Introductory Statistics Lectures

Probability density functions

The normal distribution

Anthony Tanbakuchi
Department of Mathematics
Pima Community College

(Compile date: Tue May 19 14:49:36 2009)

Contents

1 Probability density functions

1.1 Introduction

R TIP OF THE DAY: GRAPHING FUNCTIONS

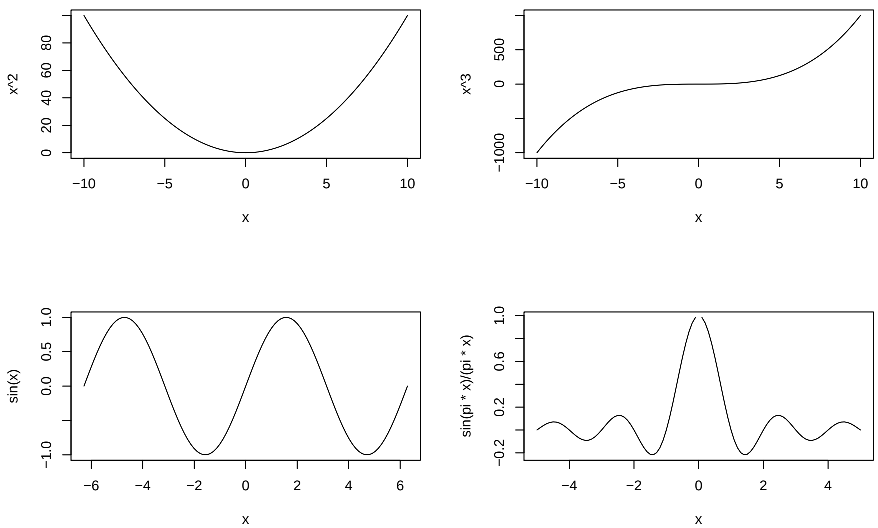
Graphing functions:

`curve(expression, xmin, xmax)`

- **expression** an expression or function involving x
- **xmin** min value of x to plot
- **xmax** max value of x to plot

R Command

```
R: par(mfrow = c(2, 2))
R: curve(x^2, −10, 10)
R: curve(x^3, −10, 10)
R: curve(sin(x), −2 * pi, 2 * pi)
R: curve(sin(pi * x)/(pi * x), −5, 5)
```



UNIFORM DISTRIBUTION

DEFINITION 1.1

UNIFORM DISTRIBUTION $f(x)$.

Occurs when the probability of a continuous random variable is equal across a range of values.

R COMMAND

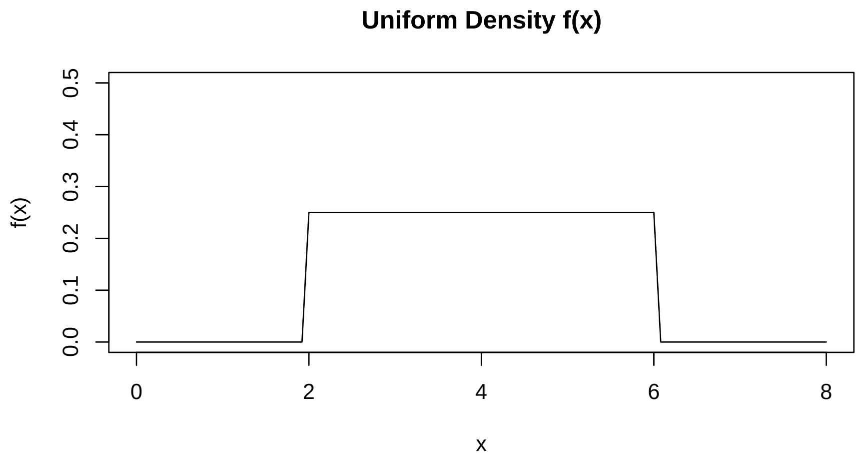
UNIFORM DENSITY:

`dunif(x, min=0, max=1)`

Useful for graphing, not useful for directly finding probabilities.

In R, all PDF's have a "d" prefix for density.

```
R: curve(dunif(x, min = 2, max = 6), 0, 8, ylim = c(0,
+     0.5), ylab = "f(x)", main = "Uniform Density f(x)")
```



Probability is area!

