$P(X \in A)$ is a probability function

Since $P(X \in A)$ is a probability function, the following properties are immediate where A and B are any subsets of $\mathbb{R}$. Either A or B may have elements that are not in R.

1. $P(X \in A) = 1 - P(X \in A^C)$. For example
   $P(X \leq x) = 1 - P(X > x)$.
2. If A and B are disjoint then
   $P(X \in A \cup B) = P(X \in A \text{ or } X \in B) = P(X \in A) + P(X \in B)$.
3. In general,
   $P(X \in A \cup B) = P(X \in A) + P(X \in B) - P(X \in A \cap B)$.
4. If $A \subset B$ then $P(X \in A) \leq P(X \in B)$.

2.2 Continuous Random Variables

Some experiments result in an outcome X that occur in a continuum of values $R \subset \mathbb{R}$ rather than a finite or countable number of outcomes. For example, subject lifetime, an ELISA score, or birthweight might all be measured on a continuous scale.

def'n: The random variable X is continuous if there exists a $f(x) \geq 0$, called the **probability density function** (pdf), or simply **density** of X, such that for any interval (a, b),

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

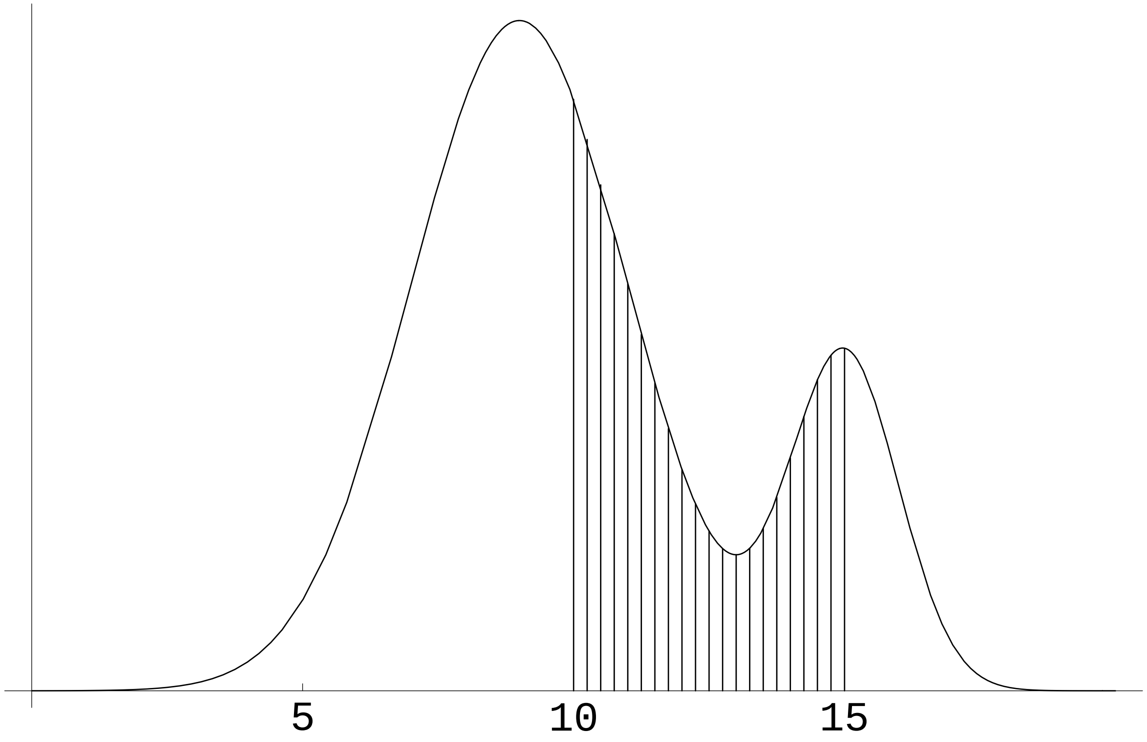

Figure 1: Probability density function $f(x)$ for one particular random variable X. Shaded region between 10 and 15 has area $\int_{10}^{15} f(x)dx = 0.3458$. That is, $P(10 \le X < 15) = 0.3458$.

- A density $f(x)$ for a continuous X is analogous to a pmf $p(x)$ for a discrete X. They both completely determine X in that $P(A)$ can be determined for any $A \subset \mathbb{R}$.
- Like a $p(x)$, a density $f(x)$ determines how probability is "spread out" over $\mathbb{R} = (-\infty, \infty)$.
- For continuous X and any fixed $a \in \mathbb{R}$,

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x)dx = 0.$$

- This implies that

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b).$$

- For Δ small, $P(a \leq X \leq a + \Delta) \approx f(a)\Delta$.
- (FTC): If $F'(x) = f(x)$, then $\int_a^b f(x)dx = F(b) - F(a)$.

A pdf $f(x)$ satisfies two properties:

1. $f(x) \geq 0$ for all $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

The first eliminates the possibility of negative probabilities, the second ensures $P(-\infty < X < \infty) = 1$. These also correspond to axioms 1 and 2 in the *definition* of probability. Axiom 3 comes out of properties of integrals.

Question: let $f(x) = kx^3$ for some $k > 0$ over $x \in R = (0, 1)$ and $f(x) = 0$ elsewhere. Can $f(x)$ be a density? If so, when? k is called a *normalizing constant.*

def'n: The **range** of a continuous random variable X is $R = \{x \in \mathbb{R} : f(x) > 0\}$.

Note that if $A \subset R^C$ then $P(X \in A) = 0$.

---

def'n: The cumulative distribution function (cdf) for *any* random variable X (discrete, continuous, or other) is defined to be

$$F_X(x) = P(X \leq x).$$

When only one random variable is being discussed, the subscript X may be dropped and the cdf written $F(x) = P(X \leq x)$.

Important: For continuous X with a density $f(x)$, the pdf is the derivative of the cdf: $f(x) = F'(x)$.

Question: what does $F(x)$ tend to when $x \to -\infty$? How about $x \to \infty$?

Using the cdf to find probabilities for continuous X

For X continuous with cdf $F(x)$,

$$P(a < X < b) = F(b) - F(a).$$

Note also for continuous random variables, including the endpoints or not doesn't matter:

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b).$$

The next three slides illustrate $F(a) = P(X \leq b)$, $F(b) = P(X \leq b)$ and $F(b) - F(a) = P(a < X < b)$.

(These are for $X \sim N(0, 1)$, $a = -1$ and $b = 0.5$.)

