Math 113 Lecture #20
§8.5: Probability

The lifetime of a randomly chosen battery of a given type is a continuous random variable X: the lifetime X is a real number that ranges over an interval of real numbers.

The probability that the lifetime of such a battery is at least 1000 hours but not more than 2000 hours is denoted by

$$P(1000 \leq X \leq 2000).$$

The probability that the lifetime of a battery is at least 500 hours is denoted by

$$P(X \geq 500).$$

The probability that the lifetime of a battery is no more than 1500 hours is denoted by

$$P(X \leq 1500).$$

Each of these probabilities is a number between 0 and 1.

How do we associate a number between 0 and 1 to a given probability?

Each continuous random variable X has a nonnegative **probability density function** $f(x)$ associated with it.

The probability that X lies between two values a and b is given by integration:

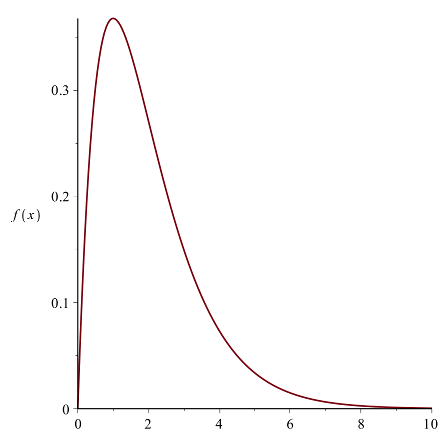
$$P(a \leq X \leq b) = \int_a^b f(x)\ dx.$$

The probability that a continuous random variable is a real number is 1, which means that

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)\ dx = 1.$$

Example. Is the nonnegative function $f(x) = xe^{-x}$ for $x \geq 0$ and $f(x) = 0$ for $x < 0$, a probability density function?

Here is the graph of $f(x)$.



To determine if it is or is not, we compute the value of the hopefully convergent improper integral to see if it is 1:

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x) &= \int_0^\infty xe^{-x}dx \\
&= \lim_{A\to\infty} \int_0^A xe^{-x}dx \\
&= \lim_{A\to\infty} \big[-xe^{-x} - e^{-x}\big]_0^A \\
&= \lim_{A\to\infty} \big[-Ae^{-A} - e^{-A} + 1\big] = 1.
\end{aligned}$$

Thus $f(x)$ is a probability density function, and we can use it to compute probabilities.

The probability that a continuous random variable X with probability density function X has a value between 1 and 2 is

$$\begin{aligned}
P(1 \leq X \leq 2) &= \int_1^2 xe^{-x}dx \\
&= \big[-xe^{-x} - e^{-x}\big]_1^2 \\
&= -2e^{-2} - e^{-2} + e^{-1} + e^{-1} \\
&\approx 0.3297530327
\end{aligned}$$

There is an approximately 33% change that the value of X is between 1 and 2.

Suppose you are waiting for a company to answer your phone call.

How long can you expect to wait?

Let $f(t)$ be the probability density function where t is time measured in minutes.

For a sample of N people who have called, we can assume that none will wait more than 60 minutes.

On a small subinterval $[t_{i-1}, t_i]$ of length Δt of $[0, 60]$ the probability that someone's call gets answers is approximately $f(\bar{t}_i)\Delta t$ for some $\bar{t}_i \in [t_{i-1}, t_i]$.

Of the N people who have called, we expect that during $[t_{i-1}, t_i]$ the number of calls that will be answered is approximately $Nf(\bar{t}_i)\Delta t$, and the time they each waited is about $\bar{t}_i$.

So the total of the waiting times over $[0, 60]$ is

$$\sum_{i=1}^n N\bar{t}_i f(\bar{t}_i)\Delta t.$$

Then the average waiting time is

$$\frac{1}{N}\sum_{i=1}^n N\bar{t}_i f(\bar{t}_i)\Delta t = \sum_{i=1}^n \bar{t}_i f(\bar{t}_i)\Delta t.$$

