LECTURE: EXPECTED VALUES

1. Median and Quartiles

The **median** is the “middle” of a probability distribution. For example, in the distribution of students’ grades, half the grades are below the median and half are above. The median is more robust to outlier values than the mean.

If Y is a random variable, then the median of Y is the value m such that $P(Y \leq m) = P(Y \geq m) = 1/2$. In term of cdf, it’s the value m for which $F(m) = P(Y \leq m) = 1/2$

Definition:

Let Y be a continuous random variable with cdf $F(y)$

Then the **median** m, **first quartile** Q_1, and **third quartile** Q_3 are numbers such that

$$F(Q_1) = P(Y \leq Q_1) = 1/4$$
$$F(m) = P(Y \leq m) = 1/2$$
$$F(Q_3) = P(Y \leq Q_3) = 3/4$$

Example 1:

Find the median of the random variable Y whose density is

$$f(y) = \begin{cases} \frac{3}{8}y^2 & 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Let m be the median of Y

STEP 1: First let's find the cdf $F(y)$ of Y

For $0 \leq y \leq 2$, which is the region we care about, we have:

$$F(y) = \int_{-\infty}^{y} f(z)dz = \int_{0}^{y} \frac{3}{8}z^2 dz = \frac{3}{8}\left[\frac{z^3}{3}\right]_0^y = \frac{y^3}{8}$$

(If $y < 0$ then $F(y) = 0$ and if $y > 2$ then $F(y) = 1$)

STEP 2: To find the median, we solve $F(m) = 1/2$ for m

$$\frac{m^3}{8} = \frac{1}{2} \Rightarrow m^3 = 4 \Rightarrow m = 2^{2/3} \approx 1.59$$

2. Expectation and Variance

Just as with discrete random variables, we can talk about the expected value and variance of a continuous random variable. They have the exact same interpretations as in the discrete case.

Expectation and variance work almost exactly the same way in the continuous case as in the discrete case, provided we replace sums with integrals:

Definition:

If Y is a continuous random variable with density $f(y)$ then

$$E(Y) = \int_{\infty}^{\infty} y f(y) dy$$

Note: And if g is a real-valued function, then the expected value of $g(Y)$ is given by

$$E[g(Y)] = \int_{\infty}^{\infty} g(y) f(y) dy$$

The variance of a continuous random variable is defined the same way as in the discrete case.

Definition:

$$\text{Var}(Y) = E[(Y - \mu)^2] \qquad \mu = E(Y)$$

Magic Variance Formula:

$$\text{Var}(Y) = E(Y^2) - [E(Y)]^2$$

Example 2:

Find $E(Y)$ and $\text{Var}(Y)$ where Y has the pdf

$$f(y) = \begin{cases} \frac{3}{8}y^2 & 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy = \int_0^2 y \left(\frac{3}{8}y^2\right) dy = \frac{3}{8}\int_0^2 y^3 dy = \frac{3}{8}\left[\frac{y^4}{4}\right]_0^2 = \frac{3}{8}\left(\frac{16}{4}\right) = \frac{3}{2}$$

For the variance, first calculate

$$
\begin{aligned}
E(Y^2) &= \int_{-\infty}^{\infty} y^2 f(y) dy = \int_0^2 y^2 \left(\frac{3}{8}y^2\right) dy = \frac{3}{8}\int_0^2 y^4 dy = \frac{3}{8}\left[\frac{y^5}{5}\right]_0^2 \\
&= \left(\frac{3}{8}\right)\left(\frac{32}{5}\right) = \frac{12}{5}
\end{aligned}
$$

Thus by the Magic Variance Formula we have:

