1.8 Dissimilarities, Metrics, and Ultrametrics

If $S = (T, H, c, D)$ is a decision system, the blocks of the quotient set $T/\text{ind}_D$ are called the decision classes of $S$; the number $|T/\text{ind}_D|$ is the rank of $S$.

The indiscernability relation can be extended to sets of attributes by defining:

$$\text{ind}_L = \bigcap \{\text{ind}_A \mid A \in L\},$$

for every $L \subseteq H$. It is clear that we have $\text{ind}_L \cap \text{ind}_K = \text{ind}_{K \cup L}$ for every $K, L \subseteq H$.

1.8 Dissimilarities, Metrics, and Ultrametrics

Definition 1.8.1 A dissimilarity on a set $S$ is a function $d : S^2 \rightarrow \mathbb{R}_{\geq 0}$ satisfying the following conditions:

(i) $d(x, x) = 0$ for all $x \in S$;
(ii) $d(x, y) = d(y, x)$ for all $x, y \in S$.

The pair $(S, d)$ is a dissimilarity space.

The set of dissimilarities defined on a set $S$ is denoted by $\mathcal{D}_S$.

Additional properties may be satisfied by dissimilarities. A non-exhaustive list include the following:

1. $d(x, y) = 0$ implies $d(x, z) = d(y, z)$ for every $x, y, z \in S$ (evenness);
2. $d(x, y) = 0$ implies $x = y$ for every $x, y$ (definiteness);
3. $d(x, y) \leq d(x, z) + d(z, y)$ for every $x, y, z$ (triangular inequality);
4. $d(x, y) \leq \max\{d(x, z), d(z, y)\}$ for every $x, y, z$ (the ultrametric inequality);
5. $d(x, y) + d(u, v) \leq \max\{d(x, u) + d(y, v), d(x, v) + d(y, u)\}$ for every $x, y, u, v$ (Buneman’s inequality).

Example 1.8.2 Consider the mapping $d : (\text{Seq}_n(S))^2 \rightarrow \mathbb{R}_{\geq 0}$ defined by

$$d(p, q) = |\{i \mid 0 \leq i \leq n - 1 \text{ and } p(i) \neq q(i)\}|,$$

for every sequences $p, q$ of length $n$ on the set $S$.

Clearly, $d$ is a dissimilarity that is both even and definite. Moreover, it satisfies the triangular inequality. Indeed, let $p, q, r$ be three sequences of length $n$ on the set $S$. If $p(i) \neq q(i)$, then $r(i)$ must be distinct from at least one of $p(i)$ and $q(i)$. Therefore,

$$\{i \mid 0 \leq i \leq n - 1 \text{ and } p(i) \neq q(i)\} \subseteq \{i \mid 0 \leq i \leq n - 1 \text{ and } p(i) \neq r(i)\} \cup \{i \mid 0 \leq i \leq n - 1 \text{ and } r(i) \neq q(i)\},$$

which implies the triangular inequality.

Theorem 1.8.3 Both the ultrametric inequality and Buneman’s inequality imply the triangular inequality.

Proof. Suppose that $d \in \mathcal{D}_S$ satisfies the ultrametric inequality. Then, $d(x, y) \leq \max\{d(x, z), d(z, y)\} \leq d(x, z) + d(z, y)$ for every $x, y, z \in S$, which shows that $d$ also satisfies the triangular inequality.
Now, let $d \in \mathcal{D}_S$ be a dissimilarity that satisfies Buneman’s inequality $d(x, y) + d(u, v) \leq \max\{d(x, u) + d(y, v), d(x, v) + d(y, u)\}$ for every $x, y, u, v$. Choosing $v = u$ we obtain $d(x, y) \leq \max\{d(x, u) + d(y, u), d(x, u) + d(y, u)\} = d(x, u) + d(u, y)$ for every $x, y, u$ which shows that $d$ satisfies the triangular inequality.

**Theorem 1.8.4** Both the triangular inequality and definiteness imply evenness.

**Proof.** Suppose that $d$ is a dissimilarity that satisfies the triangular inequality and let $x, y \in S$ be such that $d(x, y) = 0$. By the triangular inequality we have both $d(x, z) \leq d(x, y) + d(y, z) = d(y, z)$ and $d(y, z) \leq d(y, x) + d(x, z) = d(x, z)$ because $d(y, x) = d(x, y) = 0$. Thus, $d(x, z) = d(y, z)$ for every $z \in S$.

We leave to the reader to prove the second part of the statement.

**Definition 1.8.5** A dissimilarity $d \in \mathcal{D}_S$ is

1. a metric, if it satisfies the definiteness property and the triangular inequality;
2. an ultrametric, if it satisfies the definiteness property and the ultrametric inequality;
3. a tree distance, if it satisfies the definiteness property and Buneman’s inequality.

The set of metrics on a set $S$ is denoted by $\mathcal{M}_S$. The sets of ultrametrics and tree metrics on a set $S$ are denoted by $\mathcal{U}_S$ and $\mathcal{T}_S$, respectively.

