LING82100: notation

Kyle Gorman

1 Sets and membership

A *set* is an abstract, unordered collection of distinct objects, the *members* or *elements* of that set. Members of a set can be of any type, including other sets. Sets may either be finite (e.g., the set consistent of students in this class) or infinite (e.g., the set of real numbers between 0 and 1).

In this class, most of the sets we'll be concerned with consist of numbers. By convention, we use italic capitals (X, Y, Z) to denote sets—and for a few pre-defined sets, double-struck capitals such as $\mathbb{N}$ and $\mathbb{R}$—and lowercase italic letters x, y, z to denote members. Set membership is indicated with the $\in$ symbol (`\in` in LaTeX). The expression $x \in X$ is read as "x is a member of X". Some common numerical sets include:

- $\mathbb{N}$: the set of positive integers (or *counting numbers*)
- $\mathbb{N}_0$: the set of non-negative integers
- $\mathbb{R}_+$: the set of positive real numbers
- $\mathbb{R}$: the set of real numbers, and
- $[0, 1]$: the set of real numbers between 0 and 1 inclusive.

Sets can also be specified by listing—in any order—their members enclosed in curly braces. This is known as *extension notation* or *list notation*. For instance, the set of binomial random variables can be indicated as $B = \{\text{true}, \text{false}\}$.

2 Summation

Many statistical operations involve the computation of sums of numbers. A capital Greek sigma $\sum$ is used to indicate a sum. Summations can either be written as an integer range, as in

$$\sum_{i=1}^{n} (\bar{x}_i - \mu)^2 = (\bar{x}_1 - \mu)^2 + (\bar{x}_2 - \mu)^2 + (\bar{x}_3 - \mu)^2 + \ldots + (\bar{x}_n - \mu)^2$$

or as a set range, as in

$$\sum_{\bar{x} \in \bar{X}} (\bar{x} - \mu)^2$$

You may recognize the above equations as the sum of squared deviances, part of the computation of variance and standard deviation.