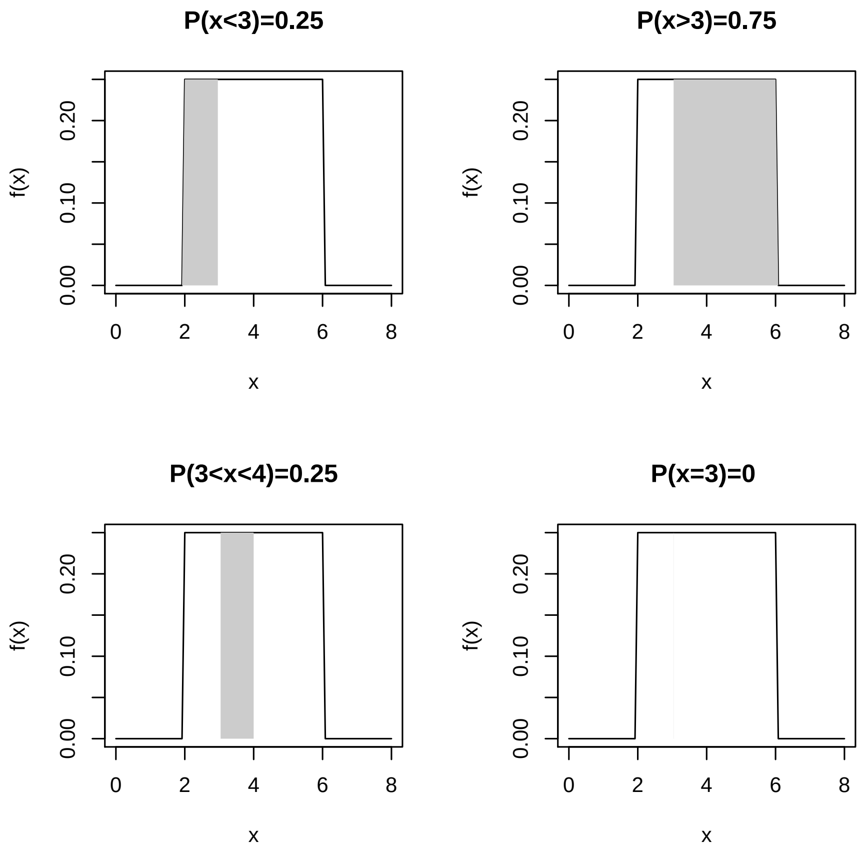
FINDING PROBABILITIES FROM DENSITY FUNCTIONS

Finding probabilities
Area represents probability!

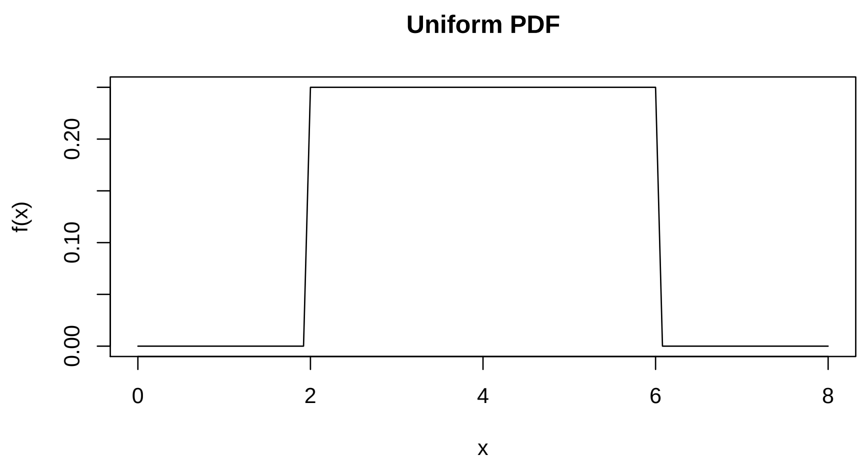
$$P(x < a) = \int_{-\infty}^{a} f(x)\,dx \qquad \text{(area to the left of } a\text{)}$$

$$P(a < x < b) = \int_{a}^{b} f(x)\,dx \qquad \text{(area between } a \text{ and } b\text{)}$$

$$P(x > a) = \int_{a}^{\infty} f(x)\,dx \qquad \text{(area to the right of } a\text{)}$$



Finding probabilities for uniform density is easy: width × height.
Use the density below to answer the following question.