If $d$ is a metric or an ultrametric on a set $S$, then $(S, d)$ is a metric space or an ultrametric space, respectively.

Theorem 1.8.3 implies $\mathcal{U}_S \subseteq \mathcal{M}_S \subseteq \mathcal{D}_S$ and $\mathcal{T}_S \subseteq \mathcal{M}_S$.

**Example 1.8.6** Let $\mathcal{G} = (V, E)$ be a connected graph. Define the mapping $d : V^2 \rightarrow \mathbb{R}_{\geq 0}$ by $d(x, y) = m$, where $m$ is the length of the shortest path that connects $x$ and $y$. Then, $d$ is a metric.

Indeed, we have $d(x, y) = 0$ if and only if $x = y$. The symmetry of $d$ is obvious.

If $p$ is a shortest path that connects $x$ to $z$ and $q$ is a shortest path that connects $z$ to $y$, then $pq$ is a path of length $d(x, z) + d(z, y)$ that connects $x$ to $y$. Therefore, $d(x, y) \leq d(x, z) + d(z, y)$.

**Example 1.8.7** The term “tree metric” comes from the fact that every tree $\mathcal{T} = (V, E)$ generates a tree metric on its set of vertices. Since $\mathcal{T}$ is connected, for any two vertices $u, v \in V$ there is a unique simple path joining $u$ to $v$. Define $d(u, v)$ as being the length of this path. We already know from Example 1.8.6 that $d$ is a metric. We shall prove that $d$ satisfies Buneman’s inequality.

Let $x, y, u, v$ be four vertices in $\mathcal{T}$. If $x = u$ and $y = v$, the inequality reduces to an obvious equality. Therefore, we may assume that at least one of the pairs $(x, u)$ and $(y, v)$ consists of distinct vertices.

Suppose that $x = u$. In this case, the inequality reduces to:

$$d(x, y) + d(x, v) \leq \max\{d(y, v), d(x, v) + d(y, x)\},$$

which is obviously satisfied. Thus, we may assume that we have both $x \neq u$ and $y \neq v$. Since $\mathcal{T}$ is a connected graph there exists a simple path $p$ that joins $x$
1.8 Dissimilarities, Metrics, and Ultrametrics

Let $1.8.1$ The Poset of Dissimilarities

It is easy to verify that $(x, y)$ for $x, y, z \in S$ be a dissimilarity on a set $S$ and let $U_d$ be the set of ultrametrics:

$$ U_d = \{ e \in \mathcal{U}_S \mid e \leq d \}. $$

The set $U_d$ has a largest element in the poset $(D_S, \leq)$.  

**Proof.** Note that the set $U_d$ is non-empty because the zero dissimilarity $d_0$ given by $d_0(x, y) = 0$ for every $x, y \in S$ is an ultrametric and $d_0 \leq d$.

Since the set $\{ e(x, y) \mid e \in U_d \}$ has $d(x, y)$ as an upper bound it is possible to define the mapping $e_1 : S^2 \to \mathbb{R}_{\geq 0}$ as:

$$ e_1(x, y) = \sup \{ e(x, y) \mid e \in U_d \} $$

for $x, y \in S$. It is clear that $e \leq e_1$ for every ultrametric $e$. We claim that $e_1$ is an ultrametric on $S$.

We prove only that $e_1$ satisfies the ultrametric inequality. Suppose that there exist $x, y, z \in S$ such that $e_1$ violates the ultrametric inequality, that is:

$$ \max \{ e_1(x, z), e_1(z, y) \} < e_1(x, y). $$

This is equivalent to

$$ \sup \{ e(x, y) \mid e \in U_d \} > \max \{ \sup \{ e(x, z) \mid e \in U_d \}, \sup \{ e(z, y) \mid e \in U_d \} \}. $$

Two cases may occur, depending on whether $p$ and $q$ have common edges.

Suppose initially that there are no common vertices between $p$ and $q$. By Theorem 1.4.12, there is a unique vertex $s$ in $p$, a unique vertex $t$ in $q$, and a simple path $r$ that joins $s$ to $t$ such that $r$ has no other vertices in common with $p$ and $q$ except $s$ and $t$, respectively, $d(s, t) = \ell(r)$ and $d(x, u) = d(x, s) + d(s, t) + d(t, u)$, $d(y, v) = d(y, s) + d(s, t) + d(t, v)$, $d(x, v) = d(x, s) + d(s, t) + d(t, v)$, $d(u, v) = d(y, s) + d(s, t) + d(t, u)$.

Thus, $d(x, u) + d(y, v) = d(x, v) + d(y, u) = d(x, s) + d(s, t) + d(t, u) + d(y, s) + d(s, t) + d(t, v) = d(x, y) + d(u, v) + 2d(s, t)$, which shows that Buneman's inequality is satisfied. ∎

1.8.1 The Poset of Dissimilarities

Let $S$ be a set. Recall that we denoted the set of dissimilarities by $D_S$. The partial order $\leq$ on $D_S$ is given by $d \leq d'$ if $d(x, y) \leq d'(x, y)$ for every $x, y \in S$. It is easy to verify that $(D_S, \leq)$ is a poset.