Recognizing this as a Riemann sum for the function $tf(t)$, we obtain in the limit that the average waiting time or mean is

$$\int_0^{60} tf(t)\ dt.$$

Passing to an improper integral we define the **mean** of a probability density function $f(t)$ to be

$$\mu = \int_{-\infty}^{\infty} tf(t)\ dt.$$

Another measure for a probability density function is its **median**, the number m that satisfies

$$\int_m^{\infty} f(x)dx = \frac{1}{2}.$$

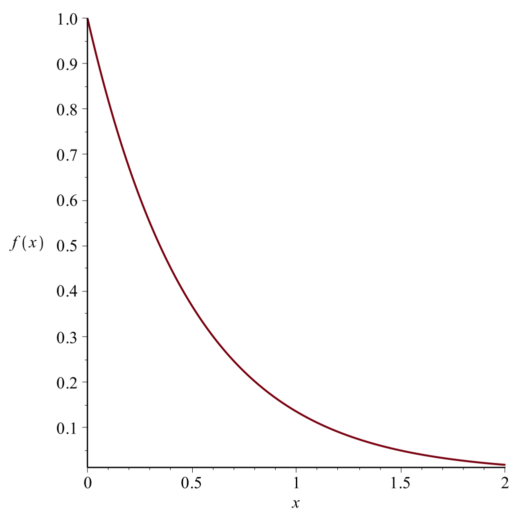
Unlike the mean, finding the median may involve a numerical approximation for the value of m.

Example. An exponentially decreasing probability density function is commonly used to model wait times or equipment failure times.

For each $c > 0$ an exponentially decreasing probability density function is

$$f(x) = \begin{cases} ce^{-cx} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}.$$

Here is the graph of $f(x)$ for $c = 2$.



For arbitrary $c > 0$, the mean of $f(x)$ is

$$\begin{aligned}\mu &= \int_{-\infty}^{\infty} xf(x)\ dx = \int_0^{\infty} cxe^{-cx}dx \\ &= \lim_{A\to\infty} \int_0^A cxe^{-cx}dx = \lim_{A\to\infty} \left[-xe^{-cx} - \frac{e^{-cx}}{c}\right]_0^A \\ &= \lim_{A\to\infty} \left[-Ae^{-cA} - \frac{e^{-cA}}{c} + \frac{1}{c}\right] = \frac{1}{c}.\end{aligned}$$

Thus the average waiting time is $1/c$.

The median m of $f(x)$ satisfies

$$\frac{1}{2} = \int_m^\infty ce^{-cx}dx = \lim_{A\to\infty} \left[-e^{-cx}\right]_m^A = \lim_{A\to\infty} \left[-e^{-cA} + e^{-cm}\right] = e^{-cm}.$$

We can solve this for

$$m = \frac{\ln 2}{c}.$$

For $c = 2$ the median is $m \approx 0.346$ which is different than the mean of 0.5.

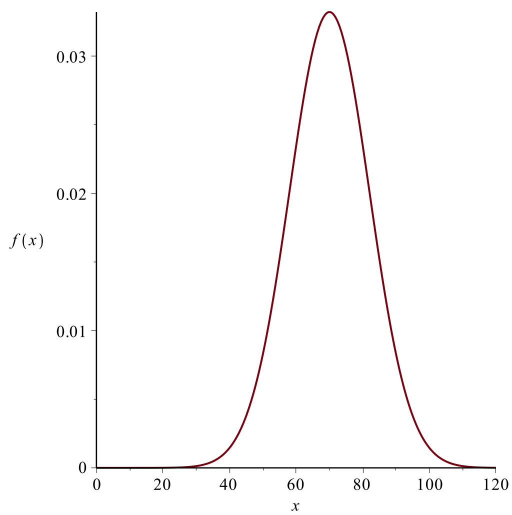
Normal Distributions. Many continuous random variables X, e.g. scores on exams, have a normal probability density distribution which is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

where μ is the mean of $f(x)$ and σ is the standard deviation that measures how spread out the values of X.

The smaller the value of σ the more concentrated the probability density is near the mean, and the larger the value of σ the less concentrated the probability density is near the mean.

Here is a "curve" for an exam with a mean of 70 and a standard deviation of 12.



Computing probability with a normal distribution requires numerical integration to approximate the integrals.

For a normal distribution with mean 70 and standard deviation of 12, the probability that a student gets an exam score of between 70 and 90 is

$$\int_{70}^{90} \frac{1}{12\sqrt{2\pi}} \exp\left(-\frac{(x-70)^2}{2(12^2)}\right) dx \approx 0.4522096478.$$

For the same normal curve, the probability that a student gets an exam score of at least 80 is

$$\begin{aligned}\int_{80}^{\infty} \frac{1}{12\sqrt{2\pi}} \exp\left(-\frac{(x-70)^2}{2(12^2)}\right) dx &\approx \int_{80}^{120} \frac{1}{12\sqrt{2\pi}} \exp\left(-\frac{(x-70)^2}{2(12^2)}\right) dx \\ &\approx 0.2023129267.\end{aligned}$$

Why could we replace ∞ with 120? Because for $x > 120$ there isn't much probability, or the area under the curve over $x > 120$ is very very small, and can safely be ignored.