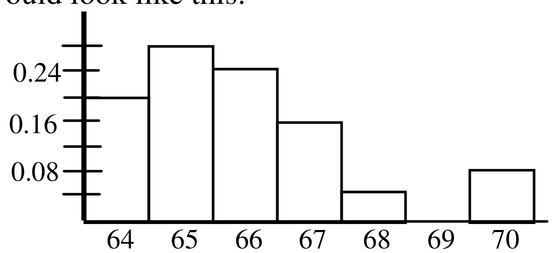
Stat 400 section 4.1 Continuous Random Variables

notes by Tim Pilachowski

Suppose we measure the heights of 25 people to the nearest inch and get the following results:

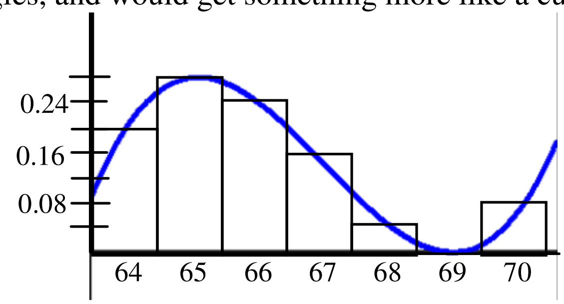
height (in.)	64	65	66	67	68	69	70
frequency	5	7	6	4	1	0	2
probability	0.2	0.28	0.24	0.16	0.04	0	0.08

The probability distribution graph would look like this:



Recall that the sum of the areas of the rectangles, which is the sum of the probabilities, equals 1.
(probability as area: area of each bar = height times width = probability times 1 = percentage of people having that height = relative frequency of that height = probability of a person having that height)

If we were to measure to the nearest half-inch, or tenth of an inch, or hundredth, or thousandth, etc., etc., etc., we'd get ever-more-narrow rectangles, and would get something more like a curve:



The area under the curve for a given interval would be the probability of people having heights within that interval.

And so we come to a definition—a *probability density function* or *probability distribution function* for a continuous random variable is a function $f(x)$ such that, on an interval $[a, b]$,

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

A pdf $f(x)$ has two necessary characteristics:

1. $f(x) \ge 0$ for all values of x in the interval $[a, b]$ (since all probabilities, and therefore areas under the curve, are zero or positive)

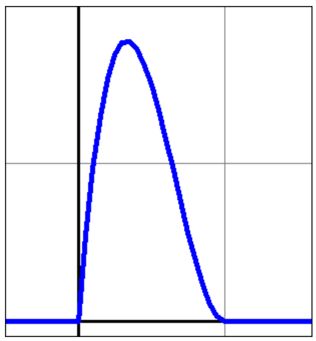
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (since the sum of all probabilities = 1 = area under the curve over the entire domain)

Example A: Verify that $f(x) = \begin{cases} 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$ is a probability density function. *answer*: Yes, it is.

Example B: Find the value of the constant k that makes $f(x) = kx(1 - x)$ a probability density function on the interval $0 \le x \le 1$. *answer*: 6

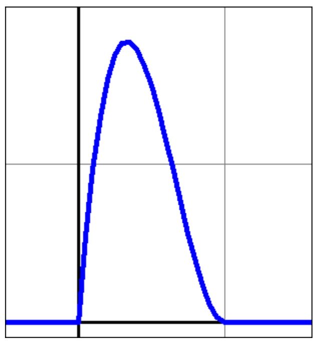
Example A revisited: A continuous random variable X has the probability density function

$f(x) = \begin{cases} 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$. Find $P\left(0 \le X \le \frac{1}{2}\right)$. *answer*: $\frac{11}{16} = 0.6875$ *Note both versions are exact*.



Example A once more: A continuous random variable X has the probability density function

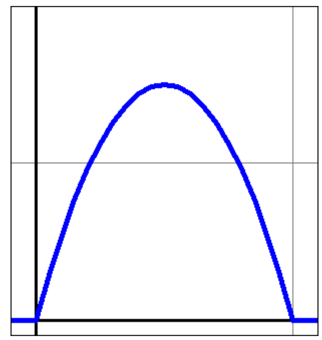
$f(x) = \begin{cases} 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$. Find $P\left(\frac{1}{2} \le X \le 1\right)$. *answer*: $\frac{5}{16} = 0.3125$



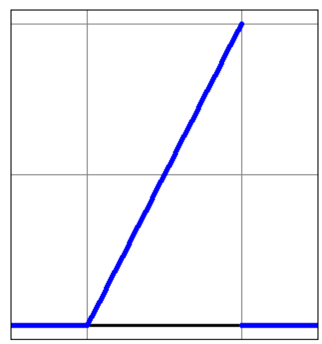
Note that we are finding probabilities for intervals as opposed to specific values. Indeed, if we tried to find $P\left(X=\frac{1}{2}\right)$, the calculus would give us $\int_{1/2}^{1/2} 12x(1-x)^2\, dx = 0$. This isn't to say that random variable X will never take on a specific value, but rather that the probability for that one specific exact value is so small that it is negligible. Parallels in life include:

- A meteorologist will predict rain for the afternoon, not rain for 2:07 pm.
- The bullseye on a dartboard is a space, not an infinitesimally small point.
- A person growing from 65 to 67 inches will at some time be exactly 66 inches, but we have no way to measure with that much accuracy to specify exactly when it happens.