Note that $\mathcal{U}_S$, the set of ultrametrics on $S$ is a subset of $D_S$.

**Theorem 1.8.8** Let $d$ be a dissimilarity on a set $S$ and let $U_d$ be the set of ultrametrics:

$$ U_d = \{ e \in \mathcal{U}_S \mid e \leq d \}. $$

The set $U_d$ has a largest element in the poset $(D_S, \leq)$.
Thus, there exists \( \hat{e} \in U_d \) such that
\[
\hat{e}(x, y) > \sup \{ e(x, z) \mid e \in U_d \},
\]
\[
\hat{e}(x, y) > \sup \{ e(z, y) \mid e \in U_d \}.
\]
In particular, \( \hat{e}(x, y) > \hat{e}(x, z) \) and \( \hat{e}(x, y) > \hat{e}(z, y) \), which contradicts the fact that \( \hat{e} \) is an ultrametric.

The ultrametric defined by Theorem 1.8.8 is known as the maximal subdominant ultrametric for the dissimilarity \( d \).

1.8.2 Metrics on \( \mathbb{R}^n \)

**Lemma 1.8.9** Let \( p, q \in \mathbb{R} \) be two numbers such that \( \frac{1}{p} + \frac{1}{q} = 1 \) and \( p > 1 \). Then, for every \( a, b \in \mathbb{R}_{\geq 0} \) we have
\[
ab \leq \frac{a^p}{p} + \frac{b^q}{q},
\]
where the equality holds if and only if \( a = b^{\frac{1}{p}} \).

**Proof.** Note that \( q = \frac{p}{p-1} \), by the hypothesis of the lemma. Consider the function \( f(x) = \frac{a^p}{p} + \frac{1}{q} - x \) for \( x \geq 0 \). We have \( f'(x) = x^{p-1} - 1 \), so the minimum is achieved when \( x = 1 \) and \( f(1) = 0 \). By taking \( x = ab^{-\frac{1}{p}} \) we obtain \( f(x) = a^{\frac{p-1}{p}} + \frac{1}{q} - ab^{-\frac{1}{p}} \geq 0 \). Multiplying both sides of this inequality by \( b^{\frac{p-1}{p}} \) gives the desired inequality.

**Theorem 1.8.10 (The Hölder Inequality)** Let \( a_0, \ldots, a_{n-1}, b_0, \ldots, b_{n-1} \) be 2\( n \) non-negative numbers, and let \( p, q \) be two numbers such that \( \frac{1}{p} + \frac{1}{q} = 1 \) and \( p > 1 \). We have:
\[
\left( \sum_{i=0}^{n-1} a_i^p \right)^{\frac{1}{p}} \cdot \left( \sum_{i=0}^{n-1} b_i^q \right)^{\frac{1}{q}} \geq \sum_{i=0}^{n-1} a_i b_i.
\]

**Proof.** Define the numbers:
\[
x_i = \frac{a_i}{\left( \sum_{i=0}^{n-1} a_i^p \right)^{\frac{1}{p}}} \quad \text{and} \quad y_i = \frac{b_i}{\left( \sum_{i=0}^{n-1} b_i^q \right)^{\frac{1}{q}}},
\]
for \( 0 \leq i \leq n-1 \). The Lemma 1.8.9 applied to \( x_i, y_i \) yields:
\[
\frac{a_i b_i}{\left( \sum_{i=0}^{n-1} a_i^p \right)^{\frac{1}{p}} \left( \sum_{i=0}^{n-1} b_i^q \right)^{\frac{1}{q}}} \leq \frac{1}{p} \sum_{i=0}^{n-1} a_i^p + \frac{1}{q} \sum_{i=0}^{n-1} b_i^q.
\]
Adding these inequalities we obtain:
\[
\sum_{i=0}^{n-1} a_i b_i \leq \left( \sum_{i=0}^{n-1} a_i^p \right)^{\frac{1}{p}} \left( \sum_{i=0}^{n-1} b_i^q \right)^{\frac{1}{q}},
\]
1.8 Dissimilarities, Metrics, and Ultrametrics

because $\frac{1}{p} + \frac{1}{q} = 1$.

