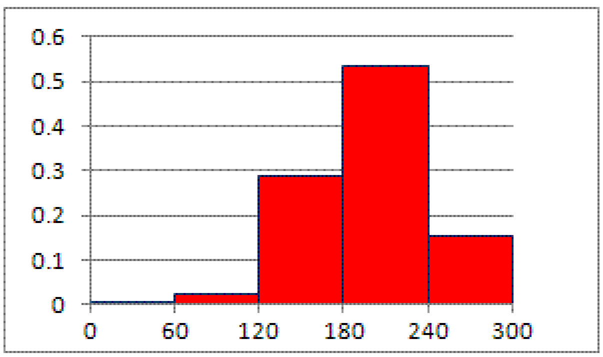
Example 6.01: [Navidi section 2.4; Devore sections 4.1-4.2]

"Exact lifetime" is a **continuous** random quantity, but
"Measured lifetime to the nearest minute" is a **discrete** random quantity.

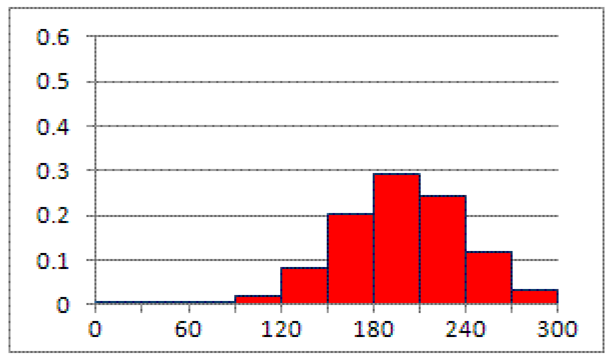
In a bar chart, the height of each bar represents the probability. Note that as the measurements become more precise, the number of intervals increases and the width, probability and height of each bar decrease. The visual effect is misleading: it appears that the total probability is decreasing to zero as the number of intervals increases to infinity.

$T$ = lifetime of a test wire in seconds.

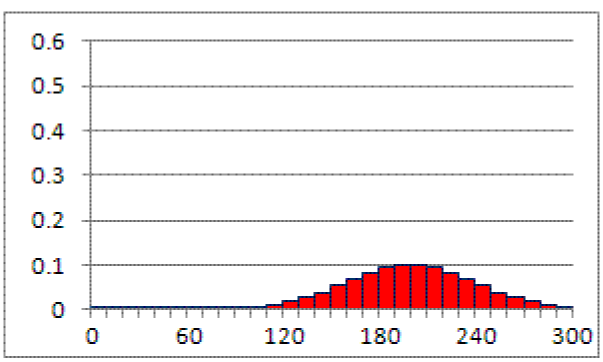
Bar chart (class width = 60 s)



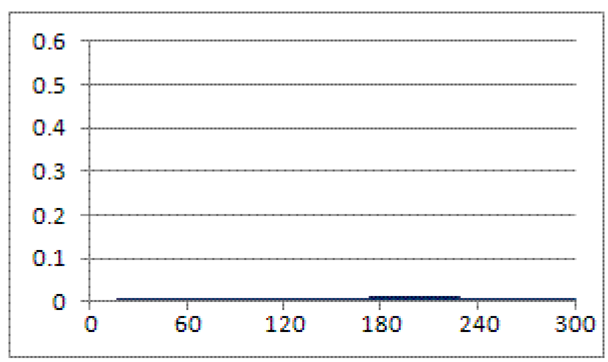

Bar chart (class width = 30 s)




Bar chart (class width = 10 s)




Bar chart (class width = 1 s)

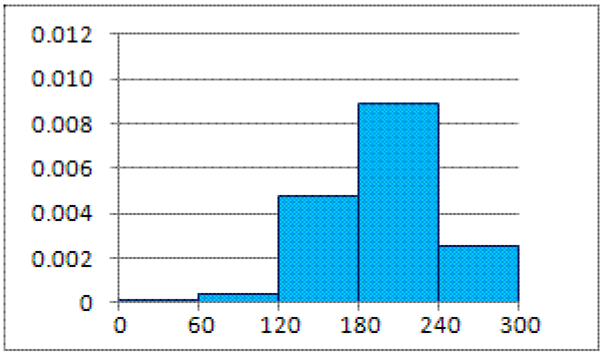



The height of each bar is the probability of observing a lifetime in that interval. As the time intervals get shorter and shorter, so the probabilities for each interval decrease – in the limit, to zero!

