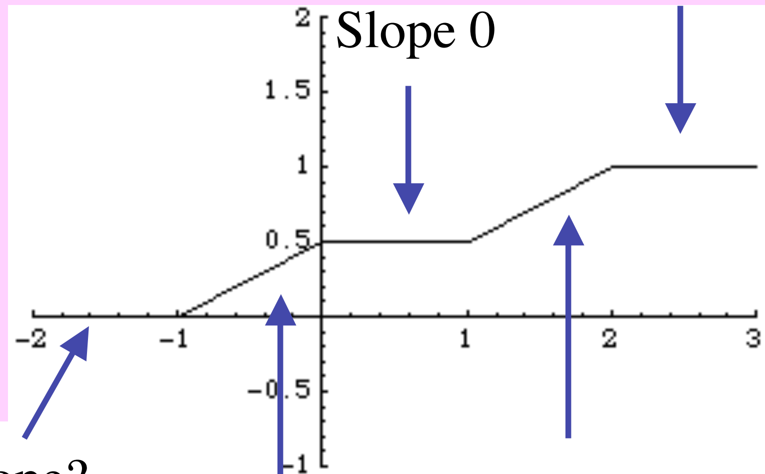
Example

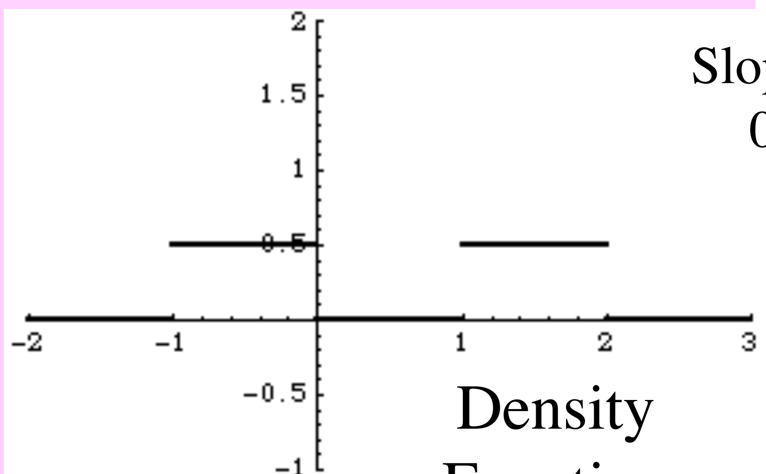
Sketch the density function for the cdf shown.



Example

Sketch the density function for the cdf shown.





DF vs. CDF

You must know which is which.

We work differently with density functions than with cumulative distribution functions.

Another Example

Suppose the cumulative distribution function for the height of trees in a forest (in feet) is given by

$$P(x) = \begin{cases} 0, & x < 0 \\ 0.1x - 0.0025x^2, & 0 \le x \le 20 \\ 1, & x > 20 \end{cases}$$

Find the height, x for which exactly half the trees are *taller* than x feet, and half the trees are *shorter* than x.

In case - Quadratic Formula: $x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

Another Example

$$P(x) = \begin{cases} 0, & x < 0 \\ 0.1x - 0.0025x^2, & 0 \le x \le 20 \\ 1, & x > 20 \end{cases}$$

Find the height, x for which exactly half the trees are *taller* than x feet, and half the trees are *shorter* than x.