Example B revisited: A continuous random variable X has the probability density function $f(x)=\begin{cases} 6x(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$. Find and compare $P\left(X \le \frac{1}{2}\right)$ with $P\left(X < \frac{1}{2}\right)$. *answer*: $\frac{1}{2} = 0.5$



Example C: A continuous random variable X has the probability density function $f(x)=\begin{cases} 2x-2 & 1 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$. a) Verify that $f(x)$ is a probability density function. b) Find b such that $P(X \le b) = 0.5$. *answer*: $1+\frac{\sqrt{2}}{2} \approx 1.7071$

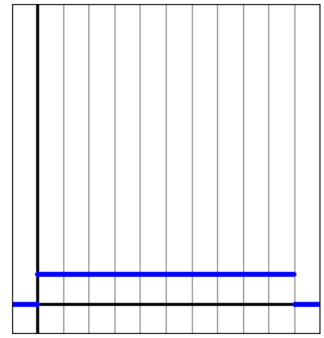


Example D: A continuous random variable X has probability density function $f(x)=\begin{cases} \frac{1}{9}x^2 & 0 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$.

a) Verify that $f(x)$ is a probability density function. b) Find $P(1 \le X \le 2)$. c) Find b such that $P(X \le b) = 0.125$.
answers: $\frac{7}{27}$, 1.5

Example E: A continuous random variable X has probability density function $f(x) = \begin{cases} \frac{1}{10} & 0 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$.

a) Verify that $f(x)$ is a probability density function. b) Find $P(2 \le X \le 8)$. c) Find b such that $P(X \le b) = 0.5$.
answers: 0.6, 0.5



Example E is a *uniform probability distribution*. Note a few characteristics.

1a. The "area under the curve" has the shape of a rectangle, so for Example E, $P(a \le X \le b)$ will always equal "length times width" $= (b - a)\left(\frac{1}{10}\right)$.

2a. For any uniform probability density function $f(x)$, $P(a \le X \le b) = (b - a) * f(x)$.

1b. $f(x) = \begin{cases} \frac{1}{10} & A \le x \le B \\ 0 & \text{otherwise} \end{cases}$ would be a probability density function as long as $B - A = 10$.

2b. A uniform probability density function has the general form $f(x;\, A,\, B) = \begin{cases} \frac{1}{B-A} & A \le x \le B \\ 0 & \text{otherwise} \end{cases}$.

3. As with all continuous random variables X, $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = P(a < X < b)$.

Example F: Gauge 4 aluminum sheets should have a thickness of 0.2043 inches, with a tolerance of ± 0.011 inches allowed. During the rolling process that produces metal sheet stock, a certain amount of "bowing" occurs in the rollers. The quality control department of an aluminum factory believes that one of the company's rolling machines is producing sheets with thicknesses varying between 0.1931 and 0.2168 inches. If we assume thickness is a uniform random variable, what is the probability that a randomly chosen sheet will be within the allowed tolerance? *answer*: ≈ 0.9283

Example G. For a particular machine, its useful lifetime (random variable T) is modeled by $f(t)=\begin{cases}0.1e^{-0.1t} & 0\le x<\infty \\ 0 & \text{otherwise}\end{cases}$. Verify that f is a probability density function, then find the probability that the machine will last a) less than 2 years, b) more than 2 but less than 4 years, and c) more than 4 years.

answers: a) $-e^{-0.2}+1\cong 0.181$, b) $-e^{-0.4}+e^{-0.2}\cong 0.148$, c) $e^{-0.4}\cong 0.670$

side note: The three probabilities add to 1, as they should. (We must add the exact values, because the approximate answers have a rounding error.)

Example G is an *exponential probability distribution*. Exponential probabilities most often describe the distance between events with uniform distribution in time, and are used widely in analysis of reliability, which deals with the amount of time a product lasts (as in Example G). The length of time a long distance phone call lasts follows an exponential distribution—there are more phone calls that last a shorter amount of time and fewer calls that last a long time. Another example is the amount of money a customer spends for one trip to the supermarket—there are more people that spend less money and fewer people that spend large amounts of money.

Note a few characteristics of exponential probability distributions.

1a. An exponential probability density function has the general form $f(x)=\begin{cases}k\,e^{-k(x-A)} & A\le x<\infty \\ 0 & \text{otherwise}\end{cases}$.

1b. In Example G, we had the case (as often happens) where $A = 0$.

2a. An exponential probability distribution is "memoryless". This means that if a random variable X is exponentially distributed, its conditional probability obeys $P(X > a + b \mid X > a) = P(X > b)$.

2b. In Example G, the probability that "the machine will last another 4 years" is the same whether we start counting when the machine is new, when it is 1 year old, when it is 3 years old, etc.

3. As with all continuous random variables X, $P(a \le x \le b) = P(a < x \le b) = P(a \le x < b) = P(a < x < b)$.

Example H. Suppose that the number of days a patient spends in ICU can be approximated with a probability distribution function $f(t) = \begin{cases} 0.4e^{-0.4(t-1)} & t > 1 \\ 0 & \text{otherwise} \end{cases}$. What is the probability that a patient will spend a) at most 3 days in ICU, b) more than 3 days in ICU? *answers*: a) ≈ 0.5507, b) ≈ 0.4493