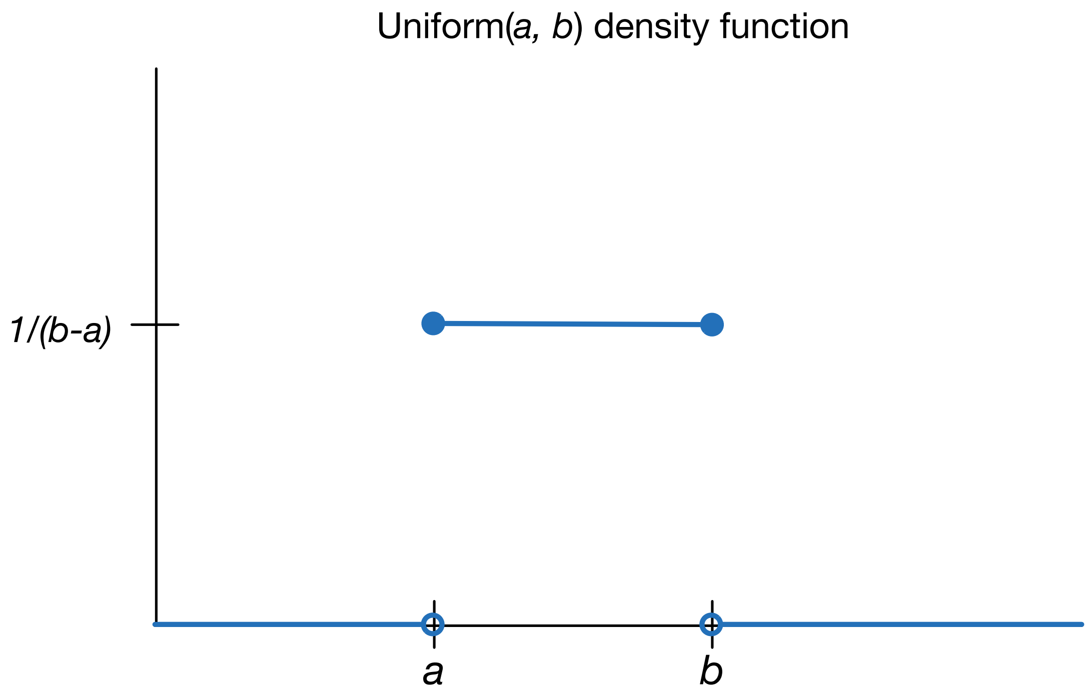
$$
\mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{12}{5} - \left(\frac{3}{2}\right)^2 = \frac{3}{20}
$$

3. Uniform Distribution

The uniform distribution describes the probability distribution on a finite interval $[a, b]$ which has the property that all sub-events are equally likely. This is a generalization of rolling a die, where all outcomes $1, 2, 3, \cdots, 6$ were equally likely.

Examples include arrival times of a bus between 8 am and $8:15$ am, which is a uniform distribution on $[0, 15]$, since each arrival time is equally likely.

What is the density function for the uniform distribution? Look at the picture below:



The uniform density is a horizontal line between a and b and is 0 otherwise. What is the height of the horizontal line? Since the integral of a probability density is 1, the area of the box must be 1. For that to be the case, the height of the box must be $\frac{1}{b-a}$

Definition:

Y has a **uniform distribution** on the interval $[a, b]$ if the pdf of Y is given by:

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

We write $Y \sim \text{Unif}(a, b)$.

Example 3:

A circle has a radius which is uniformly distributed on the interval $[0,2]$. What is the expected value of the area of the circle?

Let $Y \sim \text{Unif}(0,2)$. Then Y has density

$$f(y) = \begin{cases} \frac{1}{2} & 0 \leq y \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

The area of a circle of radius y is given by $g(y) = \pi y^2$. Then the expected value of the area is:

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy = \int_0^2 \pi y^2 \left(\frac{1}{2}\right) dy = \frac{\pi}{2}\left[\frac{y^3}{3}\right]_0^2 = \frac{\pi}{2}\left(\frac{8}{3}\right) = \frac{4}{3}\pi$$

Fact:

If $Y \sim$ Unif (a,b) then $E(Y) = \frac{a+b}{2}$ and $\text{Var}(Y) = \frac{(b-a)^2}{12}$

It makes sense that the mean of the uniform distribution is halfway between the endpoints, since Y reaches all the values between a and b with equal likelihood.

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} yf(y)dy = \int_a^b y\left(\frac{1}{b-a}\right) dy = \frac{1}{b-a}\left[\frac{y^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} \\ &= \frac{(b+a)(b-a)}{2(b-a)} = \frac{a+b}{2} \end{aligned}$$

The variance can be found similarly by computing $E(Y^2)$ and using the Magic Variance Formula.

4. Normal Distribution

The single most useful distribution is the normal distribution, also known as the Gaussian distribution or the "bell curve."

Many phenomena are approximately normally distributed. Examples include Errors in scientific measurement, due to imperfect instruments, or Scores on the SAT, in this case by design.

We will see later that, in the appropriate limit, just about everything has a normal distribution. In particular, we will see that for large enough n, we can approximate a binomial random variable with a normal distribution; this is especially useful since computations with the normal distribution are often easier than dealing with the pesky factorials in the binomial probability mass function.

