Summary of Big-O and Little-o Notation

Vikus JvRensburg

Abstract

This document provides a concise overview of the Big-O and Little-o notations, which are used to describe the asymptotic behavior of functions. These notations are fundamental in mathematical analysis, probability, and computer science for expressing growth rates, convergence, and approximation properties of functions as a variable tends to infinity or another limit point.

Big-O and Little-o Notation

Let f be a real or complex-valued function, and g a real-valued function, both defined on some unbounded subset of the positive real numbers. Assume $g(x) \neq 0$ for all sufficiently large x.

Big-O Notation

We write

$$f(x) = O(g(x)) \quad \text{as } x \to \infty$$

if there exist constants $M > 0$ and $x_0 \in \mathbb{R}$ such that

$$|f(x)| \leq M|g(x)| \quad \text{for all } x \geq x_0.$$

In words, $f(x)$ is *of order* $g(x)$ as $x \to \infty$ if the magnitude of $f(x)$ is bounded by a constant multiple of $|g(x)|$ for sufficiently large x.

Similarly, we say

$$f(x) = O(g(x)) \quad \text{as } x \to a$$

if there exist $\delta > 0$ and $M > 0$ such that

$$|f(x)| \leq M|g(x)| \quad \text{whenever } 0 < |x - a| < \delta.$$

Little-o Notation

We write

$$f(x) = o(g(x)) \quad \text{as } x \to \infty$$

if for every $\varepsilon > 0$, there exists x_0 such that

$$|f(x)| \leq \varepsilon\, g(x) \quad \text{for all } x \geq x_0.$$

Equivalently,

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0.$$

Examples:

$$2x = o(x^2), \qquad \frac{1}{x} = o(1), \quad \text{as } x \to \infty.$$

Relationship Between O and o

The difference between O and o is that:

$$f(x) = O(g(x)) \;\text{ means }\; \exists M > 0 : |f(x)| \leq M|g(x)|,$$

while

$$f(x) = o(g(x)) \;\text{ means }\; \forall \varepsilon > 0, \exists x_0 : |f(x)| \leq \varepsilon|g(x)|.$$

Thus, $f = o(g)$ implies $f = O(g)$, but not conversely. For example,

$$2x^2 = O(x^2) \quad \text{but} \quad 2x^2 \neq o(x^2).$$

Properties

- If $c \neq 0$ and $f = o(g)$, then $cf = o(g)$.
- If $f = o(F)$ and $g = o(G)$, then

$$fg = o(FG), \quad \text{and} \quad f + g = o(F + G).$$

- (Transitivity) If $f = o(g)$ and $g = o(h)$, then $f = o(h)$.

Usage in Probability

Given a sequence of random variables (X_n) in $\mathbb{R}$, and constant c. To say

$$X_n \leq c + o(1) \quad \text{with probability } 1 - o(1)$$

means

$$\forall \varepsilon > 0, \quad \mathbb{P}(X_n \leq c + \varepsilon) \to 1 \quad \text{or equivalently} \quad \mathbb{P}(X_n > c + \varepsilon) \to 0.$$

To say

$$X_n = c + o(1) \quad \text{with probability } 1 - o(1)$$

means

$$\forall \varepsilon > 0, \quad \mathbb{P}(|X_n - c| < \varepsilon) \to 1,$$

or equivalently,

$$X_n \to c \quad \text{in probability.}$$

To say

$$X_n \geq c + o(1) \quad \text{with probability } 1 - o(1)$$

means

$$\forall \varepsilon > 0, \quad \mathbb{P}(X_n > c - \varepsilon) \to 1 \quad \text{or equivalently} \quad \mathbb{P}(X_n < c - \varepsilon) \to 0.$$

Hierarchy of Decay Rates for Sequences

Let $(a_N)_{N\geq 1}$ be a sequence with $a_N \to 0$. We compare several possible rates of decay.

1. Definitions

- **Polynomial decay:**
$$a_N = O(N^{-k}) \quad \text{for some fixed } k > 0.$$
- **Superpolynomial decay:**
$$a_N = O(N^{-k}) \quad \text{for all } k > 0.$$
That is, a_N decays faster than any power of N. Typical examples include $a_N = e^{-\sqrt{N}}$ or $a_N = e^{-(\log N)^2}$.
- **Exponential (geometric) decay:**
$$a_N = O(e^{-cN}) \quad \text{for some } c > 0.$$
- **Superexponential decay:**
$$a_N = O(e^{-cN}) \quad \text{for all } c > 0.$$
Typical examples include $a_N = e^{-N^2}$.

2. Correct Hierarchy of Decay Rates

The asymptotic sizes (for $N \to \infty$) satisfy:

$$e^{-N^2} \ll e^{-cN} \ll e^{-\sqrt{N}} \ll N^{-k},$$

where "$\ll$" means "decays faster than."

Equivalently, the logical implications are:

$$\text{superexponential} \Rightarrow \text{exponential} \Rightarrow \text{superpolynomial} \Rightarrow \text{polynomial},$$

and none of these implications can be reversed.

3. Example of a Strict Separation

For instance,

$$a_N = e^{-\sqrt{N}}$$

satisfies $a_N = O(N^{-k})$ for every k, hence is superpolynomial, but

$$\frac{e^{-\sqrt{N}}}{e^{-cN}} = e^{cN-\sqrt{N}} \to \infty$$

for every $c > 0$, so a_N is not $O(e^{-cN})$. Thus, superpolynomial decay is strictly weaker than exponential decay.