Question 1. Shade the region representing $P(x < 5)$ and find the probability.

1.2 Cumulative distribution functions

CUMULATIVE DISTRIBUTION FUNCTION (CDF) $F(x)$. DEFINITION 1.2

Gives the **area to the left** of x on the probability density function.

$$P(x < a') = F(a') \quad (1)$$

$$= \int_{-\infty}^{a'} f(x)\,dx \quad (2)$$

$F(x)$ is **the** tool for finding probabilities of continuous random variables.

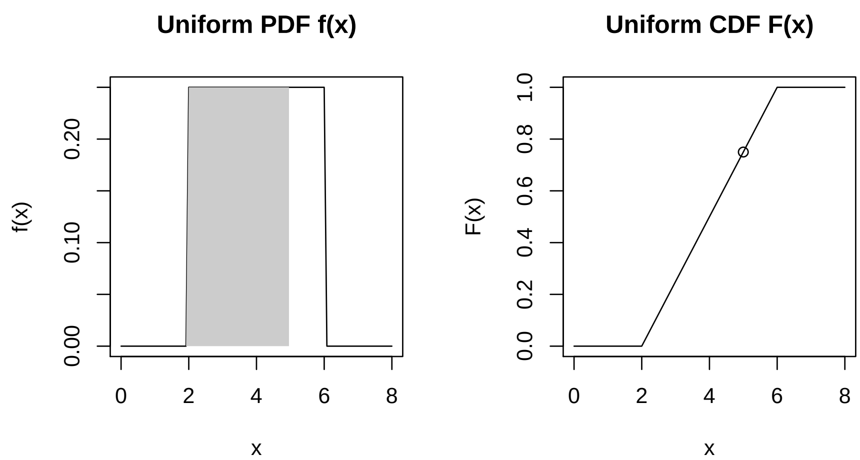
UNIFORM CDF: R COMMAND

```
punif(x, min=0, max=1)
```

Gives the **area to the left** of the uniform density at x.

In R, all CDF's have a "p" prefix for probability.

Example 1. Find $P(x < 5)$




```
R: punif(5, min = 2, max = 6)
[1] 0.75
```

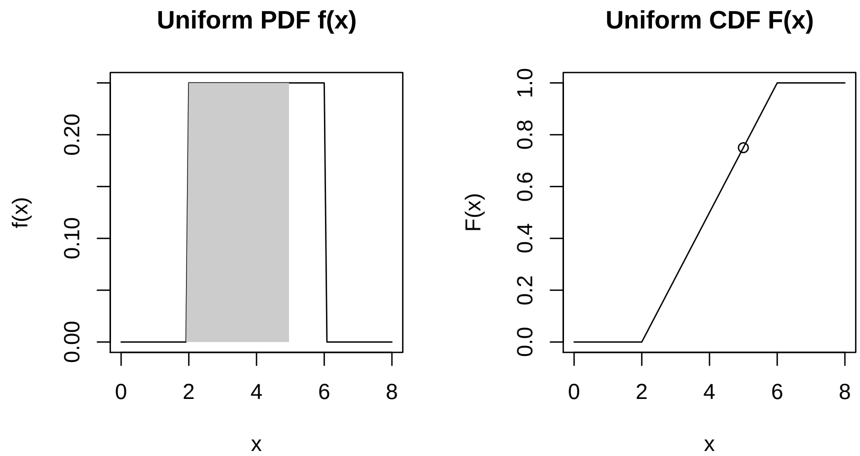
FINDING PROBABILITIES USING CDF'S

What CDF gives us
Only give area to the left!

$$P(x < a) = F(a) \qquad \text{(area to the left of } a)$$
$$P(x > a) = ? \qquad \text{(area to the right of } a)$$
$$P(a < x < b) = ? \qquad \text{(area between } a \text{ and } b)$$

Using CDF to find P(x>a)

Example 2. Find $P(x > 5)$.