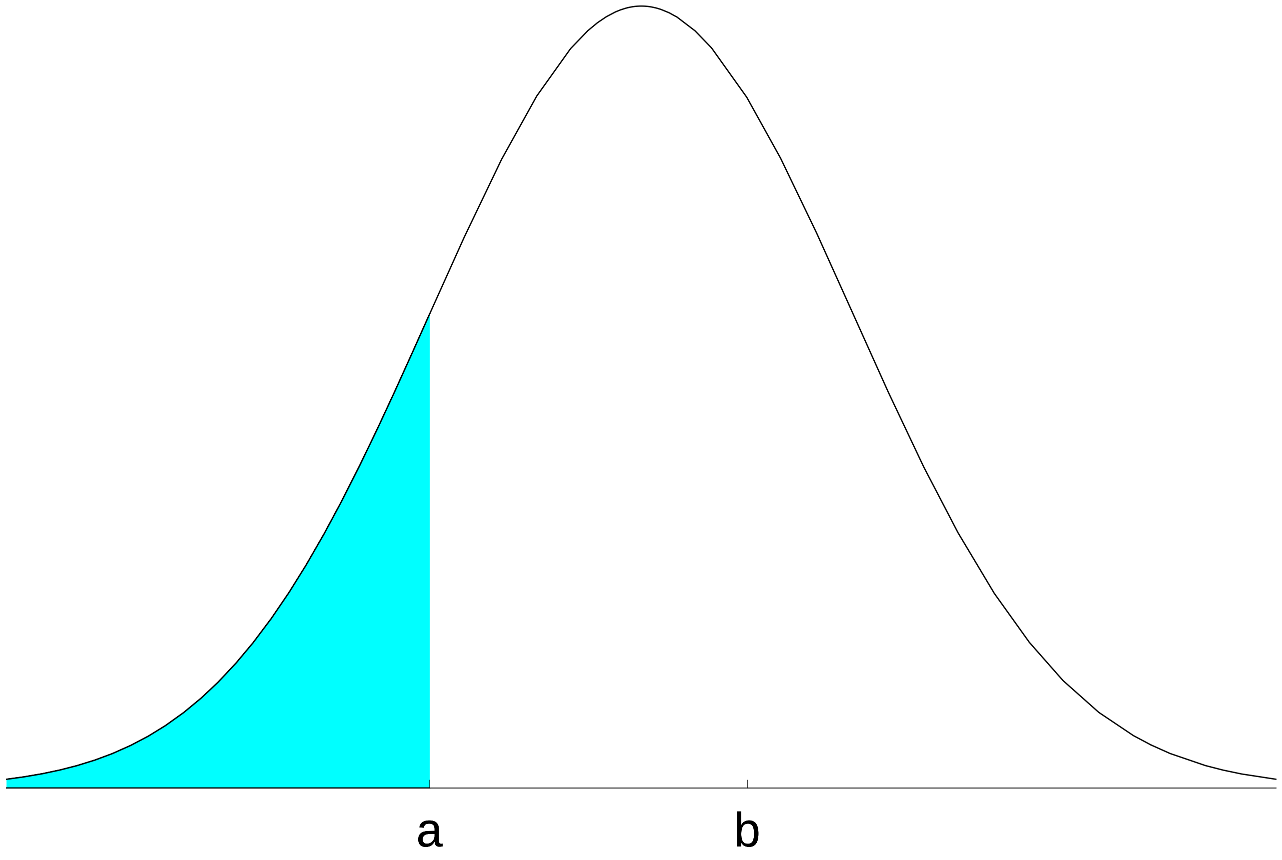

Figure 2: $F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$.

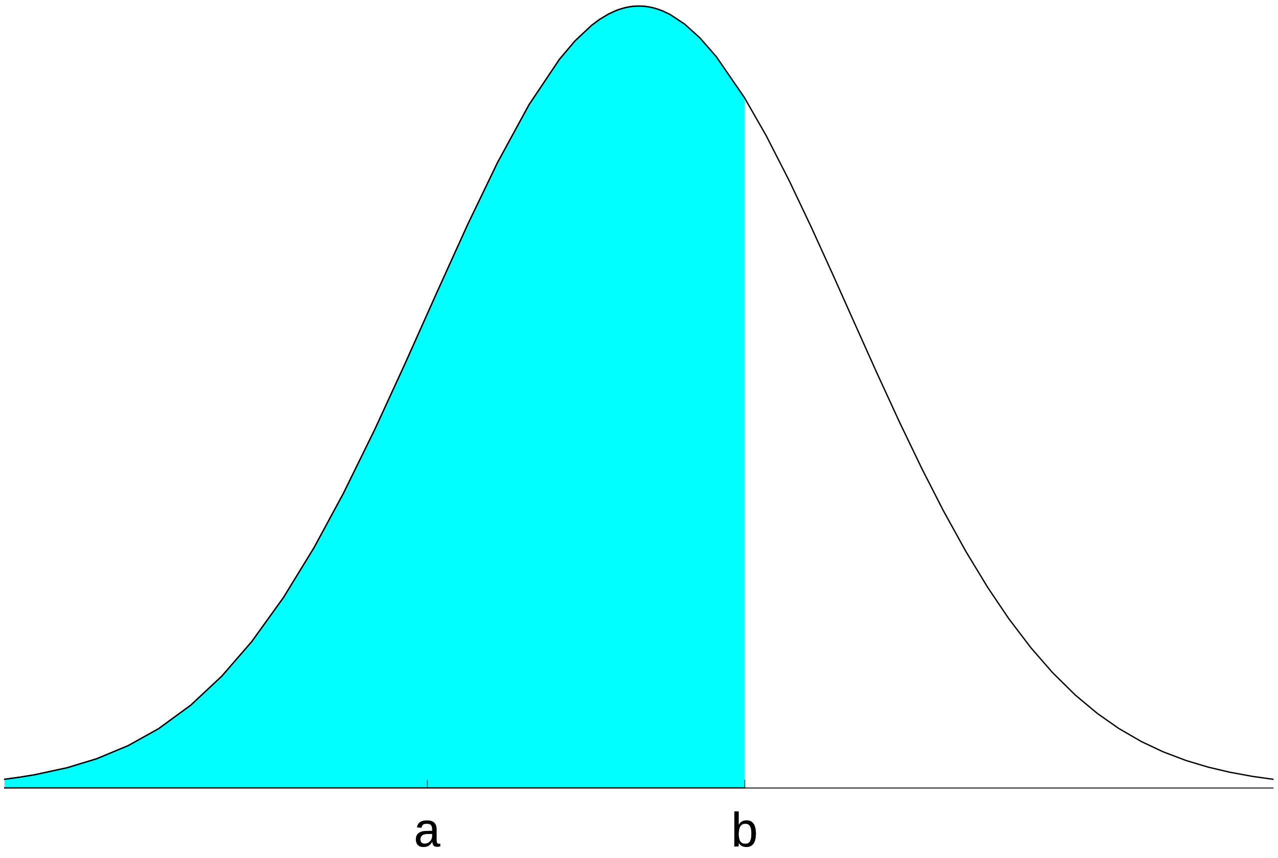

Figure 3: $F(b) = P(X \le b) = \int_{-\infty}^{b} f(x)dx$.

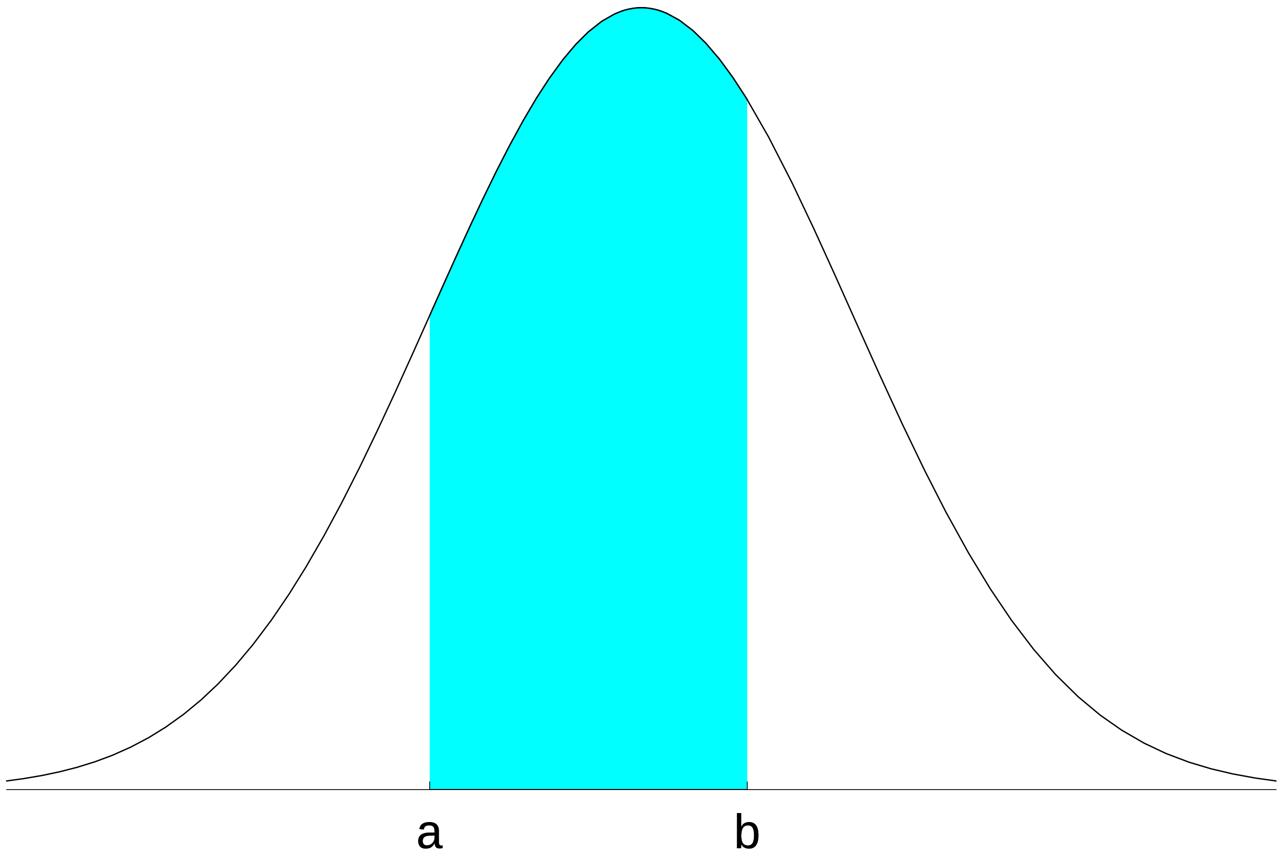

Figure 4: $F(b) - F(a) = P(a < X < b) = \int_a^b f(x)dx$. Think of subtracting the shaded part of Figure 2 from Figure 3.

Using the cdf to find probabilities for discrete X

(next two slides are optional & for your interest only...)

The cdf can be used to obtain probabilities of intervals for discrete random variables too. The following hold for *any* random variable:

$$P(X \leq x) = F(x), \qquad P(X < x) = F(x-),$$

where $F(x-) = \lim_{s \to x^-} F(s)$. For continuous random variables $F(x-) = F(x)$. These imply that

$$P(X > x) = 1 - F(x), \qquad P(X \geq x) = 1 - F(x-).$$

These can be used to compute, for example, $P(a < X \leq b)$, by noting that $(-\infty, a] \cup (a, b] = (-\infty, b]$ where the first two sets are disjoint and so

$$P(X \in (-\infty, a]) + P(X \in (a, b]) = P(X \in (-\infty, b]).$$

We can rewrite this more simply as

$$P(X < a) + P(a < X \leq b) = P(X \leq b),$$

which in terms of the cdf of X is

$$F(a-) + P(a < X \leq b) = F(b),$$

yielding

$$P(a < X \leq b) = F(b) - F(a-).$$

For continuous X this simplifies to

$$P(a < X \leq b) = F(b) - F(a).$$

For discrete random variables it's probably easier to use the pmf $p(x)$ to find probabilities.

Quantiles

The p^{th} **quantile** of a random variable, denoted x_p, divides up the real numbers $\mathbb{R}$ into two pieces, one with probability at least p and the other with probability at least $1-p$.

For continuous random variables that spread out all probability mass without gaps of “no probability” over an interval $R = (r_1, r_2)$, a quantile x_p divides up $\mathbb{R}$ into two unique disjoint intervals with probabilities p (left interval) and $1-p$ (right interval).

No gaps over a range $R = (r_1, r_2)$ amounts to requiring $F(x)$ to strictly increase over R. All four special continuous random variables in Sections 2.2.1 through 2.2.4 have ranges that are intervals on which $f(x) > 0$.

def'n: For a continuous random variable X such that $f(x) > 0$ over a range that's an interval $R = (r_1, r_2)$, a quantile x_p associated with probability p satisfies

$$P(X \leq x_p) = p,$$

i.e. $x_p = F^{-1}(p)$.

def'n: The **median** of X is $x_{0.5}$ such that $P(X \leq x_{0.5}) = 0.5$, i.e. $x_{0.5} = F^{-1}(0.5)$. The **quartiles** are $x_{0.25}$, $x_{0.5}$, and $x_{0.75}$.

The median $x_{0.5}$ chops $\mathbb{R}$ (also R) into two pieces, each with equal probability of occurring. The quartiles chop R into four pieces.

When $f(x) = 0$ for parts of R, or for discrete random variables, a quantile may not be unique and requires some careful thought...

Finding quantiles in practice:

- For a discrete random variable, find x_p such that

$$P(X \leq x_p) \geq p \text{ and } P(X \geq x_p) \geq 1 - p.$$

- For a continuous random variable, find x_p that solves

$$p = P(X \leq x_p) = F(x_p).$$

Modes:

Modes m are values that are "most likely" to happen.

- For discrete X,

$$m = \text{argmax}_{x \in \mathbb{R}}\, p(x).$$

- For continuous X

$$m = \text{argmax}_{x \in \mathbb{R}}\, f(x).$$

Uniform distribution

def'n: $X \sim U(a,b)$ if the pdf of X is $f(x) = \frac{1}{b-a}$ over the range $R = (a,b)$.

Note that we require $b > a$ for this to make sense.

The uniform random variable is useful for modeling outcomes that are equally likely across an interval.

Example: if the bus can come any time in the next five minutes and all arrival times are equally likely, then $X \sim U(0,5)$ where X is the waiting time in minutes.

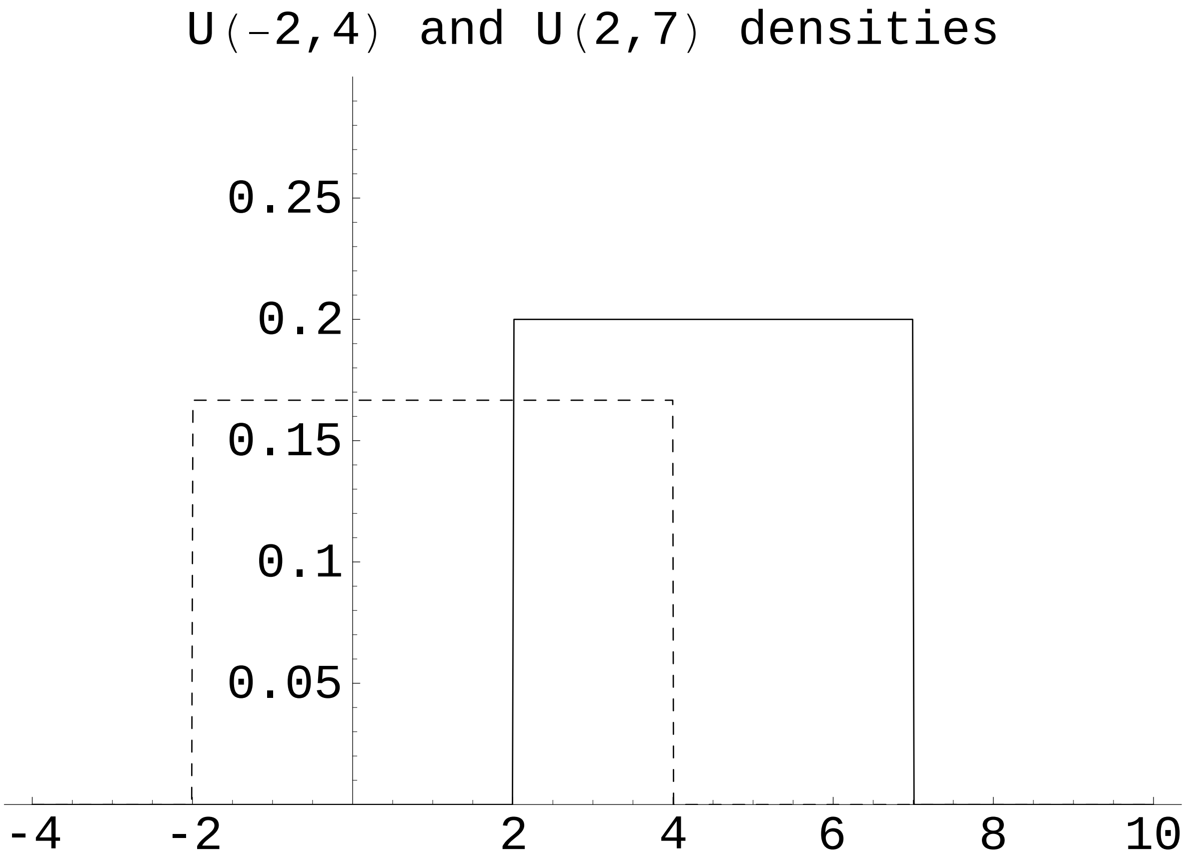

Figure 5: Uniform densities; you should match them.

The pdf of a $U(a,b)$ random variable is constant k across the interval (a,b) and zero elsewhere. Since we require $\int_{-\infty}^{\infty} f(x)dx = 1$, this amounts to $\int_a^b kdx = 1$ or $k \times (b-a) = 1$ implying $k = \frac{1}{b-a}$.

The cdf of a $U(a,b)$ random variable is derived by considering three possibilities (1) $x \leq a$, (2) $a < X < b$, and (3) $X > b$. In the three cases the probability $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)ds$ can be written as the sum of one, two, or three pieces respectively. We'll explore this in detail in class at the blackboard.

We find

$$F(x) = \left\{ \begin{array}{ll} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{array} \right\}.$$

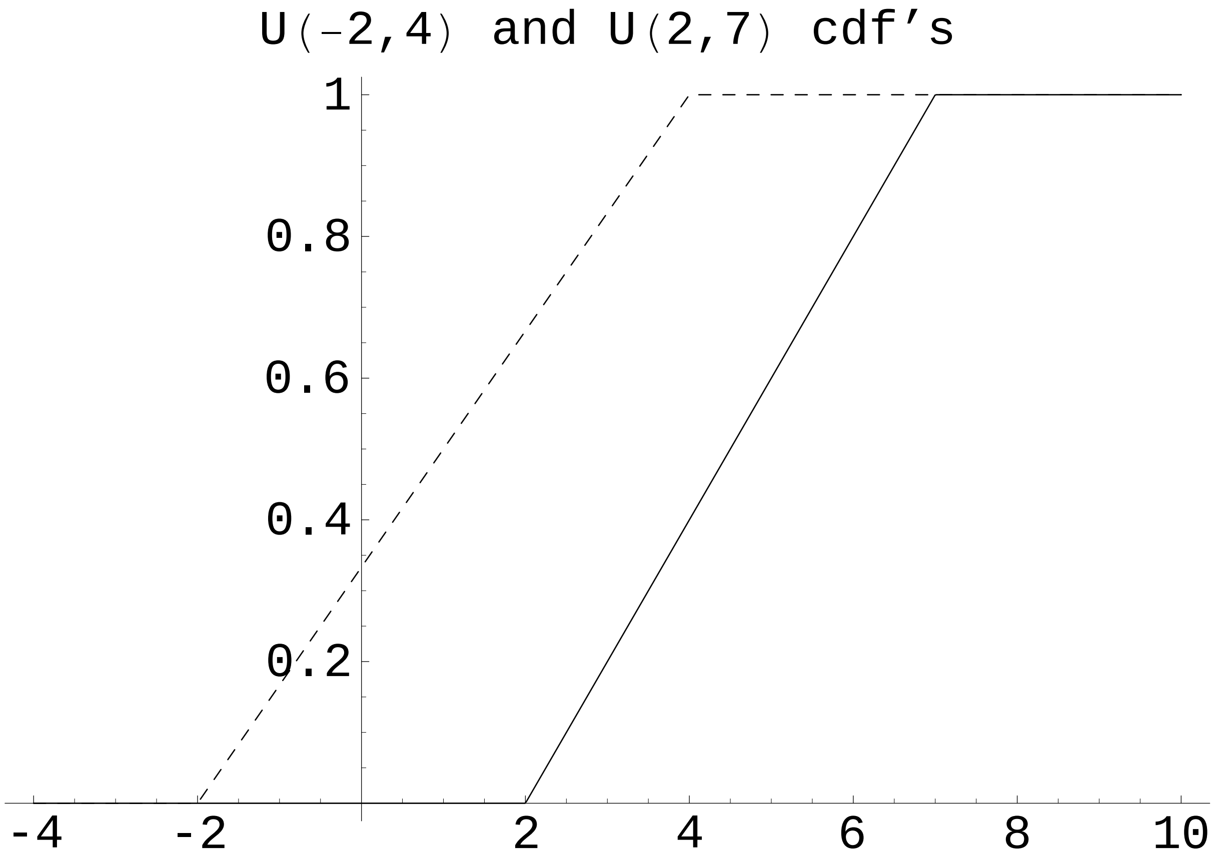

Figure 6: Uniform cdf's.

2.2.1 Exponential distribution

def'n: $X \sim \exp(\lambda)$ if X has pdf $f(x) = \lambda e^{-\lambda x}$ on $R = [0, \infty)$.

This is equivalent to writing

$$f(x) = \left\{ \begin{array}{ll} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{array} \right\}.$$

Typically used to model time to some event, e.g. waiting time until next customer at service window. Has ties to Poisson process.

Note: just as $0 \leq p \leq 1$ for Bernoulli, geometric, and binomial random variables, and $\lambda > 0$ for Poisson random variables, we need $\lambda > 0$ in an exponential random variable to get a valid probability distribution.

It is the goal of *statistical* inference to estimate p or λ in these distributions given *data*. For now we assume they're known.

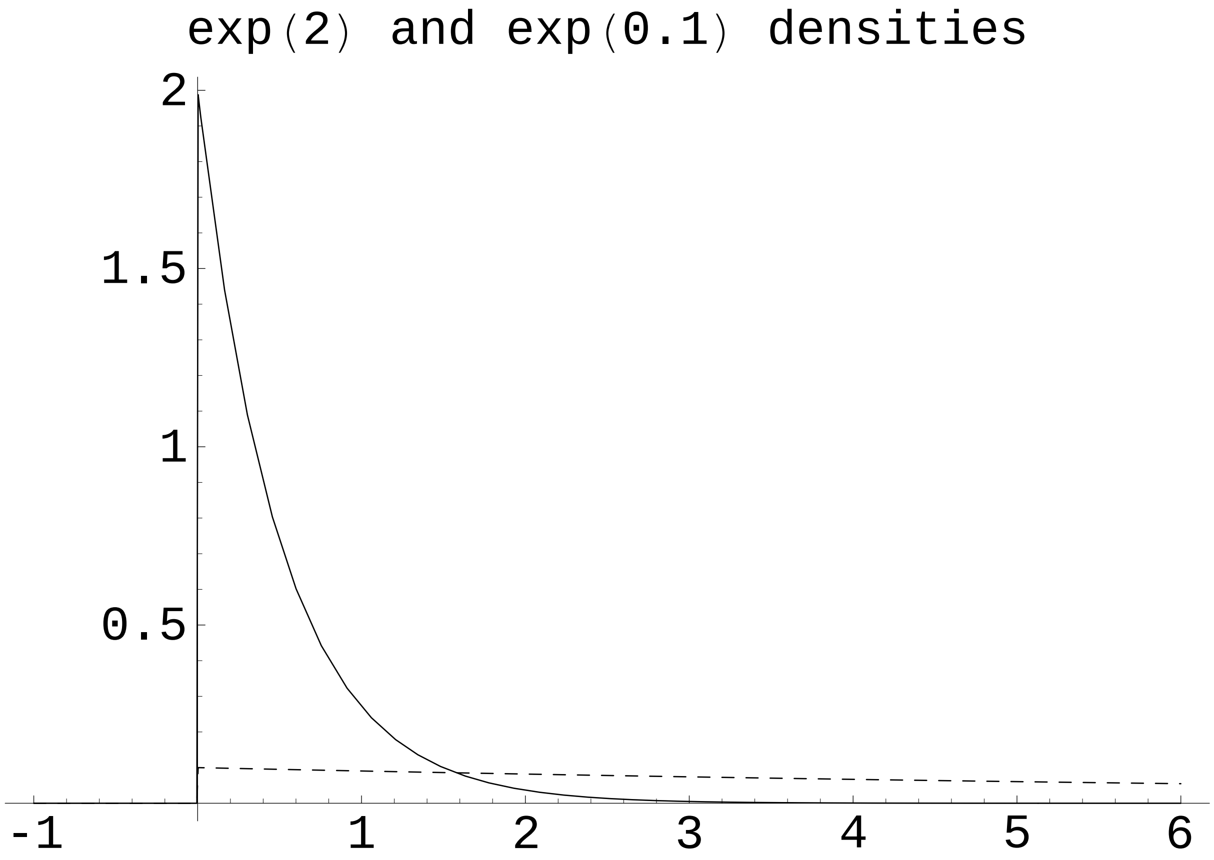

Figure 7: Exponential densities.

Note that for any $x < 0$,

$$P(X \leq x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} 0dx = 0.$$

For any $x \geq 0$,

$$\begin{aligned} P(X \leq x) &= P(X \leq 0) + P(0 < X \leq x) = 0 + \int_{0}^{x} f(x)dx \\ &= \int_{0}^{x} \lambda e^{-\lambda s} ds = \int_{0}^{x} [-e^{-\lambda s}]' ds = [-e^{-\lambda s}]_{0}^{x} \\ &= -e^{-\lambda x} - -e^{-\lambda 0} = 1 - e^{-\lambda x}. \end{aligned}$$

So

$$F(x) = \left\{ \begin{array}{ll} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{array} \right\}.$$

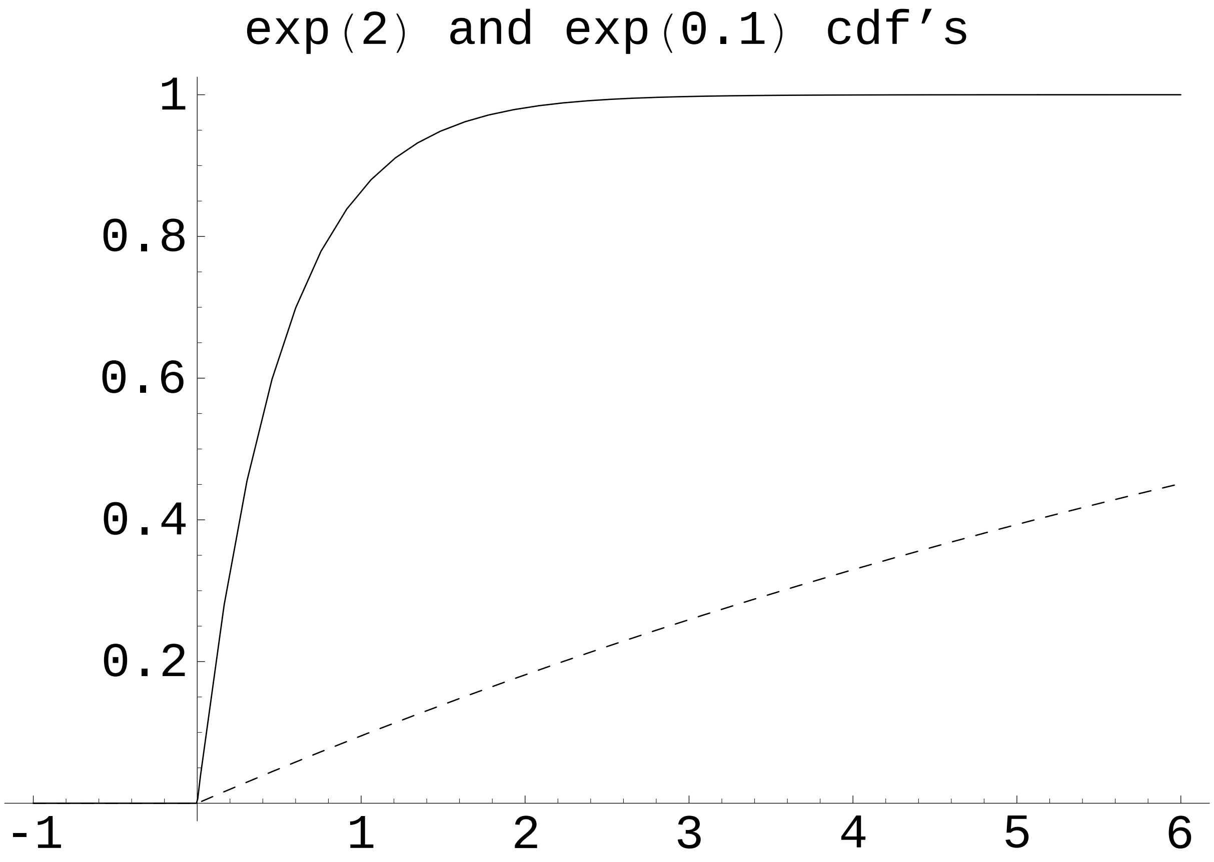

Figure 8: Exponential cdf's.

We could also just write $F(x) = 1 - e^{-\lambda x}$ on $R = [0, \infty)$ and call it a day. We've defined $F(x)$ on those values of x in the range. $F(x) = 0$ for those x that are smaller than all values in R and $F(x) = 1$ for those x that are larger than all values in R.

To find the inverse of $F(x)$ we set $p = F(x)$ and solve for x:

$$\begin{aligned} p = 1 - e^{-\lambda x} \quad &\Leftrightarrow \quad e^{-\lambda x} = 1 - p \\ &\Leftrightarrow \quad -\lambda x = \log(1-p) \\ &\Leftrightarrow \quad x = -\frac{\log(1-p)}{\lambda}. \end{aligned}$$

So $F^{-1}(p) = -\frac{\log(1-p)}{\lambda}$ for $0 < p < 1$. Note that the (statistical) range $R = [0, \infty)$ is also the (functional) range of $F^{-1}(p)$, i.e. the set of all possible values X can take on. We have $x_{0.5} = F^{-1}(0.5) = -\log(0.5)/\lambda$.

Example: a light bulb has a lifetime given by $X \sim \exp(1/1000)$ hours. What is the probability that the light bulb will last longer than 1000 hours? 2000 hours? Between 1000 and 1010 hours? What is the median lifetime of X?

$$P(X > 1000) = 1 - F(1000) = 1 - (1 - 0.001e^{-0.001\times1000}) \approx 0.368.$$

$$P(X > 2000) = 1 - F(2000) = 1 - (1 - 0.001e^{-0.001\times2000}) \approx 0.135.$$

$$P(1000 < X < 1010) = F(1010) - F(1000) = 0.00366,$$

or approximately

$$P(1000 < X < 1010) \approx f(1000) \times 10 = 0.00368,$$

where $\Delta = 1010 - 1000 = 10$.

The median lifetime is $x_{0.5} = -\log(0.5)/0.001 = 693.1$ hours.

