The O and o notation

Let \( f \) and \( g \) be functions of \( x \).

1. Big O. We define \( f = O(g) \) to mean \( f/g \) is bounded, some limit process for \( x \) being given. For example if that process is \( x \to +\infty \) then \( f = O(g) \) means that for some bound \( c \) and lower limit \( x_c \) say, \( |f(x)/g(x)| \leq c \) for all \( x \geq x_c \). \( O \) saves us from specifying \( c \) and \( x_c \) each time. Mostly \( g(x) \) will be a power of \( x \), including \( x^0 = 1 \). “Big O” is short for “Order” which in this context means size, so \( f = O(g) \) means, roughly, that \( f \) is not bigger than \( g \).

Examples
(a) \( x^p = O(x^q) \) as \( x \to +\infty \) if \( p \leq q \), but not as \( x \to 0+ \) (we need \( p \geq q \) then).
(b) \( f(x) = O(1) \) means \( f \) is bounded (for the given limit process). Similarly, \( cf = O(1) \) for any constant \( c \).
(c) \( (x - 1)^{-1} = O(1) \) as \( x \to +\infty \) (take \( c = 1 \) and \( x_c = 2 \)).
In fact we shall see below that \( (x - 1)^{-1} = O(x^{-1}) \).
Note that \( (x - 1)^{-1} \) is not bounded for all \( x \) (take \( x = 1 \)).

2. Small o. We define \( f = o(g) \) to mean \( f/g \to 0 \), again for some given limit process in \( x \). For example if the process is \( x \to +\infty \) then \( f = o(g) \) means that for each \( \varepsilon > 0 \) there is \( x_\varepsilon \) so that \( |f(x)/g(x)| < \varepsilon \) whenever \( x \geq x_\varepsilon \). “Small o” means “of smaller order”, and saves us from specifying \( \varepsilon \) and \( x_\varepsilon \) each time.

Examples
(a) \( x^p = o(x^q) \) as \( x \to +\infty \) if \( p < q \) but not if \( p = q \) (then we have \( O(x^q) \)). We need \( p > q \) if \( x \to 0+ \).
(b) \( f(x) = o(1) \) means \( f(x) \to 0 \) (for the given limit process).
Similarly, \( cf = o(1) \) for any constant \( c \).
Note that \( f(x) = o(1) \) implies \( f(x) = O(1) \). (Why?)
(c) \( (x - 1)^{-1} = o(1) \) as \( x \to +\infty \)
but not as \( x \to 0 \) (or \( 0- \) or \( 0+ \)) but then we do have \( O(1) \) instead.
(d) If \( f \) is continuous at \( x_0 \), say, then \( f(x) \to f(x_0) \), so we have
\[
f(x) - f(x_0) = o(1), \text{ as } x \to x_0.
\]
3. Calculus. O and o obey sum rules

\[
\begin{align*}
  f_1 &= O(g) \quad \text{and} \quad f_2 = O(g) \quad \Rightarrow \quad f_1 + f_2 &= O(g) \\
  f_1 &= o(g) \quad \text{and} \quad f_2 = o(g) \quad \Rightarrow \quad f_1 + f_2 &= o(g)
\end{align*}
\]

and product rules

\[
\begin{align*}
  f_1 &= O(g_1) \quad \text{and} \quad f_2 = O(g_2) \quad \Rightarrow \quad f_1 f_2 &= O(g_1 g_2) \\
  f_1 &= o(g_1) \quad \text{and} \quad f_2 = o(g_2) \quad \Rightarrow \quad f_1 f_2 &= o(g_1 g_2).
\end{align*}
\]

Examples

(a) If \( f(x) = O(x^p) \) then \( x^{-q} f(x) = O(x^{-q}) O(x^p) = O(x^{p-q}) \) as \( x \to 0^+ \). Similarly with o. Thus we can take factors out of equations like \( f(x) = o(x^p) \). The case \( p = q \) gives \( x^{-p} f(x) = o(1) \), i.e., \( x^{-p} f(x) \to 0 \).

(b) Let us improve on 1(c). We have

\[
\frac{x}{x-1} = \frac{x-1}{x-1} + \frac{1}{x-1}.
\]

By 1(c) and the sum rule, \( x(x-1)^{-1} = O(1) + O(1) = O(1) \), as \( x \to +\infty \). Now the product rule allows us to factor out \( x \) to give \( (x-1)^{-1} = O(x^{-1}) \), as promised earlier.

Differentiation is more tricky, and if for example \( f(x) = x^2 \sin(1/x) \) then \( f(x) = o(x) \) but \( f'(x) \neq o(1) \) as \( x \to 0 \).

Integration. Again some caution is needed but the following result will be enough for our purposes. Suppose that \( f = o(g) \), and that \( g \) is of one sign, in a (perhaps one-sided) neighbourhood \( N \) of \( x_0 \). Then

\[
\int_{x_0}^x f = o(\int_{x_0}^x g) \quad \text{as} \quad x \to x_0 \quad \text{within} \quad N.
\]

An example of this is given in the next section. A similar result holds if \( o \) is replaced by \( O \).

A proof of (d), which uses a bit of Analysis, is as follows.

For simplicity suppose that \( g > 0 \), \( x - x_0 > 0 \) in \( N \). For each \( \varepsilon > 0 \) there is \( x_\varepsilon \) so that \( |f(x)| < \varepsilon g(x) \) whenever \( x \) lies between \( x_0 \) and \( x_\varepsilon \). Integrating, we obtain

\[
| \int_{x_0}^x f | \leq \int_{x_0}^x |f| < \varepsilon \int_{x_0}^x g.
\]

Since \( \varepsilon \) was arbitrarily small, this completes the proof.
4. Taylor approximations

(a) The simplest of these is the zeroth order (constant) approximation $f(x) = f(x_0) + o(1)$ which holds as in 1(d) if $f$ is continuous at $x_0$, the limit process (here and below) being $x \to x_0$.

(b) If $f$ is differentiable at $x_0$ then $\frac{f(x) - f(x_0)}{x - x_0} = f'(x_0) + o(1)$, i.e.,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0)$$

by the product rule in 3.

(c) If $f$ is continuously differentiable in a neighbourhood of $x_0$, then by the mean value theorem, $f(x) = f(x_0) + f'(x_1)(x - x_0)$ where $x_1$ lies in the interval between $x$ and $x_0$. Since $f'$ is continuous we obtain, as in (a), $f'(x_1) = f'(x_0) + o(1)$ as $x$ (hence $x_1$) $\to x_0$, so substitution gives another proof of $f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0)$.

(d) If $f$ is twice differentiable at $x_0$, (perhaps one sidedly), then as in (b), $f''(x_0) = f''(x_0) + o(x - x_0)$, i.e.,

$$f''(x) - f''(x_0) = o(x - x_0).$$

Integrating and using 3(d) we obtain

$$f(x) - f(x_0) - f'(x_0)(x - x_0) - \frac{1}{2} f''(x_0)(x - x_0)^2 = o(x - x_0)^2,$$

i.e.,

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + o(x - x_0)^2.$$

(e) In general, if $f$ is $n$ times differentiable at $x_0$, (perhaps one sidedly), then similar reasoning gives

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \ldots + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n + o(x - x_0)^n.$$