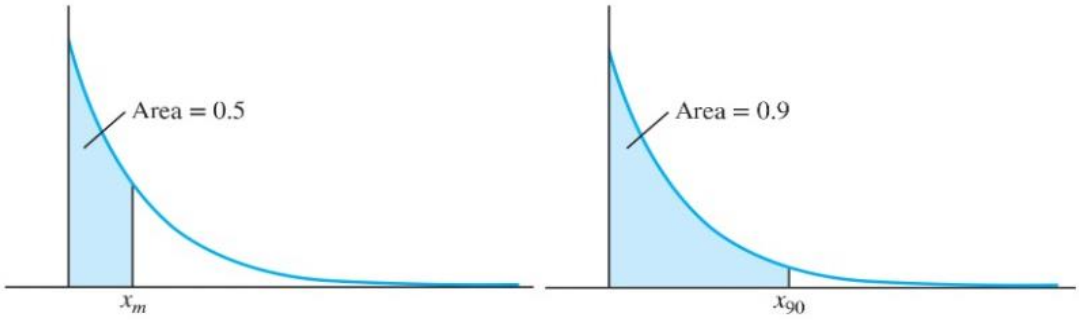
The Median and Percentiles

The **median** of a data set divides the data into two equal parts: the bottom 50% and the top 50%. In terms of a probability density function, the median (denoted by x_m) is the point at which half the area under the curve is to the left, and half the area is to the right.

More generally, for each p between 0 and 100, the ***p*th percentile** (denoted by x_p) is the number that divides the bottom $p\%$ of the data from the top $(100 - p)\%$. For example, the 90th percentile (denoted by x_{90}) is the number that divides the bottom 90% of the data from the top 10%. In terms of a probability density function, the 90th percentile is the point x_{90} at which the area under the curve to left of x_{90} is 0.9, and the area under the curve to the right of x_{90} is 0.1.



Definition:

Let X be a continuous random variable with probability mass function $f(x)$ and cumulative distribution function $F(x)$.

- The median of X is the point x_m that solves the equation

$$F(x_m) = P(X \leq x_m) = \int_{-\infty}^{x_m} f(x)dx = 0.5$$

- If p is any number between 0 and 100, the pth percentile is the point x_p that solves the equation

$$F(x_p) = P(X \leq x_p) = \int_{-\infty}^{x_p} f(x)dx = \frac{p}{100}$$

- The median is the 50th percentile.

Ex. A certain radioactive mass emits alpha particles from time to time. The time between emissions, in seconds, is random, with probability density function

$$f(x) = \begin{cases} 0.1e^{-0.1x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

(a) Find the median time between emissions.

(b) Find the 60^{th} percentile of the times.