$$0.1x - 0.0025x^2 = 0.5$$

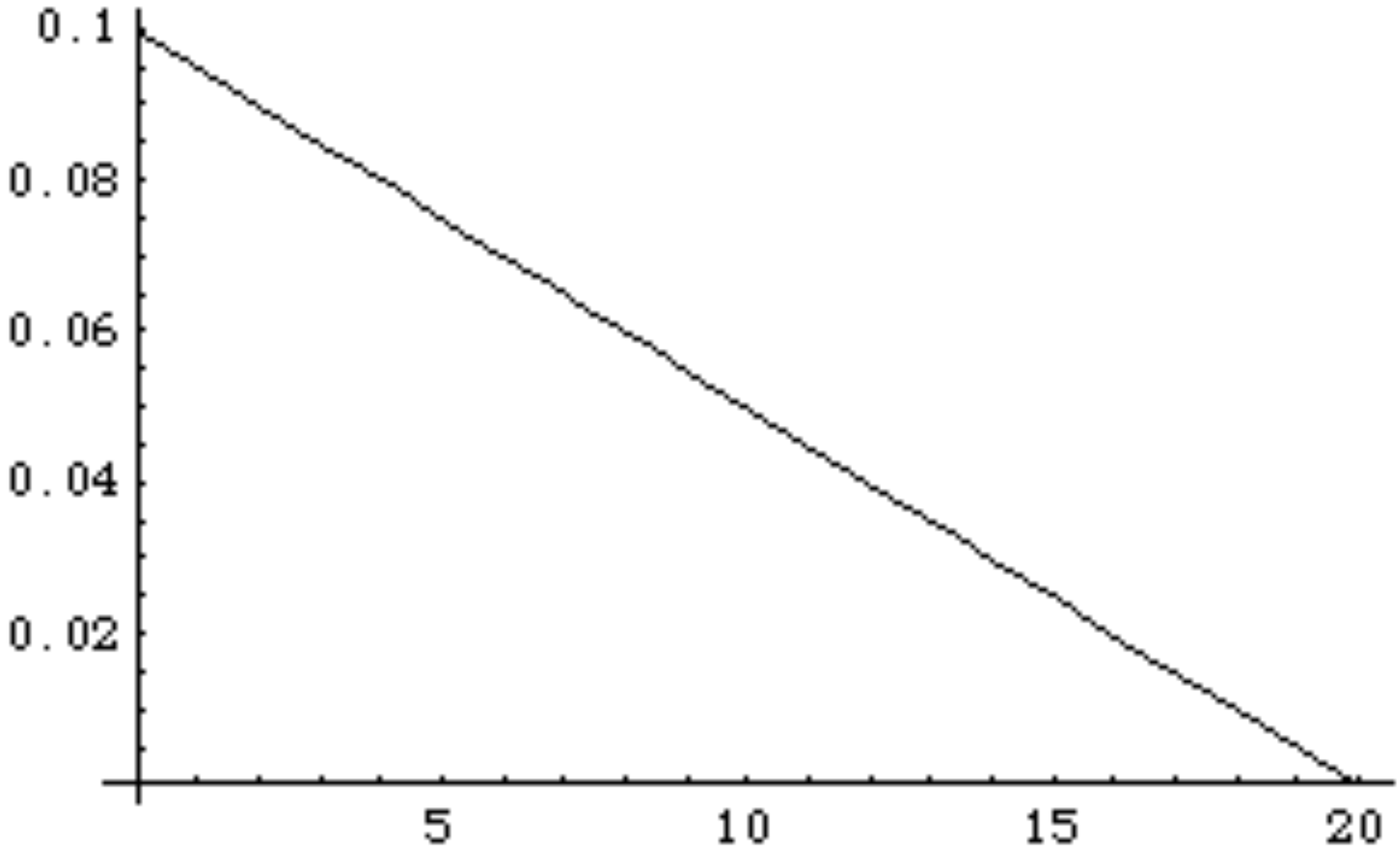
$$-0.0025x^2 + 0.1x - 0.5 = 0$$

$$-0.0025(x^2 - 40x + 200) = 0$$

$$\text{using quadratic}: \quad x = 20 \pm 10\sqrt{2} = 20 - 10\sqrt{2} \approx 5.86 \text{ ft}$$

