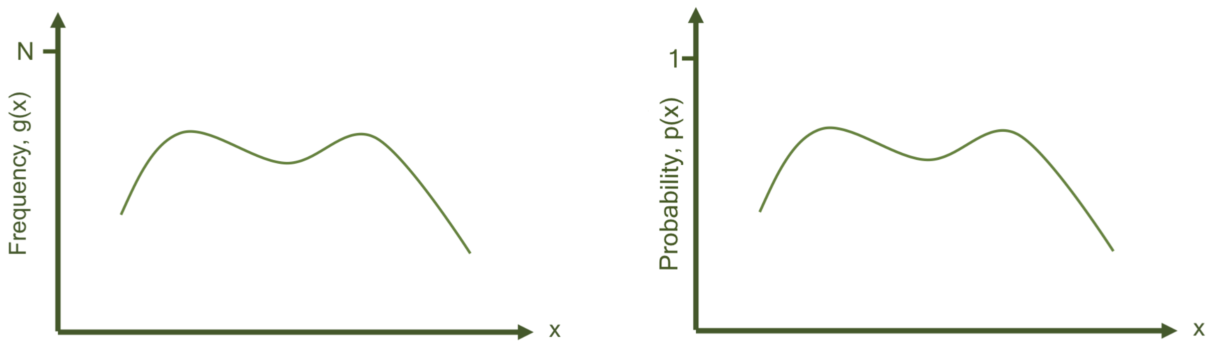
MMaD: Lecture 3 handout

Probability distribution

A histogram illustrates a frequency distribution of a variable x divided into discrete classes. If we increased the number of classes to infinitely-many, we obtain the continuous **frequency distribution** $g(x)$, a function describing the number of data points as x varies. Here, the area between $a \leq x \leq b$ gives the number of times (i.e. the frequency) that x takes a value in this interval.

If we "normalise" the data by dividing by the total number N of data points, we obtain a **probability distribution** $p(x)$. This has the same shape as the frequency distribution, but the height is rescaled from a frequency to a probability (between 0 and 1). Now, the area between $a \leq x \leq b$ gives the probability that x takes a value in this interval.



(a) Frequency $g(x)$ (b) Probability $p(x)$

$$\int_{-\infty}^{\infty} g(x)\mathrm{d}x = N \qquad \int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$$

Mean and variance for a continuous variable x with probability distribution $p(x)$:

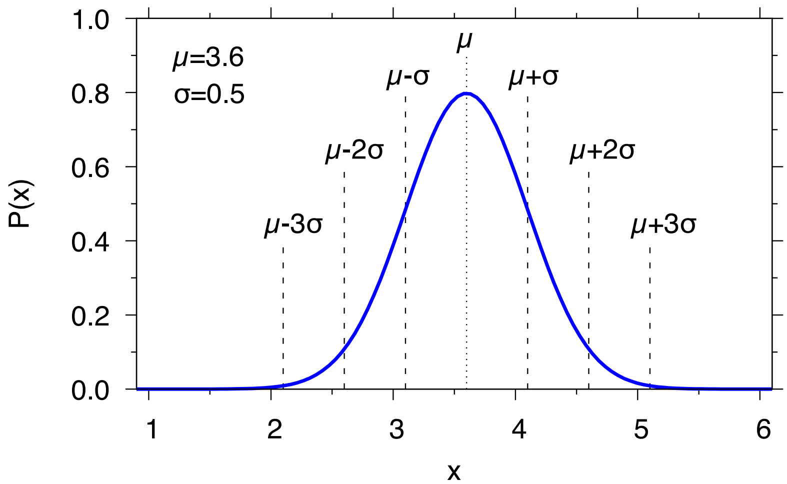
$$\bar{x} = \int_{-\infty}^{\infty} xp(x)\mathrm{d}x \qquad \sigma^2 = \int_{-\infty}^{\infty} (x - \bar{x})^2 p(x)\mathrm{d}x$$

Normal distribution

This common distribution has a probability distribution given by:

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

where, μ is the **mean** and σ is the **standard deviation**.



68% of all of the points on a normal distribution curve lie within 1 standard deviation of the mean (from $\mu - \sigma$ to $\mu + \sigma$). 95% within 2 standard deviations and 99% within 3 standard deviations.

If X is normally-distributed with a mean, μ and standard deviation, σ, we write this as:

$$X \sim N(\mu, \sigma^2)$$

Log-normal distribution

A log-normally distributed variable x has the probability density function, and properties:

$$P(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$$

Mean	$\exp\left(\mu + \frac{\sigma^2}{2}\right)$
Median	$\exp(\mu)$
Mode	$\exp(\mu - \sigma^2)$
Variance	$[\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$

Example

A company manufactures a microprocessor to control industrial robots. Existing data indicates that the life span of the microprocessors is described by a normal distribution with mean $\mu = 4000$ hours and standard deviation $\sigma = 200$ hours. Determine the probability that the life span of such a microprocessor is:

1. Less than 3700 hours.
2. Between 3700 hours and 4250 hours.
3. More than 4250 hours.