The non-negativity of the numbers $a_0, \ldots, a_{n-1}, b_0, \ldots, b_{n-1}$ can be relaxed by using absolute values. Indeed, we can easily prove the following variant of Theorem 1.8.10:

**Theorem 1.8.11** Let $a_0, \ldots, a_{n-1}, b_0, \ldots, b_{n-1}$ be $2n$ numbers, and let $p, q$ be two numbers such that $\frac{1}{p} + \frac{1}{q} = 1$ and $p > 1$. We have:

$$\left(\sum_{i=0}^{n-1} |a_i|^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i=0}^{n-1} |b_i|^q\right)^{\frac{1}{q}} \geq \left|\sum_{i=0}^{n-1} a_i b_i\right|.$$

**Proof.** By Theorem 1.8.10 we have

$$\left(\sum_{i=0}^{n-1} |a_i|^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i=0}^{n-1} |b_i|^q\right)^{\frac{1}{q}} \geq \sum_{i=0}^{n-1} |a_i||b_i|.$$

The needed equality follows from the fact that

$$\sum_{i=0}^{n-1} |a_i||b_i| \geq \left|\sum_{i=0}^{n-1} a_i b_i\right|.$$

**Corollary 1.8.12 (The Cauchy Inequality)** Let $a_0, \ldots, a_{n-1}, b_0, \ldots, b_{n-1}$ be $2n$ numbers. We have:

$$\sqrt{\sum_{i=0}^{n-1} |a_i|^2} \cdot \sqrt{\sum_{i=0}^{n-1} |b_i|^2} \geq \left|\sum_{i=0}^{n-1} a_i b_i\right|.$$

**Proof.** The inequality follows immediately from Theorem 1.8.11 by taking $p = q = \frac{1}{2}$.

**Theorem 1.8.13 (The Minkowski Inequality)** Let $a_0, \ldots, a_{n-1}, b_0, \ldots, b_{n-1}$ be $2n$ non-negative numbers. If $p \geq 1$ we have

$$\left(\sum_{i=0}^{n-1} a_i^p\right)^{\frac{1}{p}} + \left(\sum_{i=0}^{n-1} b_i^p\right)^{\frac{1}{p}} \geq \left(\sum_{i=0}^{n-1} (a_i + b_i)^p\right)^{\frac{1}{p}}.$$

If $p < 1$ the inequality sign is reversed.

**Proof.** Note that:

$$\left(\sum_{i=0}^{n-1} (a_i + b_i)^p\right)^{\frac{1}{p}} = \left(\sum_{i=0}^{n-1} a_i(a_i + b_i)^{p-1}\right)^{\frac{1}{p}} + \left(\sum_{i=0}^{n-1} b_i(a_i + b_i)^{p-1}\right)^{\frac{1}{p}}.$$
Then, by Hölder’s Inequality for $p, q$ such that $\frac{1}{p} + \frac{1}{q} = 1$ we have:
\[
\sum_{i=0}^{n-1} a_i (a_i + b_i)^{p-1} \leq \left( \sum_{i=0}^{n-1} a_i^p \right)^{\frac{1}{p}} \left( \sum_{i=0}^{n-1} (a_i + b_i)^{(p-1)q} \right)^{\frac{1}{q}}
\]
\[
= \left( \sum_{i=0}^{n-1} a_i^p \right)^{\frac{1}{p}} \left( \sum_{i=0}^{n-1} (a_i + b_i)^p \right)^{\frac{1}{q}}.
\]
Similarly, we can write:
\[
\sum_{i=0}^{n-1} b_i (a_i + b_i)^{p-1} \leq \left( \sum_{i=0}^{n-1} b_i^p \right)^{\frac{1}{p}} \left( \sum_{i=0}^{n-1} (a_i + b_i)^p \right)^{\frac{1}{q}}.
\]
Adding the last two inequalities yields:
\[
\sum_{i=0}^{n-1} (a_i + b_i)^p \leq \left( \sum_{i=0}^{n-1} a_i^p \right)^{\frac{1}{p}} + \left( \sum_{i=0}^{n-1} b_i^p \right)^{\frac{1}{q}}
\]
which is equivalent to the desired inequality:
\[
\left( \sum_{i=0}^{n-1} (a_i + b_i)^p \right)^{\frac{1}{p}} \leq \left( \sum_{i=0}^{n-1} a_i^p \right)^{\frac{1}{p}} + \left( \sum_{i=0}^{n-1} b_i^p \right)^{\frac{1}{q}}.
\]

Let $x = (x_0, \ldots, x_{n-1}), y = (y_0, \ldots, y_{n-1}) \in \mathbb{R}^n$. Define the mapping $d_p : \mathbb{R}^n \longrightarrow \mathbb{R}_{\geq 0}$ by:
\[
d_p(x, y) = \left( \sum_{i=0}^{n-1} |x_i - y_i|^p \right)^{\frac{1}{p}}.
\]

**Theorem 1.8.14** The mapping $d_p$ defined above is a metric on $\mathbb{R}^n$.

**Proof.** We limit the argument to the proof of the triangular inequality. In other words, we shall prove that
\[
d_p(x, y) \leq d_p(x, z) + d_p(z, y).
\]
We apply the Minkowski inequality to the numbers $a_i = |x_i - z_i|$ and $b_i = |z_i - y_i|$ for $0 \leq i \leq n$:
\[
\left( \sum_{i=0}^{n-1} |x_i - z_i|^p \right)^{\frac{1}{p}} + \left( \sum_{i=0}^{n-1} |z_i - y_i|^p \right)^{\frac{1}{p}}
\]
\[
\geq \left( \sum_{i=0}^{n-1} (|x_i - y_i| + |y_i - z_i|)^p \right)^{\frac{1}{p}}
\]
\[
\geq \left( \sum_{i=0}^{n-1} |x_i - z_i|^p \right)^{\frac{1}{p}},
\]

