Proability

- The **probability** of an event A is the **expected relative frequency** of **event *A*** in a **large number of trials**. The following is true of probabilities:
  - $0 \le P(A) \le 1$
  - $\sum_{\forall \text{ events}} P(E) = 1$
  - $\boxed{P(A \cup B) = P(A) + P(B) + P(A \cap B)}$. In the **special case** of **mutually exclusive events**, this becomes $P(A \cup B) = P(A) + P(B)$.
  - $\boxed{P(A \cap B) = P(A)P(B \mid A)} \Rightarrow \boxed{P(B \mid A) = \frac{P(A \cap B)}{P(A)}}$. In the **special case** of **independent events**, this becomes $P(A \cap B) = P(A)P(B)$. A few notes:
    - We can easily expand this to more than two events by noting that $P(A \cap B \cap C) = P(C)P(A \cap B \mid C)$.
    - If $A = A_1 + A_2 + \cdots + A_n$, then $P(A \mid B) = \sum_i P(A_i \mid B)$. Also, if the A_is exhaust the sample space, $P(B) = \sum_i P(A_i)P(B \mid A_i)$.
  - **Bayes' Theorem** states that

$$\boxed{P(A \mid B) = P(B \mid A)\frac{P(A)}{P(B)}}$$

    It is sometimes easier to rewrite $P(B) = P(A)P(B \mid A) + P(A')P(B \mid A')$, when P($B$) is not known, in which case it becomes

$$\boxed{P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(A')P(B \mid A')}}$$

    In fact, this is a special case of a more general form; if the sample space is not just divided into A and A', but into *any* set of **mutually exclusive events** A_i that **exhaust the sample space**, then

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{\sum_i P(A_i)P(B \mid A_i)}$$

- **Combinations and Permutations**:

Situation	Number of ways to arrange
A set of n different objects	$n!$
A set of <u>any</u> k objects from a set of n	$\boxed{{}^nP_k = \frac{n!}{(n-k)!}}$
At set of <u>any</u> k objects from a set of n <u>with replacement</u>	n^k
A set of n objects, in which n_i are identical and of type i	$\frac{n!}{\prod(n_i!)}$

A set of <u>any</u> k object from a set of n, in which <u>order</u> in the resulting arrangements does not matter	$^nP_k / k! = \boxed{^nC_k = \frac{n!}{(n-k)!k!}}$
A set of n objects into m piles, each of which contain n_i items, and in which order matters not	$\frac{n!}{\prod(n_i!)}$
A set of k objects from a set of n objects <u>with replacement</u> and where <u>order matters not</u>	$^{n+k-1}C_k = \frac{(k+n-1)!}{k!(n-1)!}$

- **Random variables**

	Discrete	Continuous
Definition	**Probability function** $P(X = x) = \begin{cases} p_i & \text{if } x = x_i \\ 0 & \text{otherwise} \end{cases}$	**Probability density function** $P(x < X \leq x + \mathrm{d}x) = f(x)\mathrm{d}x$
Total	$\sum_{\forall i} f(x_i) = 1$	$\int_{-\infty}^{\infty} f(x)\ \mathrm{d}x = 1$
Cumulative Probability function (CPF); F(x)	$P(X \leq x) = \sum_{x_i \leq x} f(x_i)$	$P(X \leq x) = \int_{-\infty}^{x} f(x)\ \mathrm{d}x$
	It is clear, from this, that $f(x) = \frac{\mathrm{d}}{\mathrm{d}x} F(x)$	
P($a < X \leq b$)	$\sum_{a < x_i \leq b} f(x_i) = F(b) - F(a)$	$\int_a^b f(x)\ \mathrm{d}x = F(b) - F(a)$
Expected value of *g(X)*, *E[g(X)]*	$\sum_{\forall i} g(x_i) P(X = x_i)$	$\int_{-\infty}^{\infty} g(x) f(x)\ \mathrm{d}x$
Mean, μ or $\langle x \rangle$	*E(X)*	
Mode	The value of x at which *f(x)* is largest. There can be more than one mode, and the mode can be "not defined" [eg: uniform distribution]	
Median	The value of x at which $F(x) = \frac{1}{2}$	
Variance, $V(X)$	$V[X] = E\left[(X-\mu)^2\right] = E\left[X^2 - 2X\mu - \mu^2\right] = E\left[X^2\right] - \left(E\left[X\right]\right)^2$	