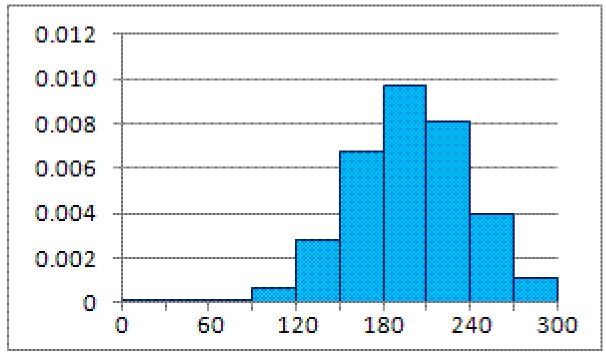
<u>Example 6.01</u> (continued)

Much more natural is the **probability histogram**, where the *area* of each bar represents the probability that the random quantity lies in the interval covered by the width of the bar. The resulting re-scaling of the vertical axis preserves the overall shape of the probability histogram as the class intervals are repeatedly subdivided. The total area thus remains 1 even as the number of intervals $\rightarrow \infty$.

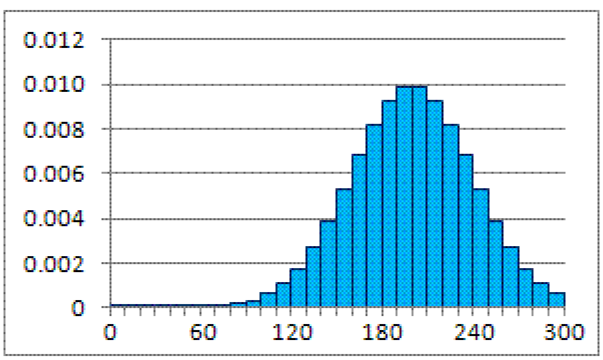
Histogram (class width = 60 s)



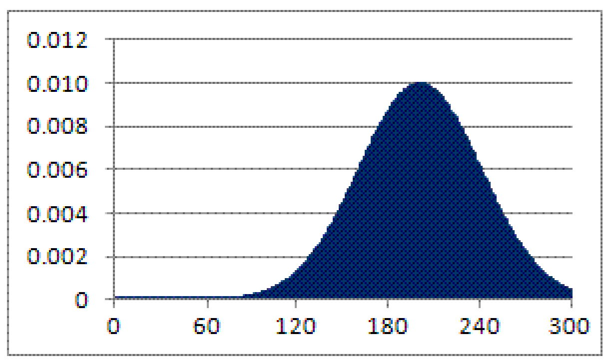

Histogram (class width = 30 s)




Histogram (class width = 10 s)




Histogram (class width = 1 s)




In the limit of infinitely many bars each of width zero, the jagged edges of the probability histogram merge into a smooth curve – the graph of the continuous probability function.

In the probability histogram,

Bar height =

As the bar width $\rightarrow 0$, bar height $\rightarrow f(x)$ = the **probability density function** (*p.d.f.*) .

The total area remains 1.

Thus two conditions for a function $f(x)$ of a continuous variable $x$ to be a valid probability density function are:

1)

2)

From a discrete probability histogram,

P[$a < X \leq b$] = the sum of the areas of the bars from $x = a$ to $x = b$
(excluding $x = a$ but including $x = b$),
= (c.d.f. at $x = b$) − (c.d.f. at $x = a$)

and P[$X = a$] = the area of the single bar centered on $x = a$.

For a continuous probability distribution, it then follows that

and P[$X = a$] =

<u>Example 6.02</u>

Verify that $f(x) = 2x \quad (0 \le x \le 1)$ is a legitimate probability density function and find $\mathrm{P}\left[-\frac{1}{2} < X < \frac{1}{2}\right]$.

Note that, by default, $f(x) = 0$ for all values of $x$ not mentioned in the definition.

---

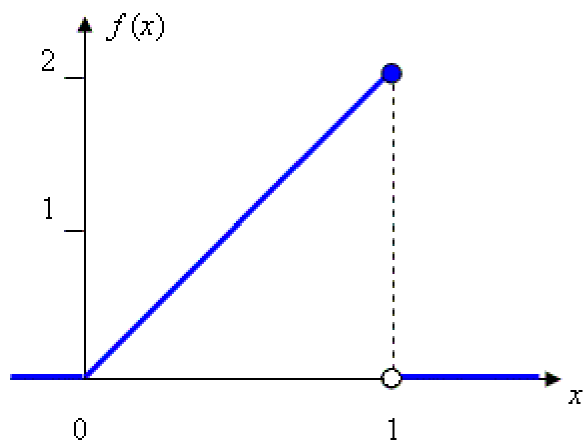
- On $0 \le x \le 1$, $f(x) = 2x \ge 0$. Elsewhere $f(x) = 0$. $\therefore f(x) \ge 0 \quad \forall x$.