1.8 Dissimilarities, Metrics, and Ultrametrics

The metric introduced in Theorem 1.8.14 is known as Minkowski metric on $\mathbb{R}^n$. If $p = 2$ we have the Euclidean metric on $\mathbb{R}^n$:

$$d_2(x, y) = \sqrt{\sum_{i=0}^{n-1} (x_i - y_i)^2}.$$ 

If $p = 1$ we have

$$d_1(x, y) = \sum_{i=0}^{n-1} |x_i - y_i|.$$ 

For the special case of $\mathbb{R}^2$ we represented these distances in Figure ?? between the points $x = (x_0, x_1)$ and $y = (y_0, y_1)$. The distance $d_2(x, y)$ is the length of the hypotenuse of the right triangle; the distance $d_1(x, y)$ is the sum of the lengths of the two legs of the triangle.

Consider the a sphere $B_p(0, 1)$ in the metric space $(\mathbb{R}^n, d_p)$ for several values of $p$. For $p = 2$ this sphere is a circle centered in the origin (shown in Figure 1.9(a)), whereas for $p = 1$ the sphere is actually a square (shown in Figure 1.9(b)).

A special metric on $\mathbb{R}^n$ is the function $d_\infty : (\mathbb{R}^n)^2 \rightarrow \mathbb{R}_{\geq 0}$ given by

$$d_\infty(x, y) = \max\{|x_i - y_i| \mid 0 \leq i \leq n\}$$

for $x = (x_0, \ldots, x_{n-1}), y = (y_0, \ldots, y_{n-1}) \in \mathbb{R}^n$.

To prove that $d_\infty$ satisfies the triangular axiom we start from the inequality:

$$|x_i - y_i| \leq |x_i - z_i| + |z_i - y_i| \leq d_\infty(x, z) + d_\infty(z, y),$$

for every $i$, $0 \leq i \leq n$. This, in turn, implies

$$d_\infty(x, y) = \max_{i \in 0 \leq i \leq n-1} |x_i - y_i| \leq d_\infty(x, z) + d_\infty(z, y),$$
which gives the desired inequality.

This metric can be regarded as a special case of the metric $d_p$. Indeed, we can write:

$$\lim_{p \to \infty} d_p(x, y) = \lim_{p \to \infty} \max_{0 \leq i \leq n-1} |x_i - y_i| \left( \sum_{i=1}^{n-1} \left( \frac{|x_i - y_i|}{\max_{0 \leq i \leq n-1} |x_i - y_i|} \right)^p \right)^{\frac{1}{p}}$$

which justifies the notation $d_\infty$.

### 1.8.3 Metrics on the collection of subsets of a finite set

A very simple metric on $P(S)$, the set of subsets of a finite set $S$ is given in the next theorem.

**Theorem 1.8.15** Let $S$ be a finite set. The mapping $\delta : (P(S))^2 \rightarrow \mathbb{R}_{\geq 0}$ defined by $\delta(X, Y) = |X \oplus Y|$ is a metric on $P(S)$.

**Proof.** The function $\delta$ is clearly symmetric and we have $\delta(X, Y) = 0$ if and only if $X = Y$. Therefore, we need to prove only the triangular inequality

$$|X \oplus Y| \leq |X \oplus Z| + |Z \oplus Y|,$$

for every $X, Y, Z \in P(S)$.

Note that $X \oplus Y = (X \oplus Z) \oplus (Z \oplus Y)$. Therefore, $|X \oplus Y| \leq |X \oplus Z| + |Z \oplus Y|$, which is precisely the triangular inequality for $\delta$.

Note that for $U, V \in P(S)$ we have $0 \leq \delta(U, V) \leq |S|$, where $\delta(U, V) = |S|$ if and only if $V = S - U$.

**Lemma 1.8.16** Let $d : S \times S \rightarrow \mathbb{R}_{\geq 0}$ be a distance and let $u \in S$ be an element of the set $S$. Define the mapping $d_u : S \times S \rightarrow \mathbb{R}_{\geq 0}$ by:

$$d_u(x, y) = \begin{cases} 0 & \text{if } x = y = u \\ \frac{d(x, y)}{d(x, y) + d(y, u) + d(u, y)} & \text{otherwise.} \end{cases}$$

Then $d_u$ is a metric on $S$. 

---

**Figure 1.9:** Spheres $B_p(0, 1)$ for $p = 1, 2$
Proof. It is easy to see that $d_u$ is symmetric and, further, that $d_u(x, y) = 0$ if and only if $x = y$.