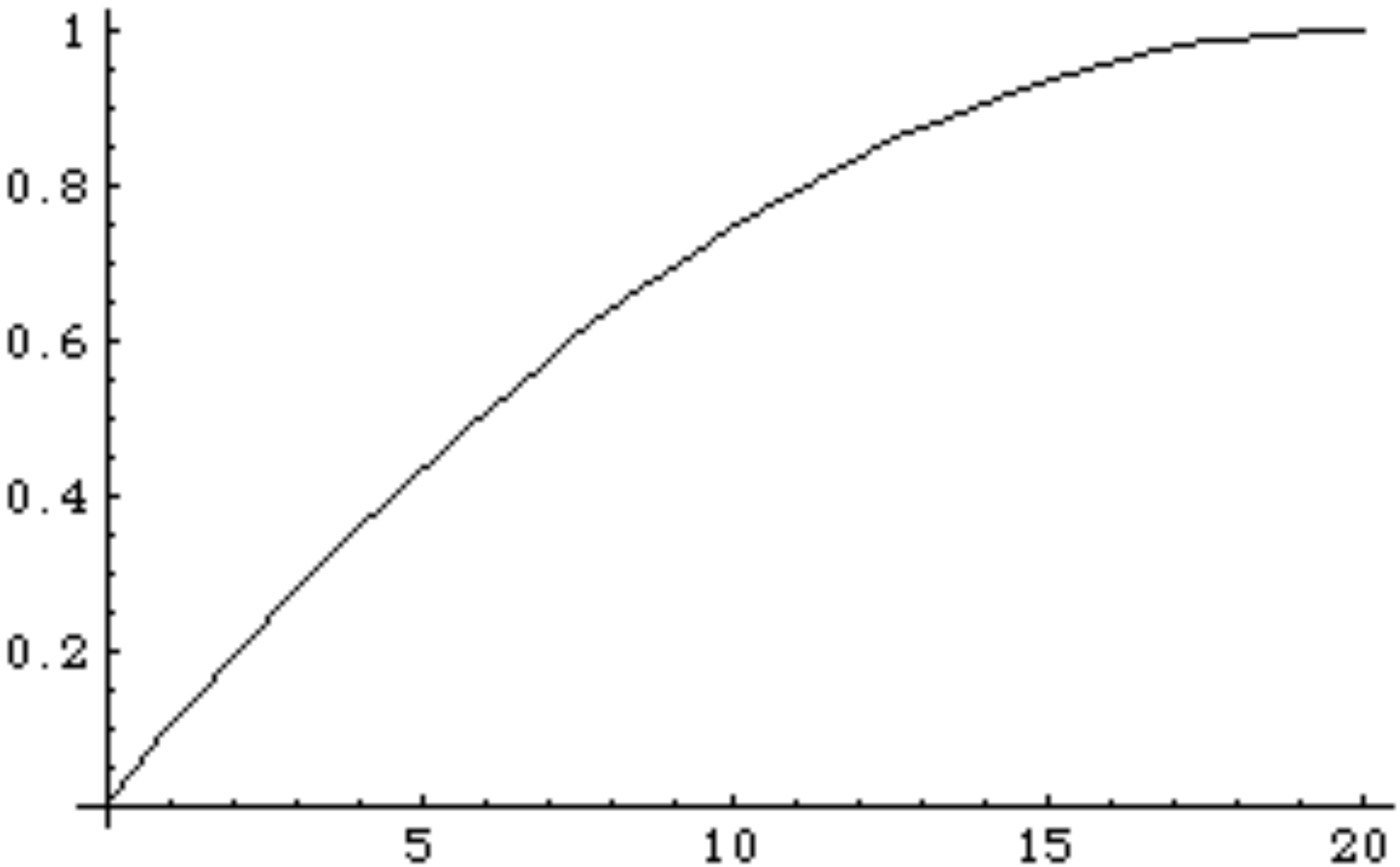
```
p[x_] := 0.1 - .005 * x

Plot[p[x], {x, 0, 20}];
```



```
P[x_] := 0.1*x - .0025*x^2

Plot[P[x], {x, 0, 20}];
```



Example

Assume $p(x)$ is the density function

$$p(x)=\begin{cases}\frac{1}{2}, & if\ 1\le x\le 3\\ 0, & if\ x<1\ \ or\ \ x>3\end{cases}$$

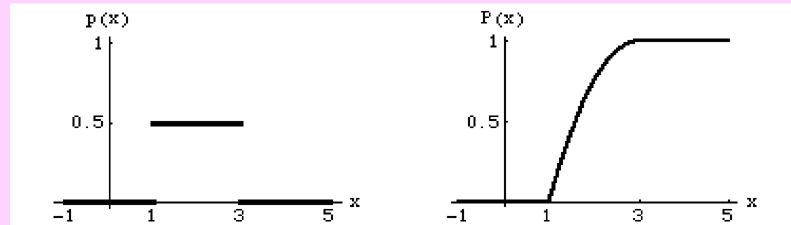
Let $P(x)$ represent the corresponding Cumulative Distribution function.