- $\int_{-\infty}^{\infty} f(x)\, dx =$

**OR:**

The graph of $f(x)$ is on or above the $x$-axis everywhere
and
the total area under the graph

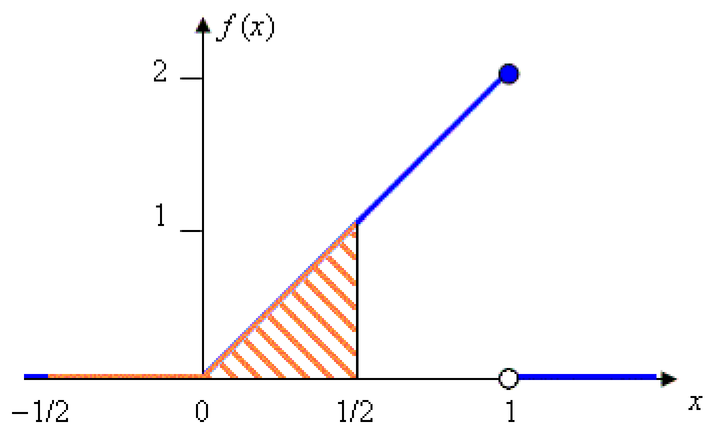
=




Therefore $f(x)$ is a valid *p.d.f.*

$\mathrm{P}\left[-\frac{1}{2} < X < \frac{1}{2}\right]$ = area under $f(x)$ between $x = -\frac{1}{2}$ and $x = \frac{1}{2}$

=

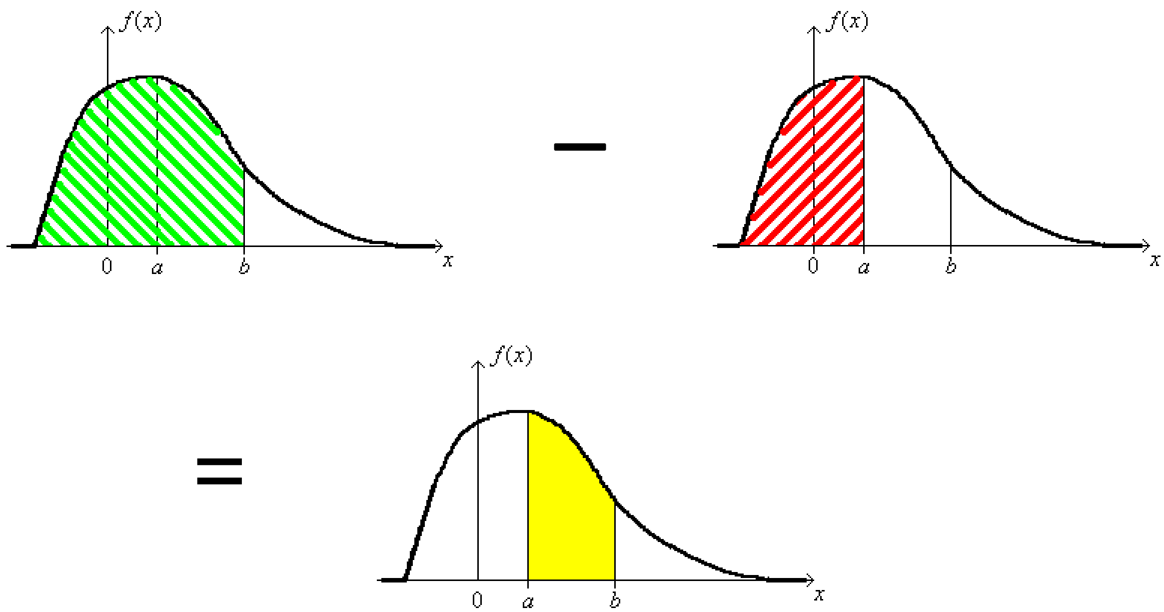



**OR:**

$$\mathrm{P}\left[-\frac{1}{2} < X < \frac{1}{2}\right] = \int_{-1/2}^{1/2} f(x)\, dx =$$