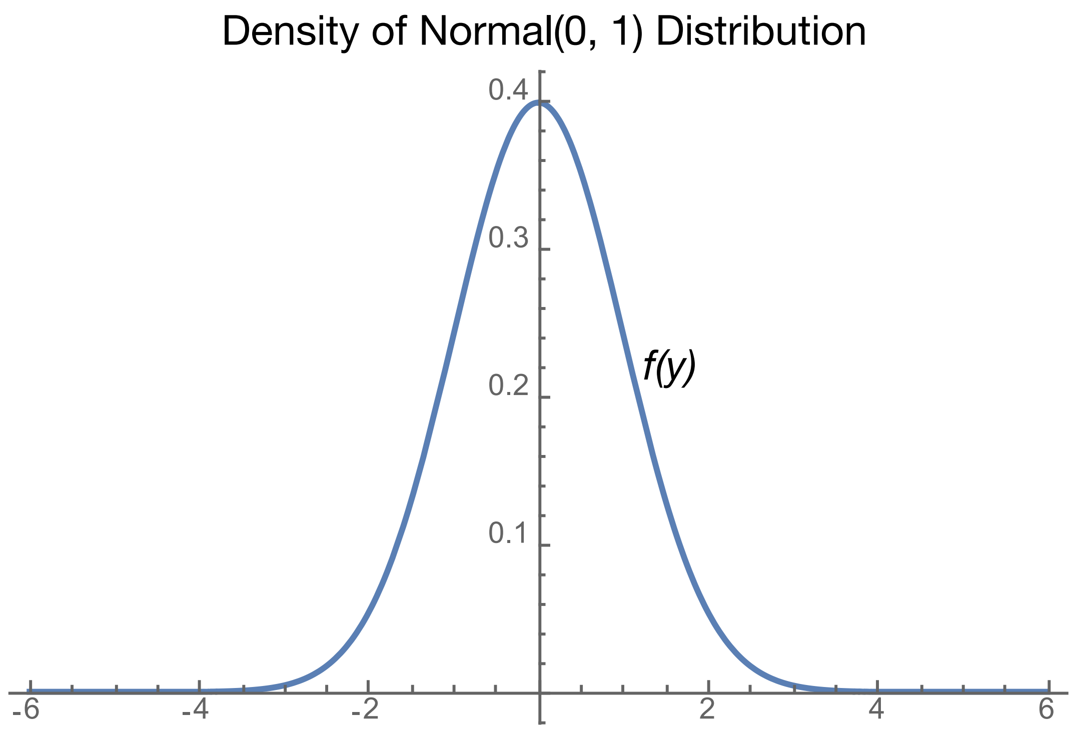
Definition:

Y has a **normal distribution** with mean μ and standard deviation σ if

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

We write $Y \sim \mathrm{N}(\mu, \sigma)$

A graph of the pdf for the normal distribution with $\mu = 0$ and $\sigma = 0$ is shown below.



Fact:

If $Y \sim$ N (μ, σ) then in fact $E(Y) = \mu$ and $\text{Var}(Y) = \sigma^2$

The fact that this is a valid pdf, i.e. that it integrates to 1, is a standard exercise in multivariable calculus, and can be found here:

Video: Gaussian Integral

Aside: If you want to see 12+ ways of doing this, check out this playlist:

Video: Gaussian Integral Playlist

The normal distribution with $\mu = 0$ and $\sigma = 1$ is so prevalent that it is called the standard normal distribution.

Definition:

The **standard normal distribution** is $Z \sim N(0, 1)$. The pdf of Z is

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Let Z be a standard normal random variable, and suppose we wish to calculate $P(-1 \leq Z \leq 1)$. Using the density of the standard normal:

$$P(-1 \leq Z \leq 1) = \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Unfortunately, there is no nice antiderivative for the integrand, so we cannot compute the integral using the fundamental theorem of calculus. One option is to use numerical integration techniques, which can be done using software packages such as Matlab or Mathematica. Another option is to use tables for the cdf of the standard normal distribution. Although this is perhaps a bit "old-school," it is important you know how to use these tables.

There are many versions of the Z-table. The one that will be provided (see website) is the actual cdf for Z, i.e. it gives values for $F(z) = P(Z \leq z)$. Since the standard normal is symmetric about 0, some tables only provide values on one side of the mean, since the others can be computed using symmetry. The table in the textbook by Wackerly et al, for example, provides $P(Z \geq z)$ for $z \geq 0$.