1B Sets

1B1 (Subsets) We have defined that $A \subseteq B$ means that every element of A is also an element of B. (If A is empty, this is automatically true.) Also, two sets are the same if they contain exactly the same elements. Working with these definitions, prove the following. (Hint: Some of these are really straightforward, but try to be precise in what you claim. Your proofs should generally have the format, "if x is any element of this set, then because so-and-so, it is also an element of that other set.")

(a) Reflexivity: $A \subseteq A$ for every set A.

(b) Antisymmetry: If $A \subseteq B$ and $B \subseteq A$, then $A = B$.

(c) Transitivity: If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

(d) There are sets A and B such that neither $A \subseteq B$ nor $B \subseteq A$.

The first three properties mean that the relation "is subset of" is a so-called *partial order*. It is somewhat similar to the usual order of, for example, the real numbers, in that sets can be "compared" to each other. However, the last property is different from real numbers, because for any two real numbers a, b, either $a \leq b$ or $b \leq a$.

1B2 (Associativity) Prove that for any three sets A, B, C, the unions $(A \cup B) \cup C$ and $A \cup (B \cup C)$ are exactly the same sets. (This means that we can simply write $A \cup B \cup C$ without worrying which union operation is performed first.)

1B3 (Big intersections) The intersection of an infinite sequence of sets $A_1, A_2, \ldots,$ is denoted

$$\bigcap_{k=1}^{\infty} A_k,$$

and is defined to contain every such object x that is an element of *every* A_k. (Thus, if there is (at least one) positive integer k such that $x \notin A_k$, then x is *not* in the intersection.)

Find the following intersections:

(a)

$$\bigcap_{k=1}^{\infty} [0, 1/k]$$

(b)

$$\bigcap_{k=1}^{\infty}]0, 1/k[$$

(c)

$$\bigcap_{k=1}^{\infty} B_k,$$

where $B_k = \{k, k+1, k+2, \ldots\}$ is the set of all integers that are greater or equal to k.

(In the first two parts, the sets are either closed or open intervals of real numbers.)

1B4 (Jaccard similarity) The *Jaccard similarity* of two finite sets A and B is

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

If both sets are empty, the similarity is defined to be 1. Find the following Jaccard similarities:

(a) $J(\{1, 3, 5\}, \{2, 4, 6\})$

(b) $J(\{1, 2, 3, 4\}, \{3, 4, 5, 6\})$

(c) $J(\{1, 2, 3, 4, 5, 6\}, \{1, 2, 3, 4, 5, 6\})$

(d) $J(\{1\}, \{1, 2, 3, 4, 5, 6\})$

1B5 (Jaccard distance) The *Jaccard distance* is $d_J(A, B) = 1 - J(A, B)$.

(a) What is $d_J(A, A)$?

(b) Prove that if $A \neq B$, then $d_J(A, B) > 0$.

(c) Prove that the Jaccard distance is *symmetric*, that is, $d_J(A, B) = d_J(B, A)$ for all finite sets A and B.

(d) What is the largest possible value of $d_J(A, B)$? When exactly does it occur?

(e) What is the smallest possible value of $d_J(A, B)$? When exactly does it occur?

(f) (** Challenging – not required for scoring this problem.) Prove the triangle inequality: If A, B, C are finite sets, then

$$d_J(A, C) \leq d_J(A, B) + d_J(B, C).$$

Jaccard distance is commonly used to define how dissimilar two objects are — based on some sets of "features", with each object having some subset of these features. The properties (a),(b),(c),(f) together show that the Jaccard distance is a proper *metric*, or a *distance function*, which is useful in many algorithms. (We might, for example, want to group a large number of objects into "clusters" such that objects within a cluster have small distances.)

1B6 (Subsets in Cartesian products) Prove that if $A \subseteq B$ and $C \subseteq D$, then $A \times C \subseteq B \times D$.

Problem 1B7 is marked with stars ** to indicate it is a "challenge". A full solution is quite challenging at this point of the course. The problem counts as "extra": When calculating the exercise points for the course, Session 1B is considered to have six exercises (and six points), but it is in fact possible to obtain *seven* points if you also solve 1B7.

1B7 (** CHALLENGE: Ordered pairs defined as sets) The lecture notes claim that "everything" in math can be defined as sets. Yet we have introduced another seemingly basic construction, the *ordered pair* (a, b). Its key property is "elementwise equality": any two ordered pairs (a, b) and (c, d) are *equal* if and only if both $a = c$ and $b = d$.

Suppose all we have is sets, and we *define* that for whatever elements a and b, the ordered pair notation (a, b) *means* the set

$$\{\{a\}, \{a, b\}\}. \tag{*}$$

Prove that the elementwise equality then always holds. Hint: To prove an "if and only if", you need to prove it both ways. (1) You must prove that if a, b, c, d are arbitrary objects and $a = c$ and $b = d$, *then* $(a, b) = (c, d)$, where each ordered pair is understood as a set according to $(*)$. (2) Then prove the opposite direction.