The **cumulative distribution function** (*c.d.f.*) is defined by

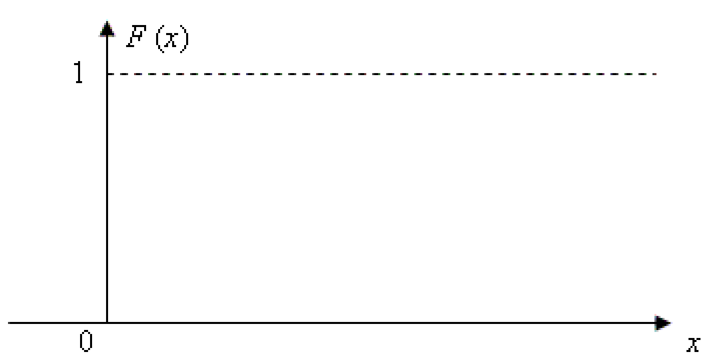
$$\boxed{F(x) \;=\; \mathrm{P}[X \le x] \;=\; \int_{-\infty}^{x} f(t)\, dt}$$




$$\boxed{\mathrm{P}\left[a < X < b\right] \;=\; F\left(b\right) - F\left(a\right)}$$

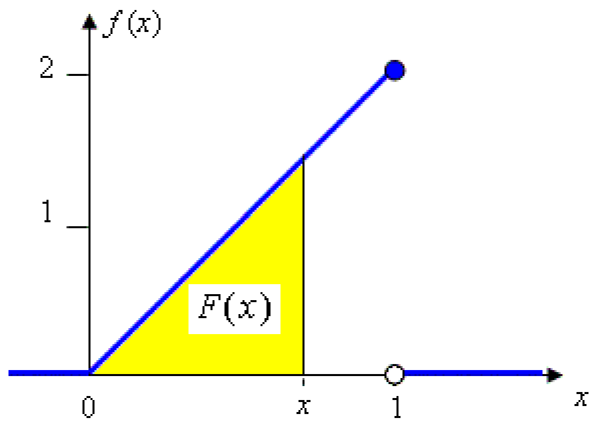
$F(-\infty) =$ $\qquad$ $F(+\infty) =$ $\qquad$ $0 \;\le\; F(x) \;\le\; 1$ for all $x$.

The *c.d.f.* is a non-decreasing function of $x$ and $\dfrac{d}{dx}\big(F(x)\big) \;=\; f(x) \;\ge\; 0 \quad \forall x$.



Example 6.02 (continued)

Find the cumulative distribution function for $f(x) = 2x \quad (0 \le x \le 1)$ .




Graphical method:

$x < 0 \quad \Rightarrow \quad F(x) =$

$0 \le x \le 1 \quad \Rightarrow \quad F(x) =$

$x > 1 \quad \Rightarrow \quad F(x) =$

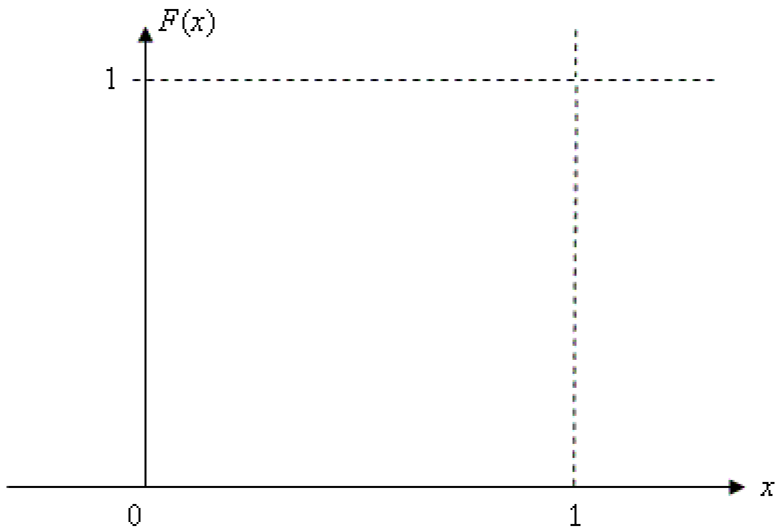
Calculus method: $F(x) = \int_{-\infty}^{x} f(t)\, dt$

