Independent events

**Definition** Events $A$, $B$ are independent if $P(A \text{ and } B) = P(A)P(B)$

**Example 9.2** Two fair dice tossed independently

Die 1, Die 2 dice can be distinguished

Let $(i,j)$ denote the outcome: $i$ dots on die 1, $j$ dots on die 2

$P(\{(i,j)\}) = \frac{1}{6} \cdot \frac{1}{6}$

E.g. $P(i+j=11) = P(\{(5,6)\} \text{ or } \{(6,5)\}) = \frac{1}{36} + \frac{1}{36} = \frac{1}{18}$

**Example 9.3** **Genetic model (Hardy-Weinberg)**

Gene - region of chromosome (including $AB$) that corresponds to some physical characteristic (e.g. flower colour).

Allele - form or type of gene (e.g. allele for red flowers, allele for white flowers).

Diploid - organisms that have paired chromosomes and therefore two alleles for gene position in chromosome. To produce offspring, for each gene, each parent contributes one copy of one of their two alleles.

Suppose we have a large population of diploid organisms, and two alleles $A, B$ for a particular gene. Each individual will be one of three genotypes $AA$, $AB$, $BB$:

Let $x, y, z$ respectively be the frequency of the genotypes $AA, AB, BB$

$x = \text{frequency of } AA \text{ genotype} = \frac{\text{no. of } AA \text{ individuals}}{\text{total no. of individuals}} = \text{prob. of randomly choosing an } AA \text{ individual}$

$y, z$ similarly defined

$x + y + z = 1$

Let $p, q$ respectively be the frequency of the alleles $A, B$ in the gene pool:

E.g. 2 individuals $AA \ AB \rightarrow 4$ alleles in gene pool $A \ A \ A \ B$

$p = \text{frequency of } A \text{ allele} = \frac{\text{no. of } A \text{ alleles}}{\text{total no. of alleles}} = \frac{2 \times \text{no. of } AA \text{ individuals} + \text{no. of } AB \text{ individuals}}{2 \times \text{total no. of individuals}}$

$= \frac{\text{no. of } AA \text{ individuals}}{\text{total no. of individuals}} + \frac{1}{2} \times \frac{\text{no. of } AB \text{ individuals}}{\text{total no. of individuals}}$

$= x + \frac{1}{2} y$

$q = \text{frequency of } B \text{ allele} = 2 + \frac{1}{2} y$

$p \times q = 1$

Consider a species of diploid organisms with non-overlapping generations (e.g. annual plants, some insects, some fish) equal no. M, F (in gene pool)

Suppose we start with a large population of individuals (parents) with freq. $p, q$ of $A, B$ alleles.
Assume
1. sex ratio is indep. of genotype
2. survival is random
3. fertility is indep. of genotype
4. survival to adulthood is indep. of genotype
5. no mutation or migration

What are $p_0, q_0$ in subsequent generations $n=1, 2, 3, \ldots$?

Females

\[
\begin{array}{c|cc}
\text{Sex} & A & B \\
\hline
\text{F} & p_0 & q_0 \\
\end{array}
\]

\[
\begin{array}{c}
\text{Next generation.} \quad \text{Females AA has probability } P(\text{get A allele from female parent}) \cdot P(\text{get A allele from male parent}) \cdot x_1 = p_0 \cdot p_0 \\
\end{array}
\]

\[
\begin{array}{c}
\text{Females BB has probability } P(\text{get B allele from female parent}) \cdot P(\text{get B allele from male parent}) \cdot x_1 = q_0 \cdot q_0 \\
\end{array}
\]

\[
\begin{array}{c}
\text{Females AB has probability } P(\text{get A allele from female parent}) \cdot P(\text{get B allele from male parent}) \cdot 2 \cdot p_0 \cdot q_0 = 2 \cdot p_0 \cdot q_0 \\
\end{array}
\]

Freq. $\text{A} \equiv p_0$: allele in next generation is

\[
\begin{array}{c}
p_1 = x_1 + \frac{1}{2} y_1 = p_0 + \frac{1}{2} \cdot p_0 \cdot q_0 = \frac{3}{4} \cdot p_0 = \frac{3}{4} \\
\end{array}
\]

unchanged!

For all $n$, $p_{n+1} = p_n + \frac{1}{2} \cdot p_n \cdot q_n = p_n$, $q_{n+1} = q_n$

This is called the Hardy-Weinberg law: under the assumptions above, the allele frequency remains constant from generation to generation.

### Random Variables

A (discrete) random variable is a function where domain is a (discrete) sample space $\Omega$. Given a random variable $X: \Omega \rightarrow \mathbb{R}$, we can ask about its probability distribution functions:

**Example 4.4**

**Tossing a die (fair or not)**

$\Omega = \{1, 2, 3, 4, 5, 6\}$

A rule $X: \Omega \rightarrow \mathbb{R}$ could be $X(\text{1}) = 1$, $X(\text{6}) = 6$, $X = \text{no. of dots on top face}$

**Example 4.5**

**Tossing two dice**

$\Omega = \{(1,1), (1,2), (1,3), \ldots, (6,6)\}$ (36 outcomes)

A rule $X: \Omega \rightarrow \mathbb{R}$ could be $X((i,j)) = i + j$, $X = \text{sum of dots on top faces of die}$

For example,

\[
\begin{align*}
P(X = 11) &= P((5,6) \text{ or } (6,5)) = \frac{1}{36} + \frac{1}{36} = \frac{1}{18} \\
P(X \geq 11) &= P((5,6) \text{ or } (6,5) \text{ or } (6,6)) = \frac{3}{36} = \frac{1}{12}
\end{align*}
\]

Possible values of $X$ are $2, 3, \ldots, 12$.

Suppose $X$ is a random variable with possible values $x_1, x_2, \ldots, x_m$.

The probability function associated with $X$ is

\[
f_i = P(X = x_i) \quad i = 1, \ldots, m
\]

Probability function must satisfy

\[
\sum_{i=1}^{m} f_i = 1
\]
Example 4.5 cont.

Two fair dice, \( X((x_1, x_2)) = \min x \)

11 possible outcomes of \( X \) are: \( x_1 = 2, x_2 = 3, \ldots, x_{11} = 12 \)

\[ x_1: \{X = x_1\} \quad f_i = P(X = x_i) \]

\[ \begin{array}{c|c|c}
2 & \{(1,1)\} & \frac{1}{36} \\
3 & \{(1,2),(2,1)\} & \frac{5}{36} \\
4 & \{(1,3),(2,2),(3,1)\} & \frac{3}{36} \\
& \vdots & \\
11 & \{(5,6),(6,5)\} & \frac{2}{36} \\
12 & \{(6,6)\} & \frac{1}{36} \\
\end{array} \]

The mean of a random variable

The **mean or expected value** of \( X \) is

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

Interpretation: for a large number of experiments, the average value of \( X \)

per experiment \[ \frac{\text{total of } N \text{ values of } X}{X} \]

is expected to be \( \overline{X} \) (\( = \overline{X} \) in this context).