Useful properties of E and Var:

- $E(aX + b) = aE(X) + b$
- $Var(aX + b) = a^2 Var(X)$

- **The Binomial Distribution**

If we have n trials, probability of success is p and X is the number of times a success is obtained in these n trials, then $X \sim B(n, p)$ and

$$P(X = x) = {}^nC_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

and

$$P(X = x+1) = \frac{p}{1-p}\left(\frac{n-x}{x+1}\right)P(X = x)$$

Also, $E(X) = np$ and $Var(X) = np(1-p)$. [To prove these results, write out the sums and take out the value we know to be *E(X)* or *Var(X)*. Then, substitute to make it look like the distribution itself, the sum of which is 1.]

Notes:

- $\sigma \propto \sqrt{n}$. This has relevance to estimation – say we do not know *p*, but carry out *N* trials of which *M* are successful. Our best estimate for *p* is

  $$p_{est} = {}^{M}\!/_{N}$$

  Now, we know that the error on *M* is σ. Thus:

  $$Np_{actual} = M \pm \sqrt{Np(1-p)}$$

  $$p_{actual} = \frac{M}{N} \pm \sqrt{\frac{p(1-p)}{N}}$$

  $$p_{actual} \approx p_{est} \pm \frac{1}{\sqrt{N}} \overbrace{\sqrt{p_{est}(1-p_{est})}}^{\text{There's a small error here}}$$

  Therefore, we find that the **error** on a parameter estimate **improves** with **repeated trials** like $1/\sqrt{N}$. This is pretty much a **universal law** for **<u>all</u>** *large enough* probability distributions ($N \geq\sim 5$). For $N < 5$, this usually underestimates the error.
- As $n \to \infty$ but p stays fixed, the binomial distribution becomes **similar** to the **normal distribution**. In other words, if $X \sim B(n,p)$, then, for big enough *n* (about $N \gtrsim 10$, but sometimes even $N \gtrsim 5$ is OK), *X* can be estimated by *Y*, where $Y \sim N(np, np[p-1])$. This estimate becomes much worse in the tails of the distribution (ie: more than $\pm 3\sigma$ from the mean). In any case, a **continuity correction** must be applied.
- On the other hand, as $n \to \infty$ but <u>*np*</u> stays fixed (at, say, $np = \lambda$), then the binomial distribution becomes similar to the **poisson distribution**.

  The **Poisson distribution** is a **discrete distribution** for counting (integer) numbers of **discrete** events in a period of time, given that the **expected number of times** the event will happen in that period of time is λ. If $X \sim Po(\lambda)$, then

  $$P(X = r) = e^{-\lambda}\frac{\lambda^r}{r!}$$

  Where there is no upper limit on *r*. Also, $E(x) = \lambda$, and $Var(x) = \lambda$.

Note that when λ is large enough ($\lambda \geq 10$ is safer, but $\lambda > 5$ is usually OK), this distribution can be estimated by $N\left(\lambda,\left[\sqrt{\lambda}\right]^2\right)$. A **continuity correction** must also be applied here.

- **The Discrete Uniform Distribution** is one in which $f(x) = 1$ for $0 \leq x \leq 1$ and 0 elsewhere. In other words, there is an equal change of getting any value between 0 and 1. The mean of the distribution is $\frac{1}{2}$ and its standard deviation is $\frac{1}{\sqrt{12}}$.
- **The normal distribution** is defined by two parameters – its mean and its variance. The probability density function for this distribution is

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

If X is normally distributed with mean μ and variance σ^2, we write $X \sim N(\mu, \sigma^2)$.

If we have a set of independent readings, our best estimates for the mean and variance are:

$$\mu_{est} = \frac{\sum x_i}{n} \quad \text{and} \quad \sigma^2_{est} = \frac{\sum\left(x_i - \mu_{est}\right)}{n-1}$$

Important percentage points of the distribution are as follows:

- 68% chance of being **1 standard deviation** from the mean.
- 95% chance of being **2 standard deviations** from the mean.
- 99.7% chance of being **3 standard deviations** from the mean.