To prove the triangular inequality observe that $a \leq a'$ implies
\[
\frac{a}{a + k} \leq \frac{a'}{a' + k},
\]
which holds for every positive numbers $a, a', k$. Then, we have:
\[
d_u(x, y) = \frac{d(x, y)}{d(x, y) + d(x, u) + d(u, y)} \\
\leq \frac{d(x, z) + d(z, y) + d(x, u) + d(u, y)}{d(x, z) + d(z, y) + d(x, u) + d(u, y)}
\]
(by Inequality 1.7)
\[
= \frac{d(x, z)}{d(x, z) + d(z, y) + d(x, u) + d(u, y)} + \frac{d(z, y)}{d(z, y) + d(z, u) + d(u, y)}
\]
\[
\leq \frac{d(x, z) + d(z, y) + d(z, u)}{d(x, z) + d(z, y) + d(z, u) + d(u, y)}
\]
\[
= d_u(x, z) + d_u(z, y),
\]
which is the desired triangular inequality. 

Theorem 1.8.17 The function $d : \mathcal{P}(S)^2 \rightarrow \mathbb{R}_{\geq 0}$ defined by:
\[
d(X, Y) = \frac{|X \oplus Y|}{|X \cup Y|}
\]
for $X, Y \in \mathcal{P}(S)$ is a metric on $\mathcal{P}(S)$.

Proof. It is clear that $d$ is symmetric and that $d(X, Y) = 0$ if and only if $X = Y$.
So, we need to prove only the triangular inequality. The mapping $\delta$ defined by $\delta(X, Y) = |X \ominus Y|$ is a metric on $\mathcal{P}(X)$, as we proved in Theorem 1.8.15. By Lemma 1.8.16, the mapping $\delta_\emptyset$ is also a metric on $\mathcal{P}(S)$. We have
\[
\delta_\emptyset(X, Y) = \frac{|X \ominus Y|}{|X \ominus Y| + |X \ominus \emptyset| + |\emptyset \ominus Y|}.
\]
Since $X \ominus \emptyset = X$, $\emptyset \ominus Y = Y$, we have
\[
|X \ominus Y| + |X \ominus \emptyset| + |\emptyset \ominus Y| = |X \ominus Y| + |X| + |Y| = 2|X \cup Y|,
\]
which means that $2\delta_\emptyset(X, Y) = d(X, Y)$ for every $X, Y \in \mathcal{P}(S)$. This implies that $d$ is indeed a metric.

Theorem 1.8.18 The function $d : \mathcal{P}(S)^2 \rightarrow \mathbb{R}_{\geq 0}$ defined by:
\[
d(X, Y) = \frac{|X \ominus Y|}{|S| - |X \cap Y|}
\]
for $X, Y \in \mathcal{P}(S)$ is a metric on $\mathcal{P}(S)$. 

Draft September, 2005
Proof. We only prove that $d$ satisfies the triangular axiom. The arguments begin, as in Theorem 1.8.17 with the metric $\delta$. Again, by Lemma 1.8.16, the mapping $\delta_S$ is also a metric on $\mathcal{P}(S)$. We have

$$
\delta_S(X,Y) = \frac{|X \oplus Y| + |S - X| + |S - Y|}{|X \oplus Y| + |S - X| + |S - Y|}
$$

because $|X \oplus Y| + |S - X| + |S - Y| = 2(|S| - |X \cap Y|)$, as the reader can easily verify.

Therefore, $d(X,Y) = 2\delta_S(X,Y)$, which proves that $d$ is indeed a distance.

The metric $\delta_\emptyset$ is also known as the Jaccard dissimilarity coefficient, whereas $1 - \delta_\emptyset(X,Y)$ is the Jaccard similarity coefficient.

A general mechanism for defining a metric on $\mathcal{P}(S)$, where $S$ is a finite set, $|S| = n$, can be introduced starting with two functions:

1. a weight function $w : S \rightarrow \mathbb{R}_{\geq 0}$ such that $\sum_{x \in S} w(x) = 1$, and
2. an injective function $\varphi : \mathcal{P}(S) \rightarrow (S \rightarrow \mathbb{R})$.

The distance defined by the pair $(w, \varphi)$ is the function $d_{w,\varphi} : \mathcal{P}(S)^2 \rightarrow \mathbb{R}_{\geq 0}$ defined by:

$$
d_{w,\varphi}(X,Y) = \left( \sum_{s \in S} w(s)|\varphi(X)(s) - \varphi(Y)(s)|^p \right)^{\frac{1}{p}},
$$

for $X, Y \in \mathcal{P}(S)$.

The function $w$ is extended to $\mathcal{P}(S)$ by:

$$
w(T) = \sum\{w(x) \mid x \in T\}.
$$

Clearly, $w(\emptyset) = 0$ and $w(S) = 1$. Also, if $P, Q$ are two disjoint subsets we have $w(P \cup Q) = w(P) + w(Q)$.