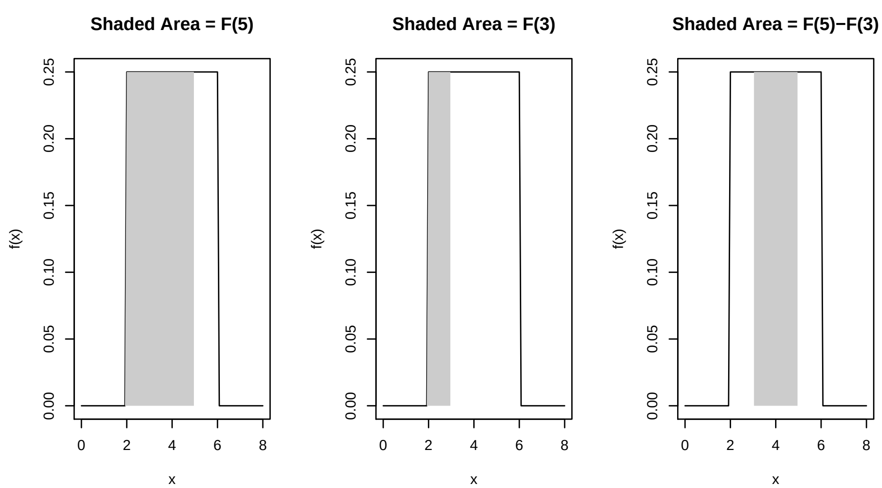


$P(x > 5) = 1 - P(x < 5) = 1 - F(5)$

```
R: 1 - punif(5, min = 2, max = 6)
[1] 0.25
```

Using CDF to find P(a<x<b)

Example 3. Find $P(3 < x < 5)$.



$P(3 < x < 5) = F(5) - F(3)$ (always subtract larger from smaller)

```
R: punif(5, min = 2, max = 6) - punif(3, min = 2,
+       max = 6)
[1] 0.5
```

Finding probabilities with CDF's

Using F(x) to find probabilities:

$$P(x < a) = F(a) \qquad \text{(area to the left of } a)$$
$$P(x > a) = 1 - F(a) \qquad \text{(area to the right of } a)$$
$$P(a < x < b) = F(b) - F(a) \qquad \text{(area between } a \text{ and } b)$$

You must know how to use this!

INVERSE CUMULATIVE DISTRIBUTION FUNCTIONS

DEFINITION 1.3

INVERSE CUMULATIVE DISTRIBUTION FUNCTIONS CDF^{-1}.
Finds the value of x that has an area p to the left. (Inverse operation of CDF).

R COMMAND

UNIFORM INVERSE CDF:

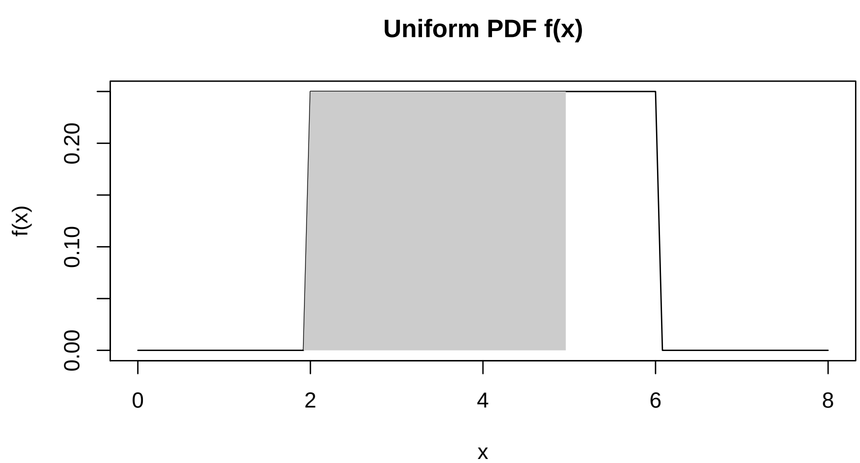
```
qunif(p, min=0, max=1)
```

Finds x with area p **to the left** on the density function.

In R, all inverse CDF's have a "q" prefix for quantile.

