**KEY FACTS AND CONCEPTS**

**Discrete random variables**

* A **random variable** is a number whose value is determined by a process, the outcome of which is open to chance. Usually, capital letters such as $X$ are used to represent random variables.
* **Discrete random variables** may take only specific values. They are usually characterised by *counting*.
* **Continuous random variables** may take any value (often within set limits). They are usually characterised by *measuring*.
* A **probability distribution** associates a probability with each possible value of a discrete random variable. It can be represented using a table, column graph, or a probability distribution function.
* If $X$ is a random variable with possible values $\left\{x\_{1},x\_{2},x\_{3},…,x\_{n}\right\}$ and corresponding probabilities $\left\{p\_{1},p\_{2},p\_{3},…,p\_{n}\right\}$ then:
	+ Each probability must be between 0 and 1.
	$0\leq p\_{i}\leq 1$ for all $ i=1,…,n$
	+ The sum of all the probabilities must be 1.
	$$\sum\_{i=1}^{n}p\_{i}=p\_{1}+p\_{2}+p\_{3}+…+p\_{n}=1$$
* For a **uniform discrete random variable,** each possible value has the same probability of occurring
$p\_{i}=\frac{1}{n}$ for all $ i=1,…,n$.
* The **mode** of a discrete probability distribution is the most frequently occurring value of the variable. This is the data value $x\_{i}$ whose probability $p\_{i}$ is the highest.
* The **median** of the distribution corresponds to the 50th percentile. If the possible values $\left\{x\_{1},x\_{2},x\_{3},…,x\_{n}\right\}$ are listed in ascending order, the median is the value $x\_{j}$ when the cumulative sum $p\_{1}+p\_{2}+…+p\_{j}$ reaches 0.5.
* If there are $n$ trials of an experiment, and an event has probability $p$ of occurring in each of the trials, then the number of times we **expect** the event to occur is $np$.
* If $X$ is a random variable with possible values $\left\{x\_{1},x\_{2},x\_{3},…,x\_{n}\right\}$ and corresponding probabilities $\left\{p\_{1},p\_{2},p\_{3},…,p\_{n}\right\}$ then the **expected value** or **mean** is:
$$E\left(X\right)=μ=\sum\_{i=1}^{n}x\_{i}p\_{i}=x\_{1}p\_{1}+x\_{2}p\_{2}+…+x\_{n}p\_{n}$$
* The expected value can be interpreted as a long-run sample mean.
* The principle purpose of the expected value is to be a measurement of the *centre* of the distribution.
* If $X$ is a random variable with possible values $\left\{x\_{1},x\_{2},x\_{3},…,x\_{n}\right\}$ and corresponding probabilities $\left\{p\_{1},p\_{2},p\_{3},…,p\_{n}\right\}$ then the **standard deviation** is:
$$σ\left(X\right)=\sqrt{\sum\_{i=1}^{n}x\_{i}^{2}p\_{i}-μ^{2}}=\sqrt{x\_{1}^{2}p\_{1}+x\_{2}^{2}p\_{2}+…+x\_{n}^{2}p\_{n}-μ^{2}}$$
* The principle purpose of the standard deviation is to be a measurement of the *spread* of the distribution.
* For a random variable, $X$, and any constants $a$ and $b$:
	+ $E\left(aX+b\right)=aE\left(X\right)+b$
	+ $σ\left(aX+b\right)=\left|a\right|σ\left(X\right)$
* A **Bernoulli trial** is an experiment which only has two possible outcomes: “success” if some event occurs or “failure” if the event does not occur.
* A **Bernoulli random variable** $X$ is the number of successes in a single Bernoulli trial.

We define $p$ as the probability of success, so $P\left(X=1\right)=p$, and $P\left(X=0\right)=1-p$, where $0\leq p\leq 1$.

* + $E\left(X\right)=p$
	+ $σ\left(X\right)=\sqrt{p\left(1-p\right)}$
* A **binomial random variable** $X$ is the number of successes in $n$ identical independent Bernoulli trials with probability of success $p$.



The distribution of $X$ is called the binomial distribution and we write $X\~B\left(n,p\right)$.

* + The probability distribution of $X$ is
	$P\left(X=k\right)=\left(\genfrac{}{}{0pt}{}{n}{k}\right)p^{k}\left(1-p\right)^{n-k}$ where $k=0,1,…,n$
	+ $E\left(X\right)=np$
	+ $σ\left(X\right)=\sqrt{np\left(1-p\right)}$