We refer to both $w$ and its extension to $\mathcal{P}(S)$ as a weight function.

The value $\varphi(T)$ of the function $\varphi$ is itself a function $\varphi(T) : S \rightarrow \mathbb{R}$ and each subset $T$ of $S$ defines such a distinct function. These notions are used in the next theorem:

**Theorem 1.8.19** If $w(x) > 0$ for every $x \in S$, then the mapping $d_{w,\varphi} : \mathcal{P}(S)^2 \rightarrow \mathbb{R}$ defined by:

$$
d_{w,\varphi}(U,V) = \left( \sum_{x \in S} w(x)|\varphi(U)(x) - \varphi(V)(x)|^p \right)^{\frac{1}{p}} \quad (1.8)
$$

for $U, V \in \mathcal{P}(S)$ is a metric on $\mathcal{P}(S)$. 

Draft September, 2005
1.8 Dissimilarities, Metrics, and Ultrametrics

Proof. It is clear that \( d_{w,\varphi}(U, U) = 0 \). If \( d_{w,\varphi}(U, V) = 0 \), then \( \varphi(U)(x) = \varphi(V)(x) \) because \( w(x) > 0 \), for every \( x \in S \). Thus, \( \varphi(U) = \varphi(V) \), which implies \( U = V \) due to the injectivity of \( \varphi \).

The symmetry of \( d_{w,\varphi} \) is immediate.

To prove the triangular inequality we apply Minkowski’s inequality. Suppose that \( S = \{x_0, \ldots, x_{n-1}\} \) and let \( U, V, W \in \mathcal{P}(S) \). Define the numbers:

\[
a_i = (w(x_i))^{\frac{1}{p}} \varphi_U(x_i), \\
b_i = (w(x_i))^{\frac{1}{p}} \varphi_V(x_i), \\
c_i = (w(x_i))^{\frac{1}{p}} \varphi_W(x_i),
\]

for \( 0 \leq i \leq n - 1 \). Then, by Minkowski’s inequality we have:

\[
\left( \sum_{i=0}^{n-1} |a_i - b_i|^p \right)^{\frac{1}{p}} \leq \left( \sum_{i=0}^{n-1} |a_i - c_i|^p \right)^{\frac{1}{p}} + \left( \sum_{i=0}^{n-1} |c_i - b_i|^p \right)^{\frac{1}{p}},
\]

which amounts to the triangular inequality \( d_{w,\varphi}(U, V) \leq d_{w,\varphi}(U, W) + d_{w,\varphi}(W, V) \). Thus, we may conclude that \( d_{w,\varphi} \) is indeed a metric on \( \mathcal{P}(S) \).

Example 1.8.20 Let \( w : S \longrightarrow [0,1] \) be a positive weight function. Define the function \( \varphi \) by

\[
\varphi(U)(x) = \begin{cases} \frac{1}{\sqrt{w(U)}} & \text{if } x \in U, \\ 0 & \text{otherwise} \end{cases}
\]

It is easy to see that \( \varphi(U) = \varphi(V) \) if and only if \( U = V \), so \( \varphi \) is an injective function.

Choosing \( p = 2 \), the distance defined in Theorem 1.8.19 becomes:

\[
d_{w,\varphi}^2(U, V) = \left( \sum_{x \in S} w(x) |\varphi(U)(x) - \varphi(V)(x)|^2 \right)^{\frac{1}{2}}.
\]

Suppose initially that neither \( U \) nor \( V \) are empty. Several cases need to be considered:

1. If \( x \in U \cap V \), then

\[
|\varphi(U)(x) - \varphi(V)(x)|^2 = \frac{1}{w(U)} + \frac{1}{w(V)} - \frac{2}{\sqrt{w(U)w(V)}}.
\]

The total contribution of these elements of \( S \) is

\[
w(U \cap V) \left( \frac{1}{w(U)} + \frac{1}{w(V)} - \frac{2}{\sqrt{w(U)w(V)}} \right).
\]

If \( x \in U - V \), then

\[
|\varphi(U)(x) - \varphi(V)(x)|^2 = \frac{1}{w(U)}
\]

and the total contribution is \( w(U - V) \frac{1}{w(U)} \).

Draft September, 2005
2. When \( x \in V - U \) then
\[
|\varphi(U)(x) - \varphi(V)(x)|^2 = \frac{1}{w(V)}
\]
and the total contribution is \( w(V - U) \frac{1}{w(V)} \).