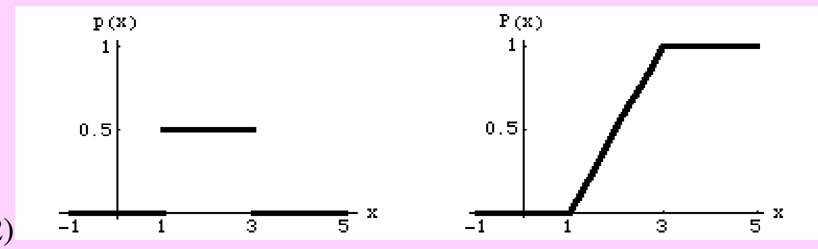
Then

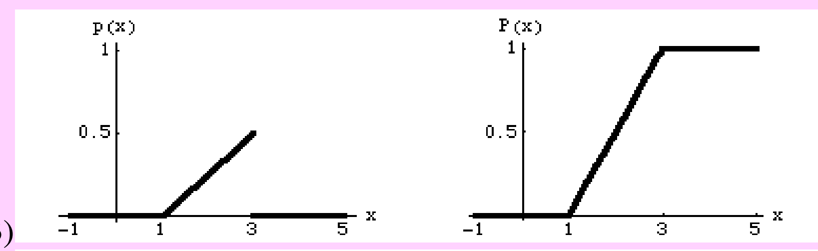
(a) For the density function given above, what is the cumulative distribution function $P(x)$ for values of x between 1 and 3?

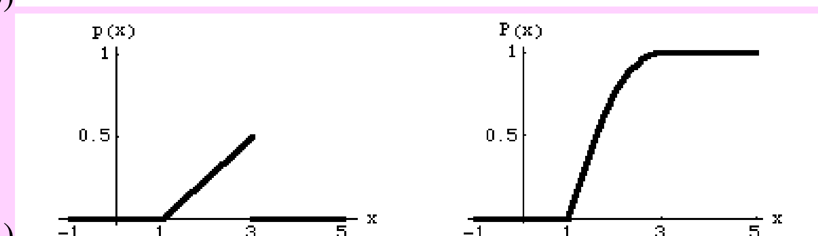
(b) P(5) = ?

(c) Choose the correct pair of graphs for $p(x)$ and $P(x)$.

(1)


(2)


(3)


(4)


In-class Assignment

The time to conduct a routine maintenance check on a machine has a cumulative distribution function $P(t)$, which gives the fraction of maintenance checks completed in time less than or equal to t minutes. Values of $P(t)$ are given in the table.

t, minutes	0	5	10	15	20	25	30
$P(t)$, fraction completed	0	.03	.08	.21	.38	.80	.98

a. What fraction of maintenance checks are completed in 15 minutes or less?

b. What fraction of maintenance checks take longer than 30 minutes?

c. What fraction take between 10 and 15 minutes?

In-class Assignment

The time to conduct a routine maintenance check on a machine has a cumulative distribution function $P(t)$, which gives the fraction of maintenance checks completed in time less than or equal to t minutes. Values of $P(t)$ are given in the table.

t, minutes	0	5	10	15	20	25	30
$P(t)$, fraction completed	0	.03	.08	.21	.38	.80	.98

a. What fraction of maintenance checks are completed in 15 minutes or less? 21%

b. What fraction of maintenance checks take longer than 30 minutes? 2%

c. What fraction take between 10 and 15 minutes? 13%

Lesson 21

Mean and Median

The Median

The median is the halfway point: Half the population has a lower value, and half a higher value.

If $p(x)$ is a density function, then the median of the distribution is the point M such that

$$\int_{-\infty}^{M} p(x)\,dx = 0.5$$

Example

Suppose an insect life span of $p(x) = \frac{1}{72}x, \quad 0 \le x \le 12$ months, and 0 elsewhere. What's the median life span?