$x < 0 \quad \Rightarrow \quad F(x) =$

$0 \le x \le 1 \quad \Rightarrow \quad F(x) =$

$x > 1 \quad \Rightarrow \quad F(x) =$

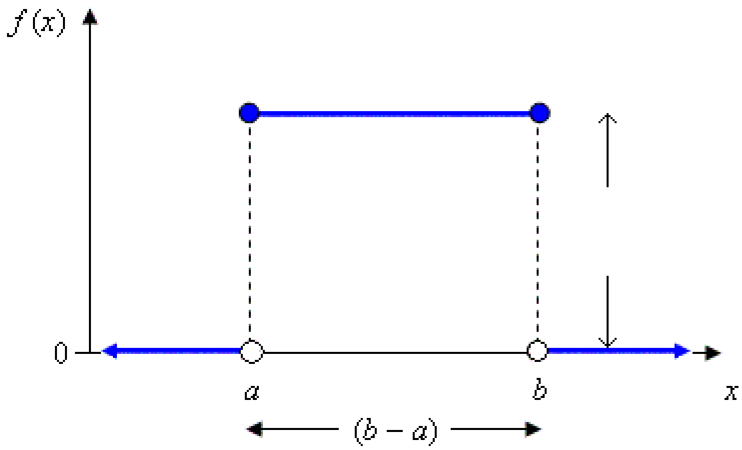
<u>Example 6.02</u> (continued)




Note how the *c.d.f.* is a non-decreasing continuous function between $F = 0$ and $F = 1$.

Example 6.03 **The Continuous Uniform Distribution**

Find the *p.d.f.* and the *c.d.f.*




The probability density function is

The cumulative distribution function is

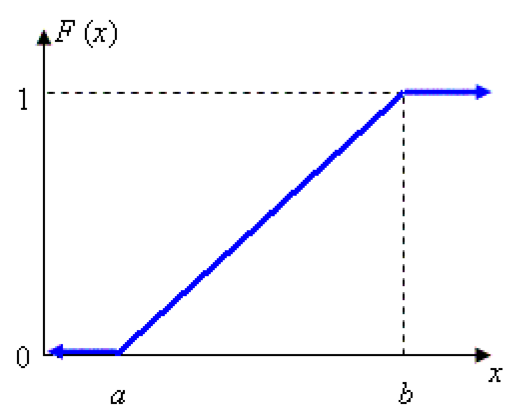
$$F(x) = \int_{-\infty}^{x} f(t)\,dt \ .$$

When $x < a$, $F(x) =$

When $x > b$, $F(x) =$

When $a \le x \le b$, $F(x) =$ **OR**

Therefore $F(x) = \Bigg\{$



## Population Mean and Population Variance for Continuous Probability Distributions

The discrete probability point masses $p_i$ are "smeared out" into infinitely many elementary masses $f(x)\,dx$ covering infinitesimal intervals $dx$. The expression for the population mean (expected value) of the random variable $X$ thus evolves from the discrete case $\mathrm{E}[X] = \sum_{\forall i} p_i x_i$ to the continuous equivalent

$$\mu = \mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

The expression for the population variance is amended in a similar manner, from $\sigma^2 = \mathrm{V}[X] = \sum_{\forall i} p_i \left(x_i - \mu\right)^2$ to

$$\sigma^2 = \mathrm{V}[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \mathrm{E}\left[X^2\right] - \left(\mathrm{E}[X]\right)^2$$

---

<u>Example 6.03</u> (continued)

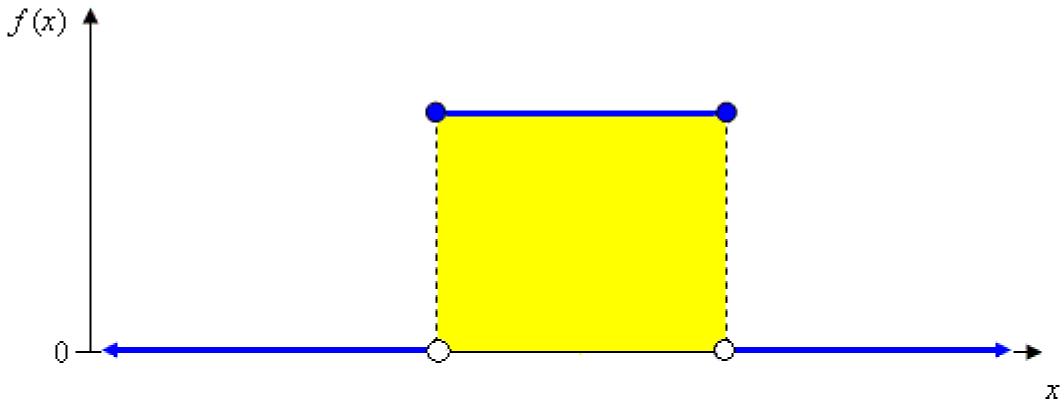
Find the population mean and variance for the continuous uniform distribution $\mathrm{U}(a, b)$ .