3. Finally, if \( x \notin U \cup V \), then
\[
|\varphi(U)(x) - \varphi(V)(x)|^2 = 0.
\]
Thus, we can write:
\[
d^2_{w,\varphi}(U, V) = w(U \cap V) \left( \frac{1}{w(U)} + \frac{1}{w(V)} - \frac{2}{\sqrt{w(U)w(V)}} \right)
+ w(U - V) \frac{1}{w(U)} + w(V - U) \frac{1}{w(V)}
= \frac{w(U \cap V) + w(U - V)}{w(U)} + \frac{w(U \cap V) + w(V - U)}{w(V)}
- \frac{2w(U \cap V)}{\sqrt{w(U)w(V)}}
= 2 \left( 1 - \frac{w(U \cap V)}{\sqrt{w(U)w(V)}} \right),
\]
where we used the fact that \( w(U \cap V) + w(U - V) = w(U) \) and \( w(U \cap V) + w(V - U) = w(V) \). Thus,
\[
d_{w,\varphi}(U, V) = \sqrt{2 \left( 1 - \frac{w(U \cap V)}{\sqrt{w(U)w(V)}} \right)}
\]
If \( U \neq \emptyset \) and \( V = \emptyset \), then it is immediate that \( d_{w,\varphi}(U, \emptyset) = 1 \). Of course, \( d_{w,\varphi}(\emptyset, \emptyset) = 0 \).

Thus, the mapping \( d_{w,\varphi} \) defined by
\[
d_{w,\varphi}(U, V) = \begin{cases} 
0 & \text{if } U = V = \emptyset, \\
1 & \text{if } U \neq \emptyset \text{ and } V = \emptyset, \\
1 & \text{if } U = \emptyset \text{ and } V \neq \emptyset, \\
\sqrt{2 \left( 1 - \frac{w(U \cap V)}{\sqrt{w(U)w(V)}} \right)} & \text{if } U \neq \emptyset \text{ and } V \neq \emptyset,
\end{cases}
\]
for \( U, V \in \mathcal{P}(S) \) is a metric, which is known as the Ochiai metric on \( \mathcal{P}(S) \).

**Example 1.8.21** Using the same notation as in Example 1.8.20 for a positive weight function \( w : S \rightarrow [0, 1] \), define the function \( \varphi \) by:
\[
\varphi(U)(x) = \begin{cases} 
\frac{1}{w(U)} & \text{if } x \in U, \\
0 & \text{otherwise}
\end{cases}
\]
It is easy to see that \( \varphi \) is an injective function.

Suppose that \( p = 2 \) in the equality 1.8. If \( U \neq \emptyset \) and \( V \neq \emptyset \) we have the following cases:
1. If \( x \in U \cap V \), then
\[
|\varphi(U)(x) - \varphi(V)(x)|^2 = \frac{1}{w(U)^2} + \frac{1}{w(V)^2} - \frac{2}{w(U)w(V)}.
\]

The total contribution of these elements of \( S \) is
\[
w(U \cap V) \left( \frac{1}{w(U)^2} + \frac{1}{w(V)^2} - \frac{2}{w(U)w(V)} \right).
\]
If \( x \in U - V \), then
\[
|\varphi(U)(x) - \varphi(V)(x)|^2 = \frac{1}{w(U)^2}
\]
and the total contribution is \( w(U - V) \frac{1}{w(U)^2} \).

2. When \( x \in V - U \) then
\[
|\varphi(U)(x) - \varphi(V)(x)|^2 = \frac{1}{w(V)^2}
\]
and the total contribution is \( w(V - U) \frac{1}{w(V)^2} \).

3. Finally, if \( x \notin U \cup V \), then \( |\varphi(U)(x) - \varphi(V)(x)|^2 = 0 \).

Summing up these contributions we can write:
\[
d_{w,\varphi}^2(U, V) = \frac{1}{w(U)} + \frac{1}{w(V)} - 2 \frac{w(U \cap V)}{w(U)w(V)}
\]
\[
= \frac{w(U) + w(V) - 2w(U \cap V)}{w(U)w(V)}
\]
\[
= \frac{w(U \oplus V)}{w(U)w(V)}.
\]
If \( V = \emptyset \), \( d_{w,\varphi}(U, \emptyset) = \sqrt{\frac{1}{w(U)}} \); similarly, \( d_{w,\varphi}(\emptyset, V) = \sqrt{\frac{1}{w(V)}} \).

We proved that the mapping \( d_{w,\varphi} \) defined by:
\[
d_{w,\varphi}(U, V) = \begin{cases} 
\sqrt{\frac{w(U \oplus V)}{w(U)w(V)}} & \text{if } U \neq \emptyset \text{ and } V \neq \emptyset \\
\frac{1}{w(U)} & \text{if } U \neq \emptyset \text{ and } V = \emptyset \\
\frac{1}{w(V)} & \text{if } U = \emptyset \text{ and } V \neq \emptyset \\
0 & \text{if } U = V = \emptyset,
\end{cases}
\]
for \( U, V \in \mathcal{P}(S) \), is a metric on \( \mathcal{P}(S) \) known as the \( \chi^2 \) metric. \( \square \)

1.8.4 Ultrametrics and Hierarchies

Ultrametrics represent a strengthening of the notion of metric, where the triangular inequality is replaced by a stronger requirement.

**Definition 1.8.22** A mapping \( d : S^2 \to \mathbb{R}_{\geq 0} \) is a **ultrametric** on \( S \) if it satisfies the following properties: