1. Probabilities and Events:

A random experiment is an experiment whose outcome cannot be predicted in a deterministic fashion. Each random experiment has a set of possible outcomes that form the sample space of the experiment. For instance, the outcome of rolling a dice on a table can be any integer value between 1 and 6.

An event $E$ is an arbitrary subset of this sample space, like obtaining an even value after rolling a dice. Since events are sets, the usual set operations apply to them.

We can associate with each event $A$, a probability $P(A)$ measuring its relative likelihood. This probability should obey to the three following axioms:

1. For any event $A$, $P(A) \geq 0$.
2. If $S$ denotes the sample space of an experiment then $P(S) = 1$.
3. If events $A$ and $B$ are disjoint ($A \cap B = \emptyset$) then $P(A \cup B) = P(A) + P(B)$

From these axioms one can show that $P(\emptyset) = 0$ and $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

The conditional probability $P(A|B)$ is the probability of the occurrence of an event $A$ given that event $B$ has already occurred. It is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

As a result, we have $P(A \cap B) = P(A|B)P(B)$. Two events are said to be independent if

$$P(A \cap B) = P(A)P(B)$$

2. Random Variables

A random variable (rv) $X$ over a sample space $S$ is a function that assigns a real number $X(S)$ to any event $E$ in $S$. A discrete random variable is a random variable whose values are always integer.

All random variables have a cumulative distribution function (cdf) defined as

$$F(z) = P(X \leq z)$$

Discrete random variables

Assume that a discrete rv $X$ can have $n$ possible values $\{x_1, \ldots, x_n\}$. To each of these values we can associate a probability $p_i$. These $p_i$'s define the probability mass function (pmf) of $X$ and will always verify the relation

$$\sum_{i=1}^{n} p_i = 1$$
The cdf of \( X \) is then given by:

\[
F(z) = \sum_{x_i \leq z} p_i ,
\]

its mean by

\[
\mu = E(X) = \sum_{i=1}^n p_i x_i ,
\]

and its standard deviation by

\[
\sigma^2 = E(X - \mu)^2 = \sum_{i=1}^n p_i (x_i - \mu)^2 .
\]

The variance measures the degree of dispersion of the rv. A rv with a zero variance would have its mean as sole possible value. The standard deviation \( \sigma \) is the square root of the variance.

**Discrete random variables**

Continuous random variables have a *probability density function* (pdf) \( f(x) \) such that

\[
f(x) = \lim_{\Delta x \to 0} P(x < X \leq x + \Delta x)
\]

We will always have

\[
\int_{-\infty}^{\infty} f(x)dx = 1
\]

The cdf of a continuous rv is the integral of its pdf

\[
F(z) = \int_{-\infty}^{z} f(x)dx
\]

The mean of a continuous rv is given by

\[
\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx
\]

and its variance by

\[
\sigma^2 = E(X - \mu)^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx
\]

**Covariance and Correlation Coefficient**

If two variables \( X \) and \( Y \), with respective means \( \mu_X \) and \( \mu_Y \), are not independent, their covariance

\[
\text{cov}(X,Y) = E(X - \mu_X)(Y - \mu_Y)
\]
will be different from zero. While variances are always positive, covariances can have negative values. The correlation coefficient \( \rho_{XY} \) is often used to measure the strength of a possible linear dependence between two r.v's:

\[
\rho_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}
\]

A coefficient of correlation equal to one indicates a perfect linear dependence between the two r.v's. Conversely a coefficient of correlation equal to zero indicates the absence of any linear dependence between the two r.v's but does not guarantee that the two r.v's are independent.

### 3. Sample Statistics

If \( x_1, x_2, \ldots, x_n \) are \( n \) observations of the value of an unknown quantity \( X \), they constitute a sample of size \( n \) for the population on which \( X \) is defined. This sample will have

1. a sample mean \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \)

2. a sample variance \( s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \)

### 4. The Central Limit Theorem

If the \( n \) mutually independent random variables \( x_1, x_2, \ldots, x_n \) have the same distribution, and if \( \mu = E(x_i) \) and \( \sigma^2 = E((x_i - \mu)^2) \) exist, then the r.v

\[
\frac{1}{\sigma/\sqrt{n}} \sum_{i=1}^{n} x_i - \mu
\]

is distributed according to the standard normal distribution (zero mean and unit variance).

### 5. Estimation

**Estimating a mean**

Assume that we have a sample \( x_1, x_2, \ldots, x_n \) consisting of \( n \) independent observations of a given population. The sample mean \( \bar{x} \) is an unbiased estimator of the mean \( \mu \) of the population.

For large values of \( n \), the (1-\( \alpha \))% confidence interval for \( \mu \) is given by

\[
\left[ \bar{x} - \frac{z_{\alpha/2} \sigma}{\sqrt{n}}, \bar{x} + \frac{z_{\alpha/2} \sigma}{\sqrt{n}} \right]
\]

where \( z_{\alpha/2} \) satisfies \( F(u_{\alpha/2}) = 1 - \frac{\alpha}{2} \) for the standard normal distribution. For \( \alpha = .05 \), \( z_{0.025} = 1.96 \).

---

1. This is the critical assumption. Without it, you cannot apply the formula.
In nearly all cases, the population variance $\sigma^2$ is unknown. We can construct a similar confidence interval by replacing the unknown $\sigma$ in the preceding formula by the standard-deviation $s$ of the sample but the value $z_{\alpha/2}$ must now be read from a table of Student's $t$-distribution with $n-1$ degrees of freedom whenever $x < 30$.

**Estimating a proportion**

Mean response times are a poor estimator of customer satisfaction. Quantiles and proportions are a much better index of the actual performance of the system. A $\alpha$ quantile represents the value $x$ such that $P(X > x) = 1 - \alpha$ as in “95% of the customers have to wait less than 50 seconds for the video of their choice.” A proportion $p$ represents the probability $P(X \leq \theta)$ for some fixed threshold $\theta$ as in “97% of our customers have to wait less than a minute.”

The main advantage of proportions over quantiles is that their confidence intervals are much easier to obtain. Assume that we have $n$ independent observations $x_1, x_2, \ldots, x_n$ of a given population variable $X$ and that this variable has a continuous distribution. Let $p$ represent the proportion we want to estimate, say $P(X \leq \theta)$, and $k$ represent the number of observations that are $\leq \theta$. The rv $k$ is distributed according to a binomial distribution

$$P(k \text{ out of } n) = \binom{n}{k} p^k (1-p)^{n-k}$$

with mean $np$ and variance $np(1-np)^{n-k}$.

The sample random variable $\hat{p} = \frac{k}{n}$ has a mean equal to $p$ and a variance equal to $\frac{p(1-p)}{n}$. For $n \geq 30$, the distribution of $\hat{p}$ is approximately normal and we have

$$P\left[ \hat{p} - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} < p \leq \hat{p} + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right] = 1 - \alpha$$