We need

$$\frac{1}{2} = \int_0^M \frac{1}{72}x\,dx = \frac{1}{144}x^2\Big|_0^M = \frac{1}{144}M^2$$

Then solve for *M*:

$$\frac{1}{2} = \frac{1}{144}M^2$$

$$72 = M^2$$

$$M = \sqrt{72} \approx 8.49$$ months

Median From CDF

If we have a CDF $P(x)$, the median M is the point where

$$P(M) = 0.5$$

(This of course says 1/2 of the population has a value less than M!)

Example

CDF: $P(t) = t^2, \ 0 \leq t \leq 1.$

What's the median of this distribution?

Example

CDF: $P(t) = t^2,\ \ 0 \le t \le 1.$

What's the median of this distribution?

$$P(M) = \tfrac{1}{2}$$

$$M^2 = \tfrac{1}{2}$$

$$M = \sqrt{\tfrac{1}{2}} \approx 0.707$$

The Mean

What most people think of as the average:

Add up n values and divide by n.

How should we represent this for something represented by a density function $p(x)$?

The Mean

We define the mean of a distribution given by density function $p(x)$ to be

$$\int_{-\infty}^{\infty} x\, p(x)\, dx$$

Example

Previous insect population example:

$p(x) = \frac{1}{72}x, \quad 0 \le x \le 12$ and $p(x) = 0$ elsewhere.

What's the mean lifespan?

$$\begin{aligned}\int_{-\infty}^{\infty} x\,p(x)\,dx &= \int_0^{12} x\left(\tfrac{1}{72}x\right)dx \\ &= \int_0^{12} \tfrac{1}{72}x^2\,dx \\ &= \tfrac{1}{216}x^3\Big|_0^{12} \\ &= 8\end{aligned}$$

Example

Previous insect population example:

$p(x) = \frac{1}{72}x, \quad 0 \le x \le 12$ and $p(x) = 0$ elsewhere.

What's the mean lifespan? 8 months.

The mean is slightly below the median. This means

Example

Previous insect population example:

$p(x) = \frac{1}{72}x, \quad 0 \le x \le 12$ and $p(x) = 0$ elsewhere.

What's the mean lifespan? 8 months.

The mean is slightly below the median. This means more than half the insects live longer than the mean lifespan.

Example 2

From "Another Example",

$$p(x) = 0.1 - 0.005x$$

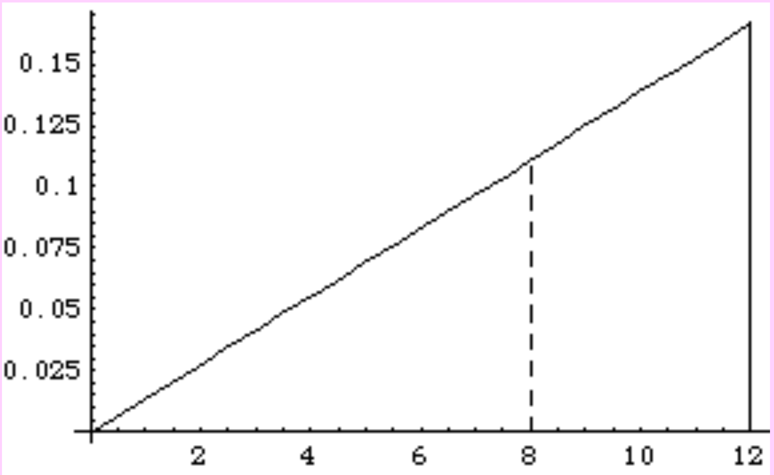
$$mean = \int_{-\infty}^{\infty} x\,p(x)dx = \int_0^{20} x(0.1 - 0.005x)dx =$$

$$\int_0^{20} 0.1x - 0.005x^2 dx = 0.05x^2 - 0.001\overline{6}x^3\Big|_0^{20}$$

$$= 6.64 \text{ ft}$$

Geometry

Turns out: the mean is the point where the graph of the distribution would balance if we cut it out.



Example

Find the mean and median for $p(x) = \tfrac{1}{9}x^2,\ 0 < x < 3$ and zero elsewhere.

Median: Solve for *M*

$$\int_0^M \tfrac{1}{9}x^2 dx = \tfrac{1}{2}$$

$$\tfrac{1}{27}M^3 = \tfrac{1}{2}$$

$$M = \left(\tfrac{27}{2}\right)^{1/3} \approx 2.38$$

Example

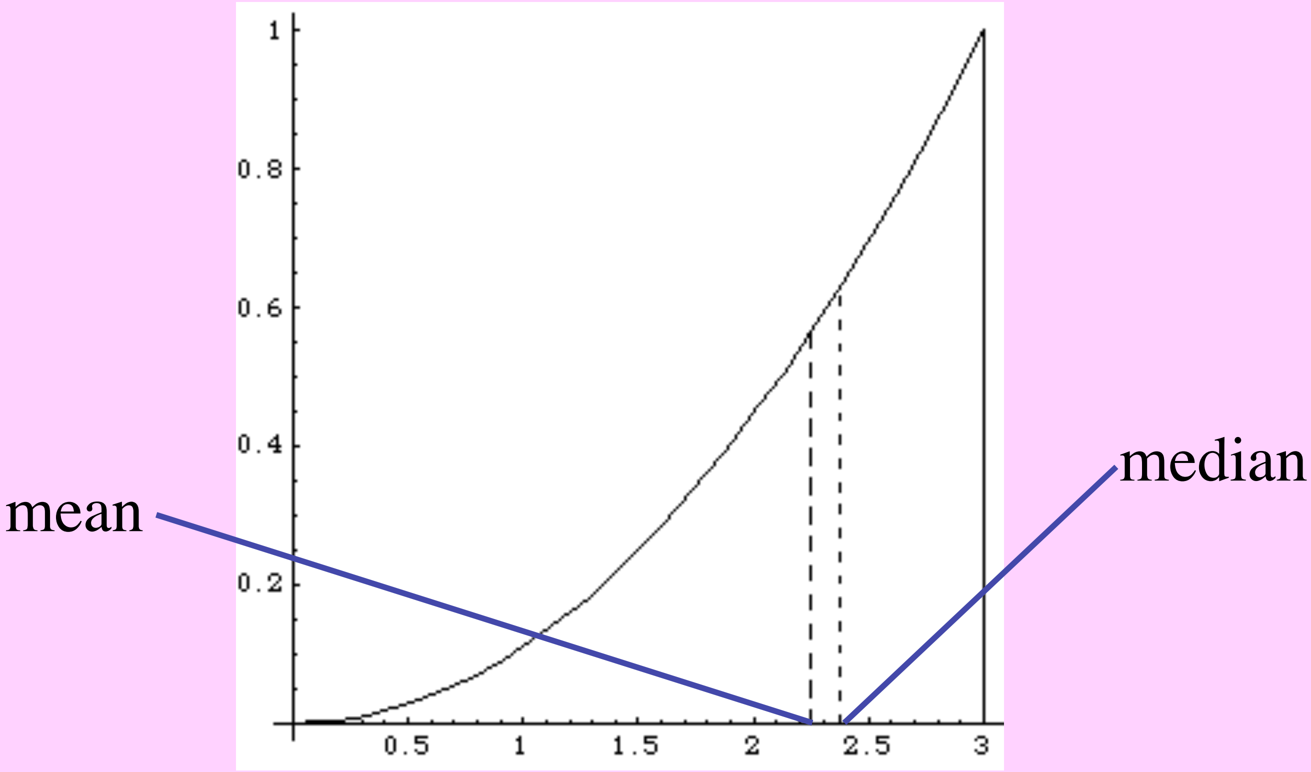
Find the mean and median for $p(x) = \frac{1}{9}x^2,\ 0 < x < 3$ and zero elsewhere.

Mean: Evaluate

$$\begin{aligned}
\int_{-\infty}^{\infty} x\, p(x)\, dx &= \int_0^3 x\left(\tfrac{1}{9}x^2\right)dx \\
&= \int_0^3 \tfrac{1}{9}x^3 dx \\
&= \tfrac{1}{36}x^4\Big|_0^3 = 2.25
\end{aligned}$$

Example

Find the mean and median for $p(x) = \frac{1}{9}x^2,\ 0 < x < 3$ and zero elsewhere.