Using inverse CDF to find x given p

Example 4. Find x' such that $P(x < x') = 0.75$ (the value of x that has an area to the left of 0.75).




```
R: qunif(0.75, min = 2, max = 6)
 [1] 5
```

Question 2. Find x' such that $P(x > x') = 0.25$ (the value of x that has an area to the right of 0.25).

1.3 Normal distribution

NORMAL PROBABILITY DENSITY FUNCTION $f(x)$. DEFINITION 1.4

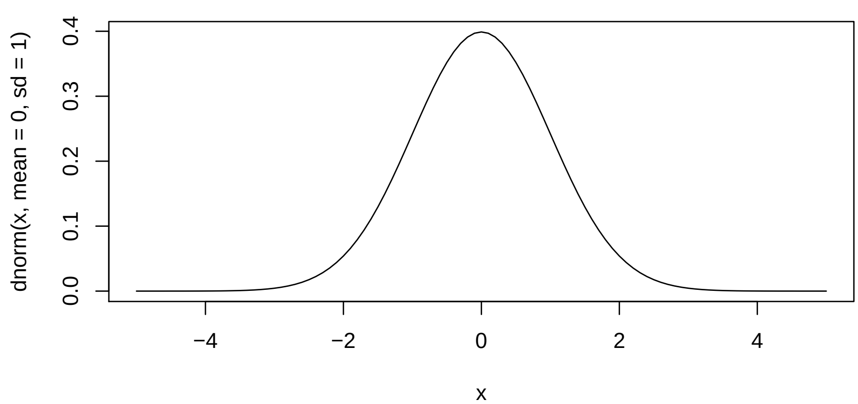
$$f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \quad (3)$$

characterized by μ and σ.
Occurs frequently in nature.

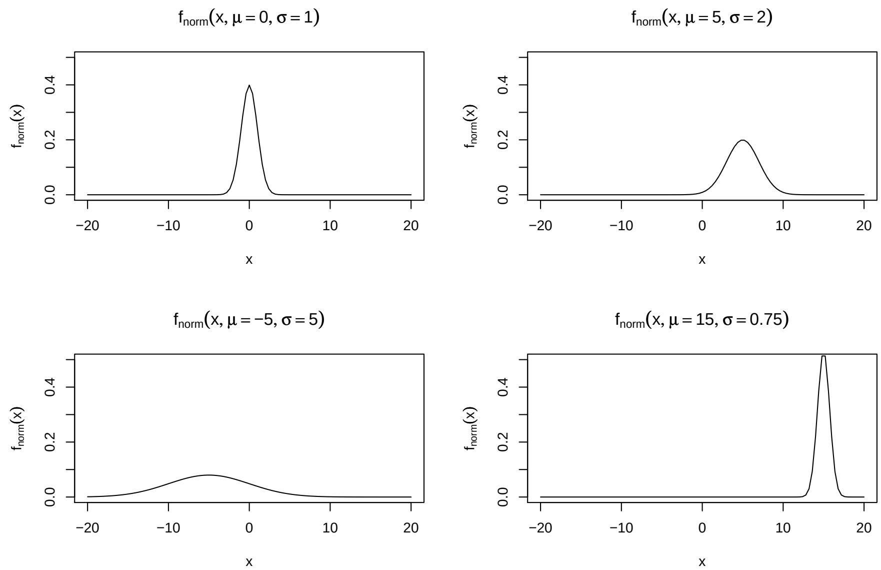
NORMAL DENSITY: R COMMAND
`dnorm(x, mean=0, sd=1)`
By default it is the standard normal density.

Visualizing the normal distribution

```
R: curve(dnorm(x, mean = 0, sd = 1), -5, 5)
```



Visualizing effect of μ, σ



Area under curve is always 1.

