Continuous random variables

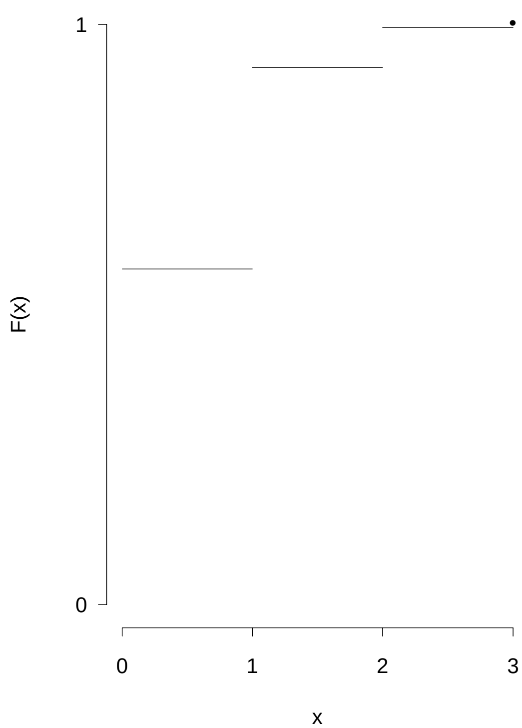
- Can take on an uncountably infinite number of values
- Any value within an interval over which the variable is definied has some probability of occuring
- This is different from the discrete case, in which every point in a given interval may not be a possible value

Cumulative dist. functions: continuous case

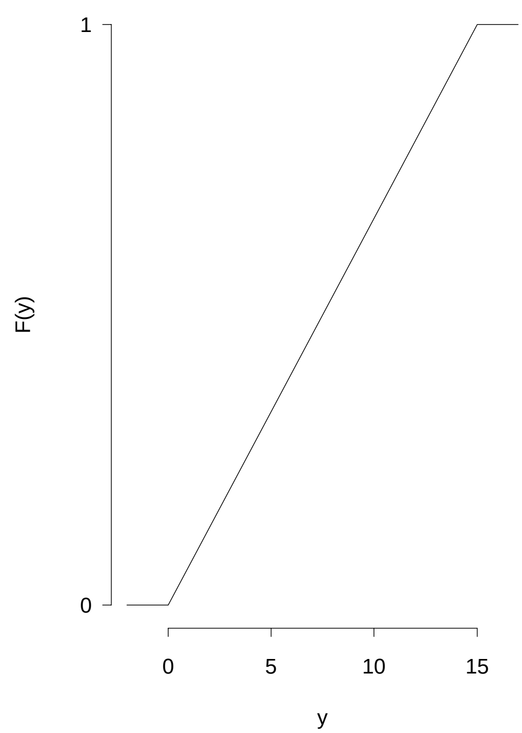
$$F(x) \quad = \quad P(X \leq x)$$

- In continuous case, cumulative distribution function (a.k.a. distribution function) doesn't look like a step function
- This is because in each interval, there are points of positive probability contributing to F(x)
- The c.d.f is a continuous, nondecreasing function
- $F(-\infty) = 0$, $F(\infty) = 1$

Discrete vs. cont. example



Binomial c.d.f with n=3, p=1/6



Uniform c.d.f with a=0 and b=15

Probability Density Functions

- Probability density function (p.d.f) for X is derivative of the c.d.f: $f(x) = \frac{dF(x)}{dx} = F'(x)$
- It follows that the c.d.f can be written: $\int_{-\infty}^{x} f(t)dt$
- $f(x)$ must be continuous (exception: see note on p. 139)
- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

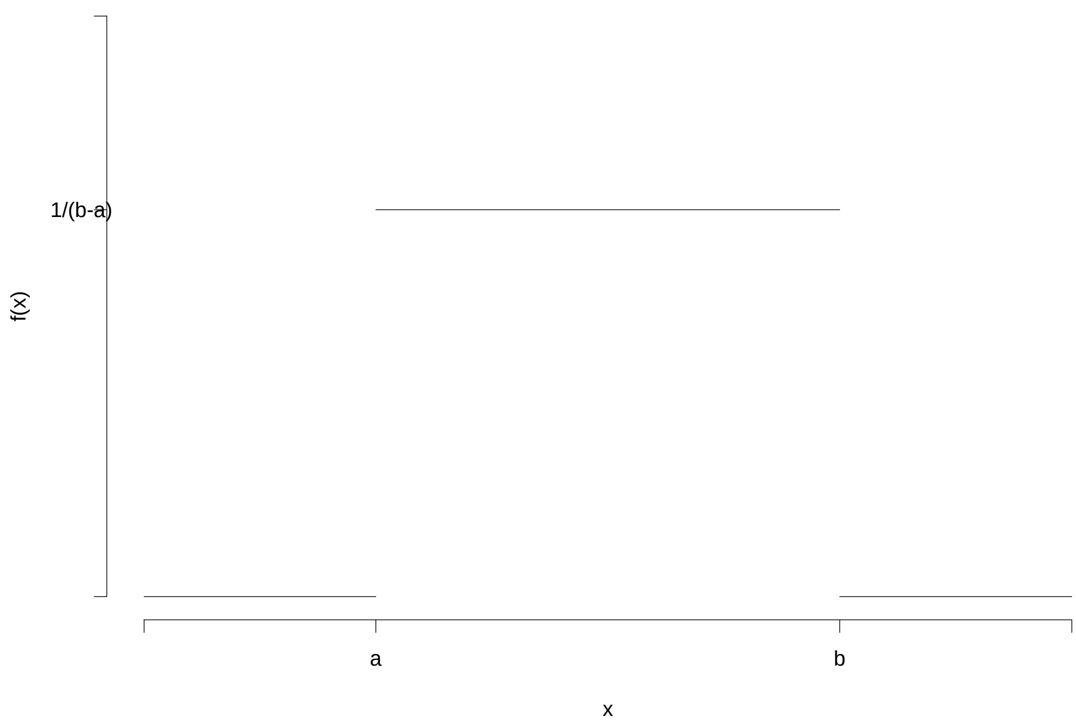
Expected value for continuous random variables

- The continuous case uses integrals instead of sums (as in discrete case)
- Continuous X: $E(X) = \int_{-\infty}^{\infty} xf(x)dx$
- Discrete X: $E(X) = \sum xf(x)$
- Likewise for the expected value of a function of x: $E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx$

Uniform distribution

If X is distributed uniformly on the interval (a, b), then X has p.d.f:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if} \quad a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$



Uniform p.d.f

Mean of the uniform distribution

If X is distributed uniformly on the interval (a, b), $E(X) = \frac{a+b}{2}$.

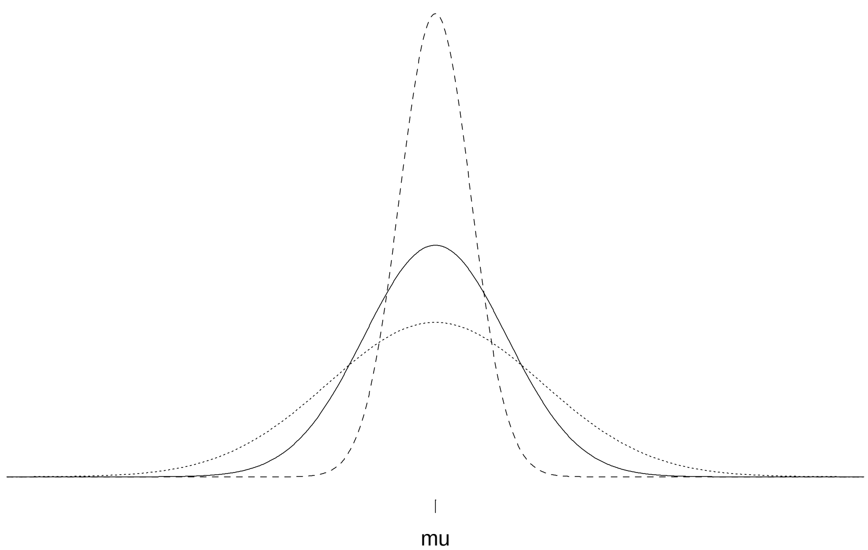
Show this using the definition we have for the expected value of a continuous random variable X: $E(X) = \int_{-\infty}^{\infty} x f(x) dx$

Example using the uniform

Let's say that Jenise doesn't get up as soon as her alarm goes off. The extra time she sleeps in is given by the random variable X, which is distributed uniformly on the interval (0 min, 15 min).

- What's the probability that her extra sleeping time is less than 5 min?
- What's the probability that her extra sleeping time is less than 10 min, but more than 7 min?

Normal distribution



- Has p.d.f $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{\frac{-1}{2\sigma^2}(x-\mu)^2\}$, for $-\infty < x < \infty$
- Mean is given by μ, variance by $\sigma^2 > 0$
- Has the classic bell-curve shape
- Forms the basis of the empirical rule
- Used to approximate many real-life processes

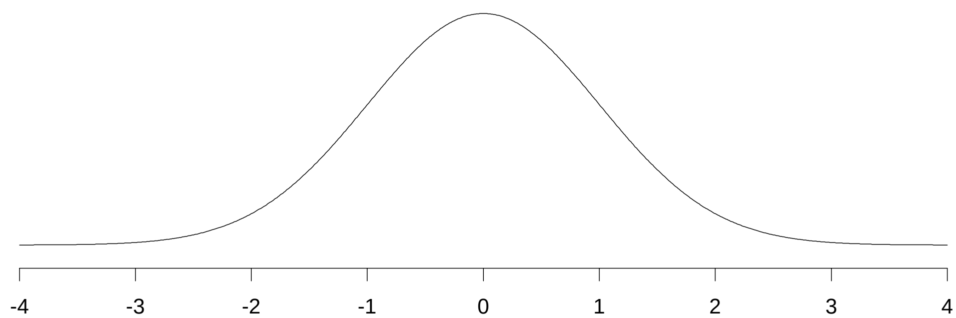
Areas under the normal curve

- To find $P(a \leq X \leq b)$, we need to evaluate $\int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\{\frac{-1}{2\sigma^2}(x-\mu)^2\}dx$
- But there is no closed-form solution to the integral!
- Numerical integration methods must be used in order to find $P(X \geq x)$ for various values of x
- How to do this, since there are infinite possibilities for μ and σ?

Areas under the normal curve

- Random variable Z has a standard normal distribution if it is distributed normally with $\mu = 0$ and $\sigma = 1$
- Values of Z correspond to how many standard deviations away from the mean they are
- All other normally distributed RVs can be transformed to the standard normal using this idea of "how many std. dev. from the mean is it"
- For $X \sim N(\mu, \sigma^2)$, we can transform X into standardized scores (z-scores) using $Z = \frac{X-\mu}{\sigma}$

Using z-scores and the normal table



Standard normal p.d.f

- Areas under the curve have been calculated and recorded in the normal table (see inside cover of textbook)
- For each z you look up in the table, you will get $P(Z \geq z)$
- Since the standard normal is symmetric around $\mu = 0$, there are no negative values for z on the table
- Symmetry is an important property to remember when using the table!!!

Using the std. normal table

Suppose the variable Z has a standard normal distribution, i.e. $Z \sim N(0, 1)$. Find the following probabilities:

- $P(-1 \le Z \le 1)$

- $P(Z \le -1.96)$

- $P(-0.50 \le Z \le 1.25)$

Using the std. normal table

Suppose the variable Z has a standard normal distribution, i.e. $Z \sim N(0, 1)$.

- Which value marks the 95th percentile?

- Which values are the boundaries for the middle 80% of the data?

Example using the normal

The time required to complete a college achievement test was found to be normally distributed, with a mean of 110 minutes and a standard deviation of 20 minutes.

- What percentage of students will finish within 2 hours?

- What percentage of students will finish after 1.5 hours but before 2.5 hours?

Example using the normal (cont.)

The time required to complete a college achievement test was found to be normally distributed, with a mean of 110 minutes and a standard deviation of 20 minutes.

- When should the test be terminated to allow just enough time for 90% of students to complete the test?

- What are the boundaries for the IQR of the time it takes to complete the test?

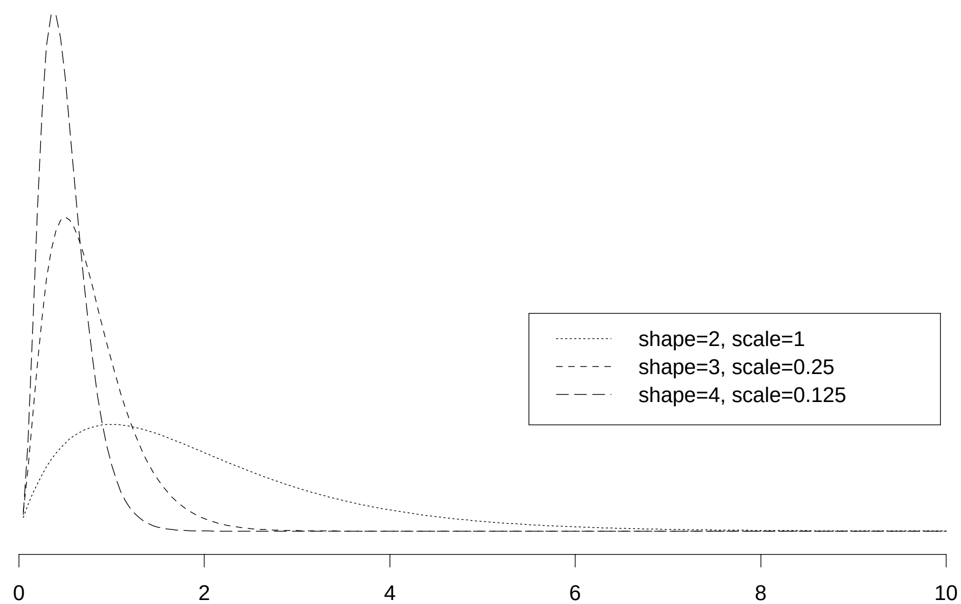
Gamma distribution

If X has a gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, then X has p.d.f.:

$$f(x) = \begin{cases} \frac{x^{\alpha-1}\exp(\frac{-y}{\beta})}{\beta^{\alpha}\Gamma(\alpha)} & 0 \le x \le \infty \\ 0 & \text{otherwise} \end{cases}$$

- $\Gamma(\alpha) = (\alpha - 1)!$ *only* in the case that α is an integer
- Only when α is an integer can we integrate this p.d.f. over an interval and get a closed-form expression
- $E(X) = \alpha\beta$
- Two special cases of the gamma have their own names
  1. An exponential with parameter β is a gamma with $\alpha = 1$ and β
  2. A chi-squared with ν degrees of freedom is a gamma with $\alpha = \frac{\nu}{2}$ and $\beta = 2$

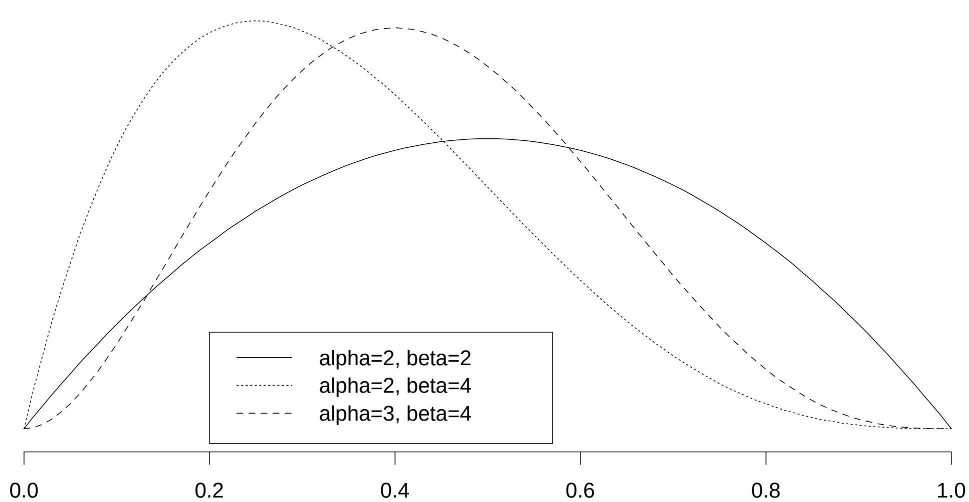
Various gamma p.d.f.s



Beta distribution

If X has a beta distribution with parameters $\alpha > 0$ and $\beta > 0$, then X has p.d.f.:

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



Ex: fueling station

A gas station operates 2 pumps, each of which can pump up to 10,000 gallons of gas in a month. The total amount of gas pumped at the station in a month is a random variable Y (measured in 10,000 gallons) with a probability density function given by

$$f(y) = \begin{cases} y, & 0 < y < 1 \\ 2 - y, & 1 \leq y < 2 \\ 0, & \text{elsewhere} \end{cases}$$

- Find $F(y)$.

Fueling station (cont.)

- Find $P(8 < Y < 12)$.

- Given that the station pumped more than 10,000 gallons in a particular month, find the probability that the station pumped more than 15,000 gallons during the month.