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_{-\infty}^{a} 0\, dx + \int_{a}^{b} x\left(\frac{1}{b-a}\right) dx + \int_{b}^{\infty} 0\, dx =$$

$$\mathrm{V}[X] = 0 + \int_a^b \left\{x - \left(\frac{a+b}{2}\right)\right\}^2 \left(\frac{1}{b-a}\right) dx + 0 = \left(\frac{1}{b-a}\right)\left[\frac{1}{3}\left\{x - \left(\frac{a+b}{2}\right)\right\}^3\right]_a^b$$

$$= \frac{1}{3}\left(\frac{1}{b-a}\right)\left\{\left(\frac{b-a}{2}\right)^3 - \left(\frac{a-b}{2}\right)^3\right\} = \frac{1}{3}\,\frac{(b-a)^3}{(b-a)}\,\frac{2}{8}$$

$$\therefore \quad \sigma^2 = \boldsymbol{\frac{(b-a)^2}{12}} \quad \text{and} \quad \sigma = \frac{\text{(range)}}{\sqrt{12}} .$$


f(x)
0
x

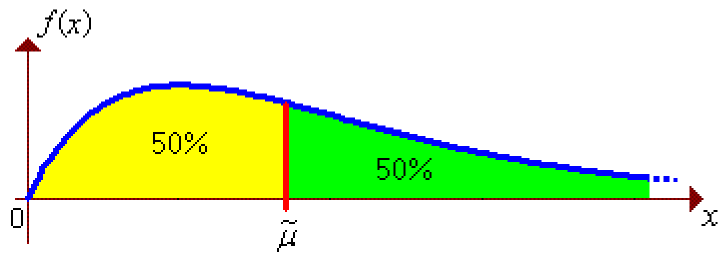
## The Population Median

The sample median is the "half-way" value in an ordered set of values.

The population median $\tilde{\mu}$ is the value that divides the area under a probability curve (or histogram) into two equal halves:

$$\text{P}\left[X \le \tilde{\mu}\right] = \text{P}\left[X \ge \tilde{\mu}\right] = \frac{1}{2}$$

$$\Rightarrow \quad F\left(\tilde{\mu}\right) = \frac{1}{2}$$




The population quartiles are defined in a similar way:

The lower quartile $x_L$ is such that $\text{P}\left[X \le x_L\right] = F\left(x_L\right) = \frac{1}{4}$

The upper quartile $x_U$ is such that $\text{P}\left[X \le x_U\right] = F\left(x_U\right) = \frac{3}{4}$

---

<u>Example 6.04</u>

For the random quantity $X$ whose probability density function is

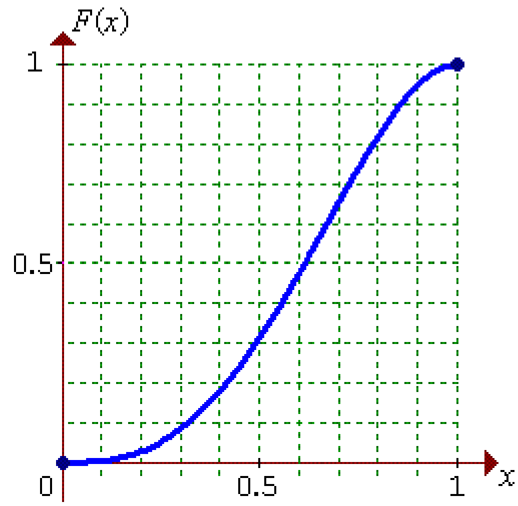
$$f\left(x\right) = 12x^2\left(1-x\right) \qquad \left(0 \le x \le 1\right)$$

find

(a) the population mean;
(b) the population standard deviation;
(c) the population median;

---

Example 6.04 (continued)




The value $x_\alpha$ to the left of which a proportion $\alpha$ of all values lies is a quantile, which satisfies $\mathrm{P}\left[X \le x_\alpha\right] = F\left(x_\alpha\right) = \alpha$ .

The median is the 50$^{\text{th}}$ percentile. The quartiles are the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles.

[End of Chapter 6]