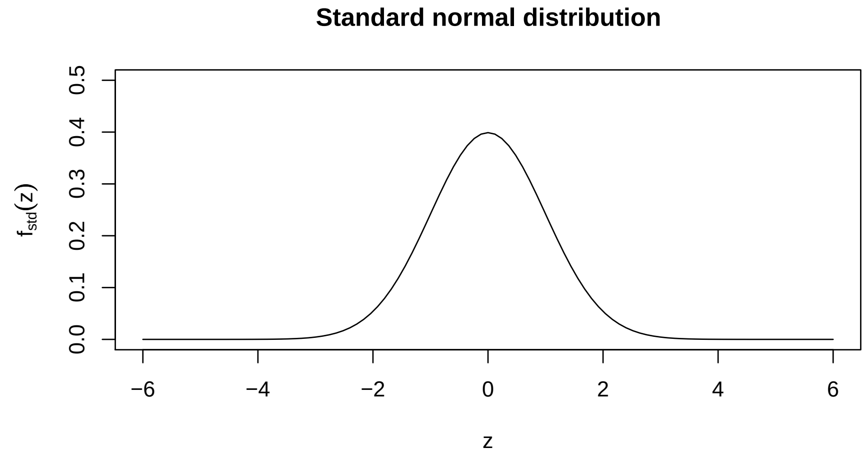
STANDARD NORMAL DISTRIBUTION

DEFINITION 1.5 STANDARD NORMAL DISTRIBUTION $f(z)$.

A normal distribution with $\mu = 0$ and $\sigma = 1$. If you convert normally distributed x data into z-scores, you will have a standard normal distribution.

Since there are an infinite set of normal distributions, historically we converted x to z and then only had **one** standard normal distribution and **one** standard normal cumulative distribution $F(z)$. A single table of $F(z)$ could then be used to solve most probability questions involving normal distributions.

With computers, we can directly use any specific normal cumulative distribution $F(x)$ and very accurately find probabilities.



FINDING PROBABILITIES INVOLVING THE NORMAL DISTRIBUTION

NORMAL CDF:
`pnorm(x, mean=0, sd=1)`
Gives the **area to the left** of the normal density at x.

R COMMAND

NORMAL INVERSE CDF:
`qnorm(p, mean=0, sd=1)`
Finds x with area p to **the left on** the density function.

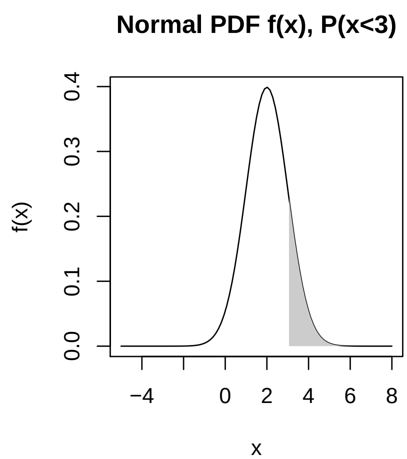
R COMMAND

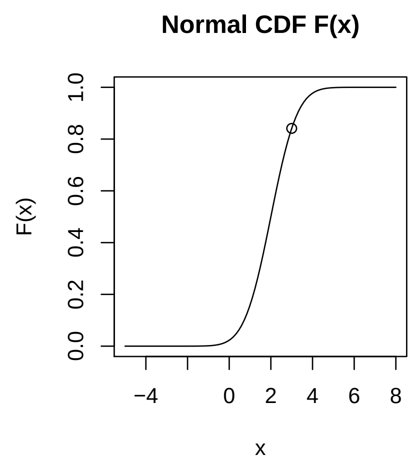
Tips for solving probabilities involving normal dist.

1. Determine μ and σ.
2. Sketch the PDF & area representing probability.
3. If asked to find probability use CDF: R function `pnorm(x,...)` to find probability p.
4. If asked to find value of x corresponding to probability use CDF^{-1}: R function `qnorm(p,...)` to find the value of x.
5. If working with **upper tail** be sure to take compliment!
   Be careful, if you want to find the value of x that has an area p to the right you need to use `qnorm(1-p, ...)` .

EXAMPLES

Example 5. Given $\mu = 2$ and $\sigma = 1$, find $P(x > 3)$.





$P(x > 3) = 1 - F(3)$

```
R: p = 1 - pnorm(3, mean = 2, sd = 1)
R: p
[1] 0.15866
```

Thus, $P(x > 3) = 0.159$

Now lets look at the inverse problem:

Example 6. Given $\mu = 2$ and $\sigma = 1$, what value of x' satisfies $P(x > x') = 0.159$?