2.2.2 Gamma distribution

def'n: $X \sim \text{gamma}(\alpha, \lambda)$ if X has pdf $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$ over the range $R = [0, \infty)$.

As in the exponential, the range is $[0, \infty)$, suitable for modeling times, such as lifetimes. In fact, the exponential random variable is obtained from the gamma when $\alpha = 1$, i.e. $X \sim \exp(\lambda)$.

The function $\Gamma(\alpha)$ is called the gamma function and does not have a nice, closed form. Values of the function for a given x are computed from its definition:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \qquad x > 0.$$

Useful fact: if n is an integer then $\Gamma(n) = (n-1)!$.

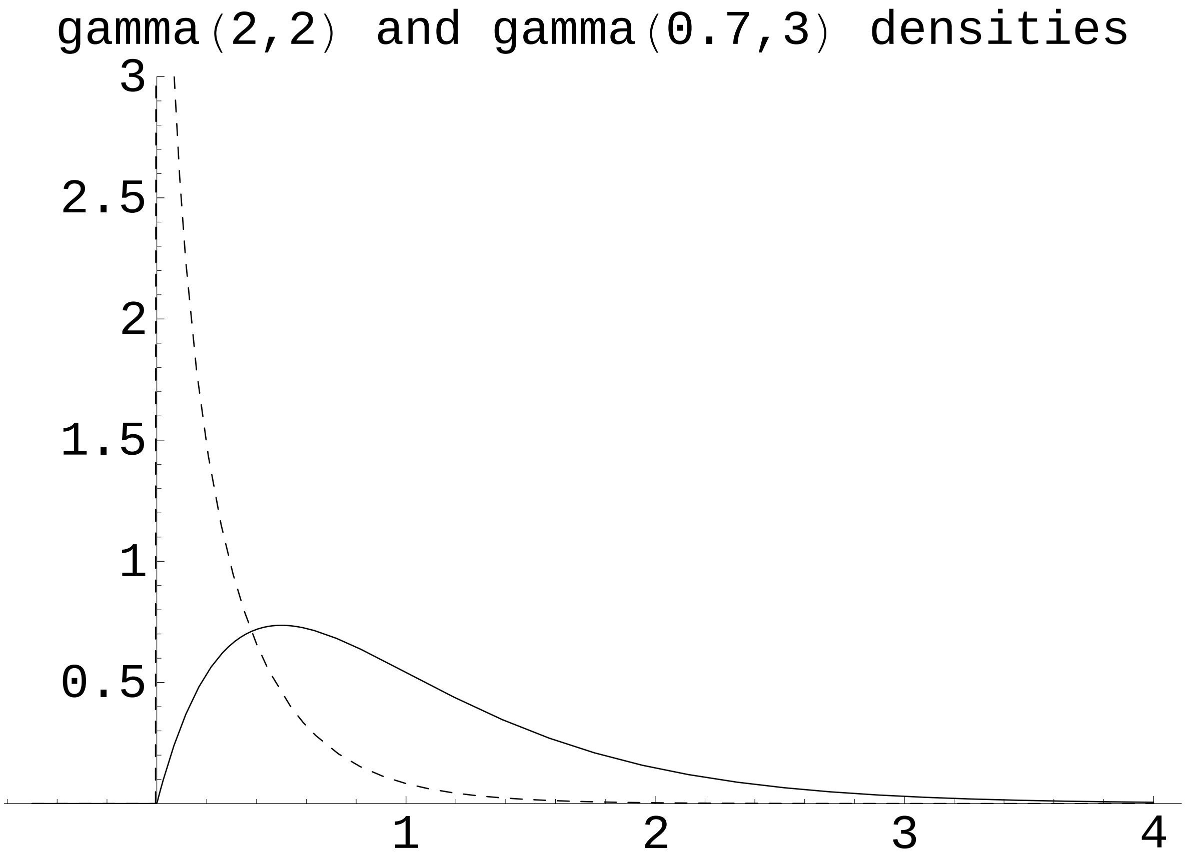

Figure 9: Gamma densities.

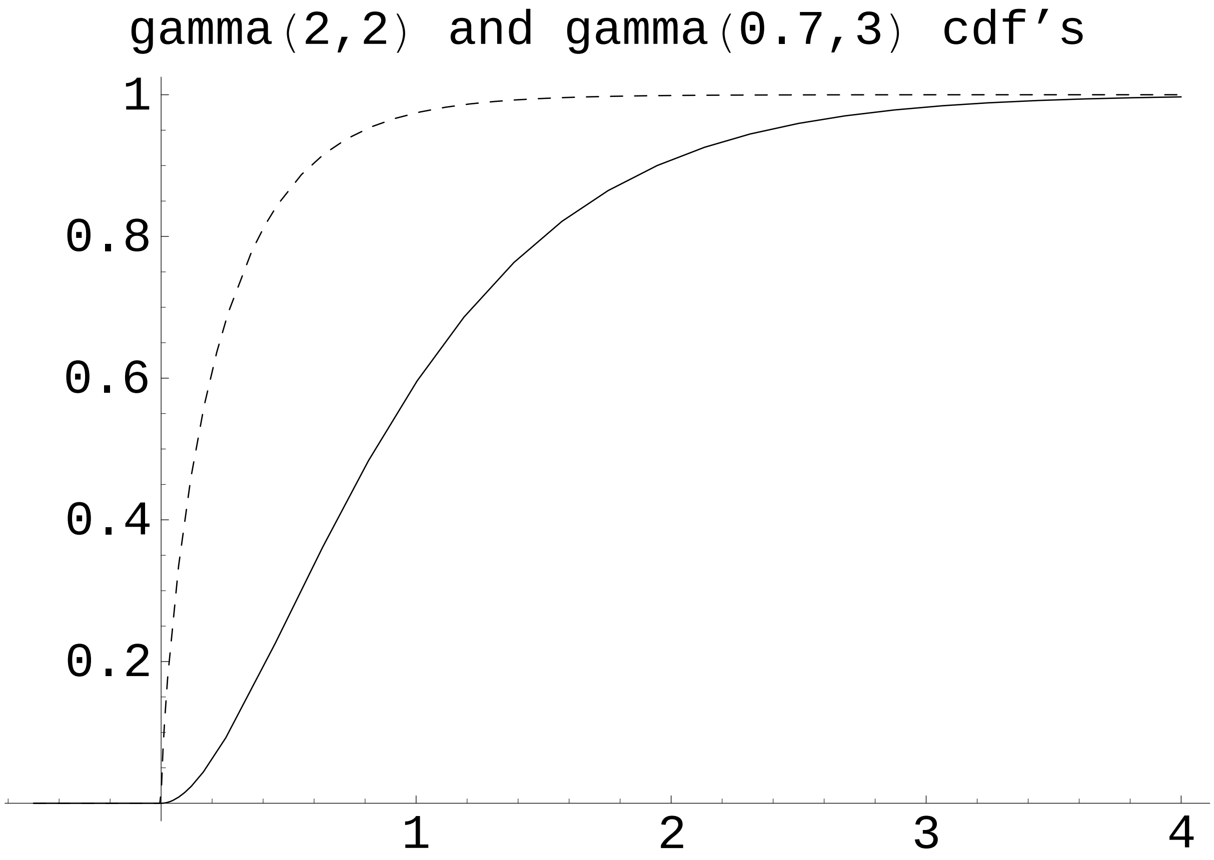

Figure 10: Gamma cdf's.

The gamma density does not have a nice closed form for the cdf $F(x) = P(X \leq x)$ like the exponential does. So it is not easy to get quantiles $x_p = F^{-1}(x)$. Quantiles and probabilities such as $P(a \leq X \leq b)$ are obtained numerically.

We require both $\alpha > 0$ and $\lambda > 0$ to have a valid probability distribution defined for X.

Example A on pp. 53-54 shows that a fitted gamma density is better than the simpler exponential density for modeling the waiting times between earthquakes. This implies that time between earthquakes is not memoryless. See text.

Not in text: Weibull distribution

def'n: $X \sim \text{Weib}(\alpha, \lambda)$ if X has pdf $f(x) = \frac{\alpha}{\lambda}\left(\frac{x}{\lambda}\right)^{\alpha-1} e^{-\left(\frac{x}{\lambda}\right)^{\alpha}}$ on $R = [0, \infty)$.

The Weibull density has shapes similar to the gamma, but *does* have a nice form for the cdf, $F(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^{\alpha}}$. So quantiles can be computed as $x_p = F^{-1}(p)$, and probabilities of X being in intervals can be calculated easily.

Note that like the gamma, an exponential random variable is obtained when $\alpha = 1$, $X \sim \exp(1/\lambda)$.

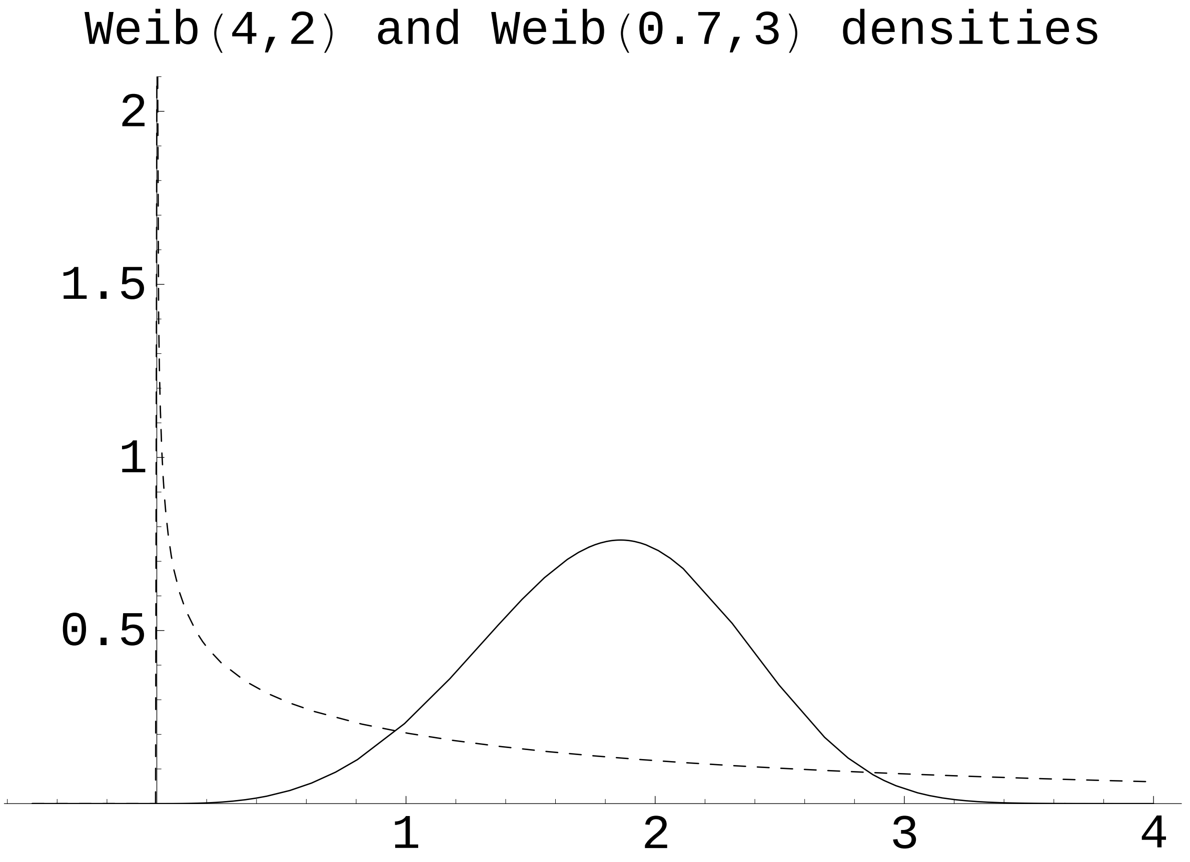

Figure 11: Weibull densities.

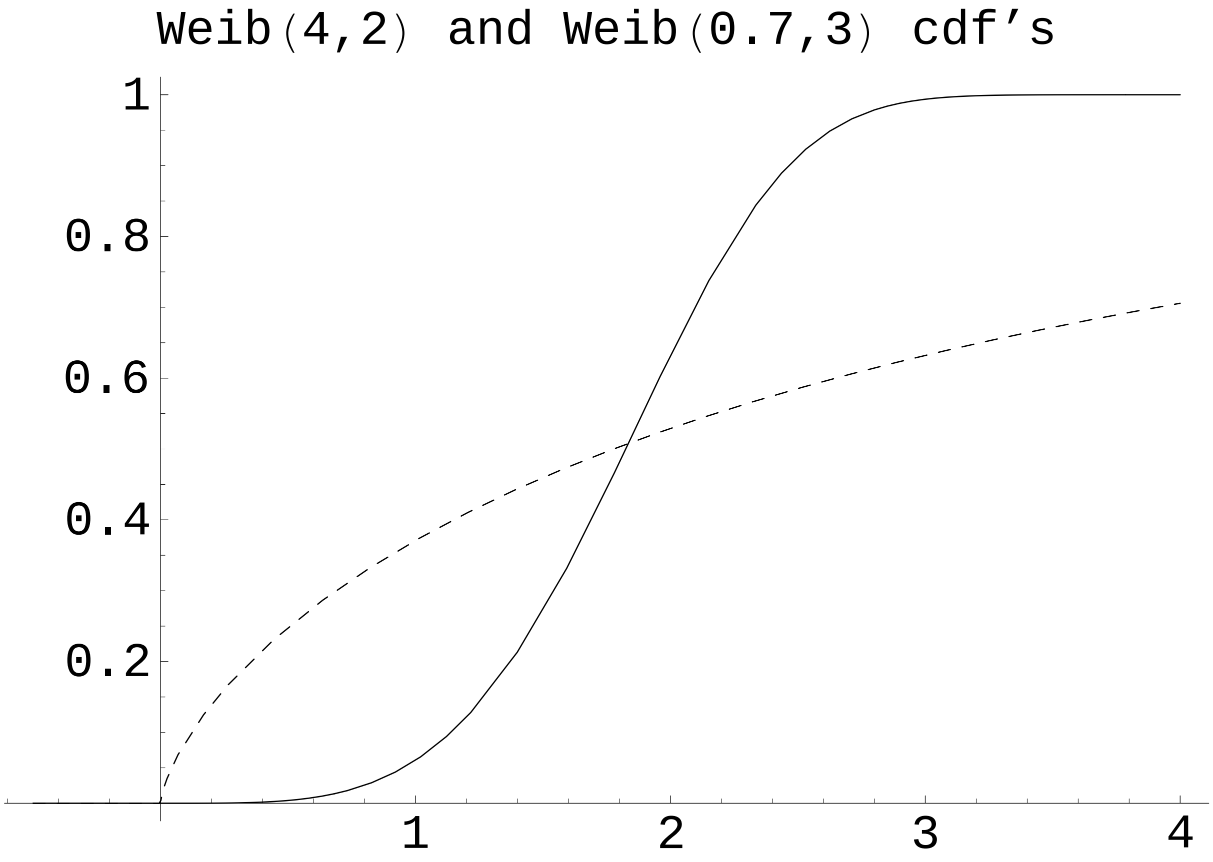

Figure 12: Weibull cdf's.

2.2.4 beta distributions

def'n: $X \sim \text{beta}(\alpha, \beta)$ if X has the pdf
$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ over the range $R = (0, 1)$.

Since the possible values for X are between zero and 1, i.e. $P(0 \leq X \leq 1) = 1$ for a beta random variable, the beta distribution is useful for modeling probabilities.

Beta random variables do not have nice closed forms for cdf's and hence for quantiles. They are obtained numerically except for in certain simple cases.

Note that if $X \sim \text{beta}(1, 1)$ then $X \sim U(0, 1)$.

Exercise: find the cdf of the random variable $X \sim \text{beta}(3, 2)$. Is it easy to find the quantile function $F^{-1}(p)$?

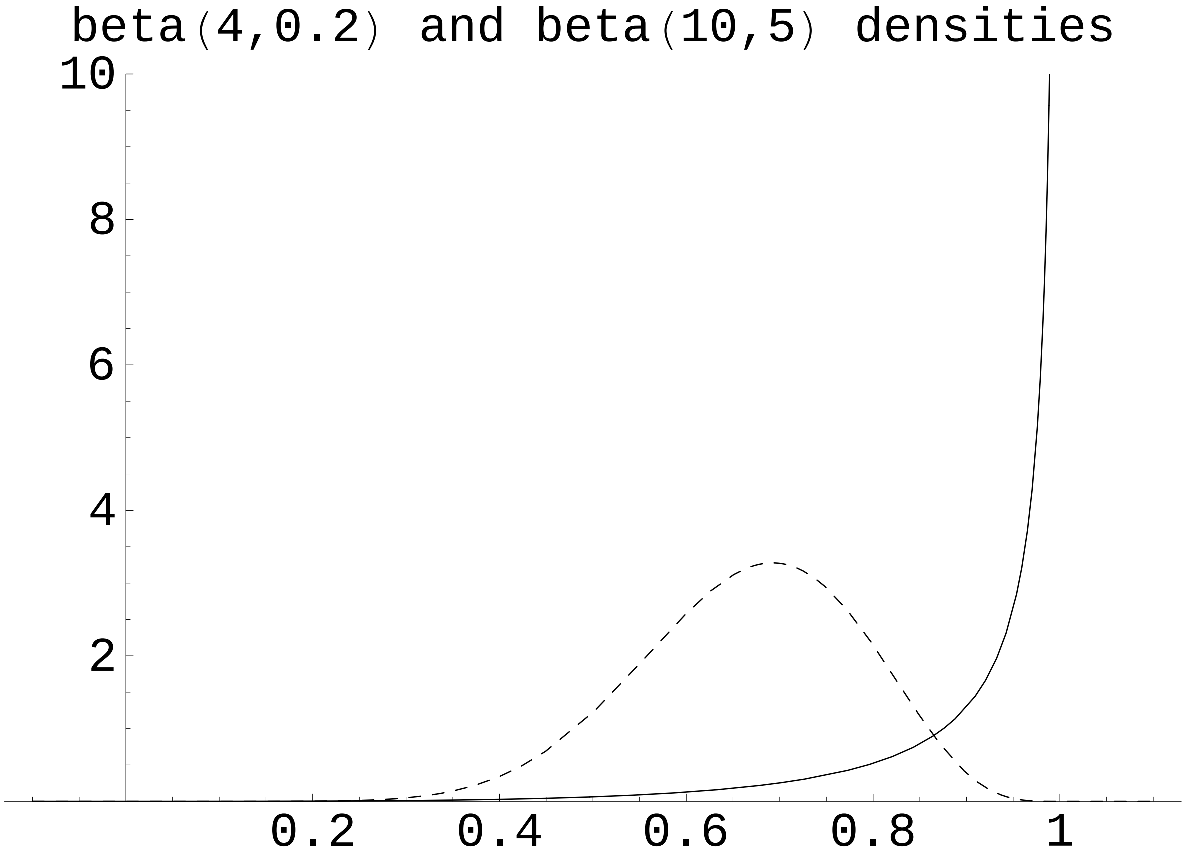

Figure 13: beta densities.

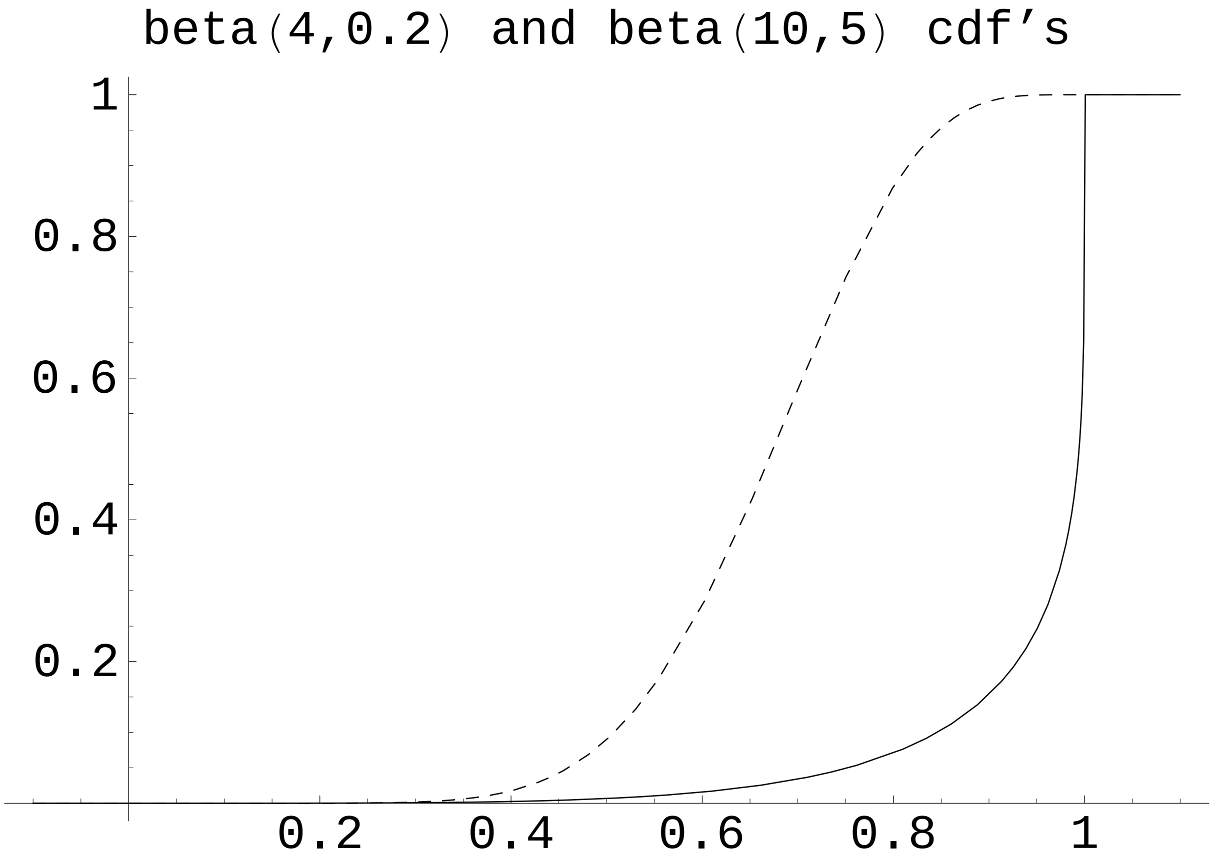

Figure 14: beta cdf's.

Homework, Chapter 2: 1, 2*, 5, 7, 33 (this is a Weibull in disguise!), 34*, 35, 40* (what kind of random variable is this?), 41 (i.e. find $x_{0.25} = F^{-1}(0.25)$ and $x_{0.75} = F^{-1}(0.75)$), 45, 46*, 47 (hint: use the first derivative)

*hand in.

Next time: calculus review. Then: normal distributions, functions of random variables, review, and problems.