The Normal Distribution

A special distribution:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2}$$

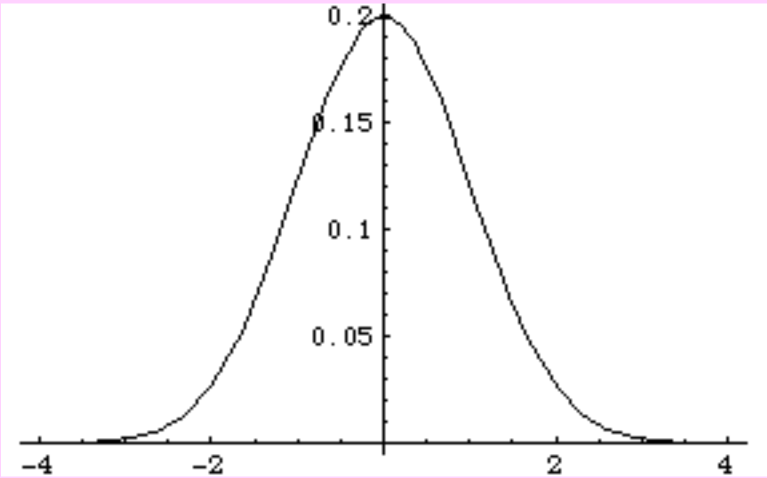
The Normal Distribution

A special distribution:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2}$$

The Normal Distribution

Here's an example with $\mu = 0$ and $\sigma = 1$:



The distribution is symmetric about $x = \mu$.

So the mean is μ, and so is the median.

σ is called the *standard deviation,* and describes how spread out the curve is.

The Normal Distribution

Unfortunately, e^{-x^2} has no elementary antiderivative.

So we must use numerical means to evaluate integrals involving the normal distribution.

Example

Suppose a density function for number of people (in millions) on the internet at one time were

$$p(x) = \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8}$$

What are the mean and median?

Example

Suppose a density function for number of people (in millions) on the internet at one time were

$$p(x) = \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8}$$

What are the mean and median?

Here, μ = 10, so both the mean and median are 10 million people.

Example

Suppose a density function for number of people (in millions) on the internet at one time were

$$p(x) = \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8}$$

What is the standard deviation?

Example

Suppose a density function for number of people (in millions) on the internet at one time were

$$p(x) = \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8}$$

What is the standard deviation?

Here, we see that $\sigma = 2$, so 2 million people.

Example

Suppose a density function for number of people (in millions) on the internet at one time were

$$p(x) = \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8}$$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

What is the standard deviation?

Here, we see that σ = 2, so 2 million people.

Example

Suppose a density function for number of people (in millions) on the internet at one time were

$$p(x) = \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8}$$

Write an integral for the fraction of time between 8 and 12 million people are on the internet.

Example

Suppose a density function for number of people (in millions) on the internet at one time were

$$p(x) = \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8}$$

Write an integral for the fraction of time between 8 and 12 million people are on the internet.

$$\int_8^{12} \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8} dx$$

Evaluating the Integral

How to calculate the required integral?

$$\int_{8}^{12} \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8} dx$$

Evaluating the Integral

How to calculate the required integral?

$$\int_{8}^{12} \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8} dx$$

We could use our Simpson's rule calculator:

Evaluating the Integral

How to calculate the required integral?

$$\int_8^{12} \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8} \, dx$$

We could use our Simpson's rule calculator:

n	Δ*x*	Result
2	2	0.693237
4	1	0.683058
8	0.5	0.682711
16	0.25	0.682691
32	0.125	0.68269

Evaluating the Integral

How to calculate the required integral?

$$\int_8^{12} \frac{1}{2\sqrt{2\pi}} e^{-(x-10)^2/8}\,dx$$

We could use our Simpson's rule calculator:

n	**Δ*x***	**Result**
2	2	0.693237
4	1	0.683058
8	0.5	0.682711
16	0.25	0.682691
32	0.125	0.68269

Seems to converge to about 0.68.

In Class Assignment

Find the mean and median of the distribution with density function:

(calculations should be to 2 decimal places)

$$p(x) = \frac{1}{4}x^3 \quad \text{for} \quad 0 < x < 2 \text{ and } 0 \text{ elsewhere}$$