$x' = F^{-1}(1 - 0.159)$

```
R: p
[1] 0.15866
R: x = qnorm(1 - p, mean = 2, sd = 1)
R: x
[1] 3
```

Note where we had to take the compliment!

Thus, the value of x that has an area 0.159 to the right is 3!

1.4 Summary

- For discrete random variables, probability is given by the distribution $p = P(x_i)$. ("d" prefix in R.)
  - For the binomial distribution, the probability of a specific number of successes x is `p = dbinom(x,n,p)` .
- For continuous variables, probability is **area** on density $f(x)$.
  - Use CDF's $F(x)$ to find probabilities. ("p" prefix in R)

$$P(x < x') = F(x') \qquad \text{(area to the left of } x')$$
$$P(x > x') = 1 - F(x') \qquad \text{(area to the right of } x')$$
$$P(a < x < b) = F(b) - F(a) \qquad \text{(area between } a \text{ and } b)$$

  - Use inverse CDF's $F^{-1}(p)$ to find specific value of x' in $p = P(x < x')$ given probability p. ("q" prefix in R)
    $x' = F^{-1}(p)$

For the normal distribution:

- CDF $p = F(x)$: `p=pnorm(x, mean=0, sd=1)`

$$P(x < x') = \texttt{pnorm(x',...)}$$
$$P(x > x') = 1 - \texttt{pnorm(x',...)}$$
$$P(a < x < b) = \texttt{pnorm(b,...)} - \texttt{pnorm(a,...)}$$

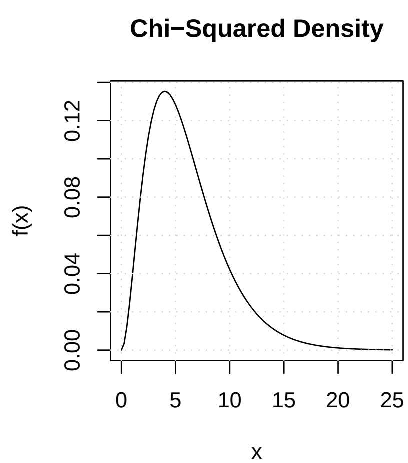
  where "..." is " `mean=`μ`, sd=`σ ".
- Use inverse CDF's $x = F^{-1}(p)$ to find x given probability p. ("q" prefix in R)
  To find x' in $p = P(x < x')$: `x'=qnorm(p, ...)`
  To find x' in $p = P(x > x')$: `x'=qnorm(1-p, ...)`

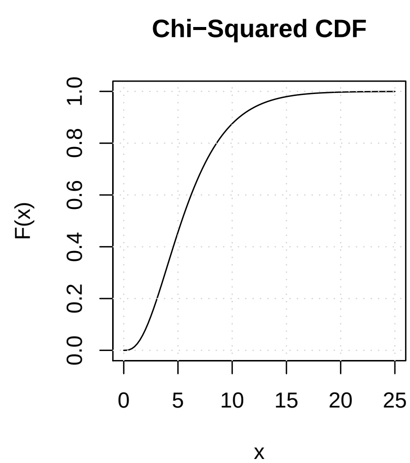
For continuous variables in general:

- Carefully determine location of area: to left, to right, interval.
- Always make a sketch when doing problems.
- CDF assumes areas to the left. Take the compliment when finding upper tail!

1.5 Additional Examples

Given the following density function on the left and it's corresponding CDF for the χ^2 distribution, answer the following questions.





Question 3. Find $P(x > 10)$

Question 4. Find P_{25}

The SAT-I scores for females is normally distributes with a mean of 998 and a standard deviation of 202 (based on data from the college board).

Question 5. If a female is randomly selected, what is the probability that her score is greater than 1100?

Question 6. What would the score be for P_{75}?

Question 7. What proportion of students scored between 500-1100?

Replacement times for CD players are normally distributed with a mean of 7.1 years and a standard deviation of 1.4 years.

Question 8. Find the probability that a randomly selected CD player will have a replacement time less than 8 years.

Question 9. If you want to provide a warranty so that on only 2% of the CD players will be replaced before the warranty expires, what is the time length of the warranty?