DERIVATIVES II

9.1 Maximum: In mathematics, the maximum and minimum (plural: maxima and minima) of a function, known collectively as extrema (singular: extremum), are the largest and smallest value that the function takes at a point either within a given neighborhood (local or relative extremum) or on the function domain in its entirety (global or absolute extremum). Pierre de Fermat was one of the first mathematicians to propose a general technique (called adequality) for finding maxima and minima.

More generally, the maximum and minimum of a set (as defined in set theory) are the greatest and least element in the set. Unbounded infinite sets such as the set of real numbers have no minimum and maximum.

Definition

A real-valued function f defined on a domain X has a global (or absolute) maximum point at $x*$ if $f(x*) \geq f(x)$ for all x in X. Similarly, the function has a global (or absolute) minimum point at $x*$ if $f(x*) \leq f(x)$ for all x in X. **The value of the function at a maximum point is called the maximum value of the function and the value of the function at a minimum point is called the minimum value of the function.**

If the domain X is a metric space then f is said to have a local (or relative) maximum point at the point $x*$ if there exists some $\varepsilon > 0$ such that $f(x*) \geq f(x)$ for all x in X within distance ε of $x*$. Similarly, the function has a local minimum point at $x*$ if $f(x*) \leq f(x)$ for all x in X within distance ε of $x*$. A similar definition can be used when X is a topological space, since the definition just given can be rephrased in terms of neighbourhoods. Note that a global maximum point is always a local maximum point, and similarly for minimum points.

In both the global and local cases, the concept of a strict extremum can be defined. For example, $x*$ is a strict global maximum point if, for all x in X with $x \neq x*$, we have $f(x*) > f(x)$, and $x*$ is a strict local maximum point if there exists some $\varepsilon > 0$ such that, for all x in X within distance ε of $x*$ with $x \neq x*$, we have $f(x*) > f(x)$. Note that a point is a strict global maximum point if and only if it is the unique global maximum point, and similarly for minimum points.

A continuous real-valued function with a compact domain always has a maximum point and a minimum point. An important example is a function whose domain is a closed (and bounded) interval of real numbers

Finding functional maxima and minima

Finding global maxima and minima is the goal of mathematical optimization. If a function is continuous on a closed interval, then by the extreme value theorem global maxima and minima exist. Furthermore, a global maximum (or minimum) either must be a local maximum (or minimum) in the interior of the domain, or must lie on the boundary of the domain. So a method of finding a global maximum (or minimum) is to look at all the local maxima (or minima) in the interior, and also look at the maxima (or minima) of the points on the boundary; and take the biggest (or smallest) one.

Local extrema of differentiable functions can be found by Fermat's theorem, which states that they must occur at critical points. One can distinguish whether a critical point is a local maximum or local minimum by using the first derivative test, second derivative test, or higher-order derivative test, given sufficient differentiability.

For any function that is defined piecewise, one finds a maximum (or minimum) by finding the maximum (or minimum) of each piece separately; and then seeing which one is biggest (or smallest).

Examples

The global maximum of occurs at x = e.
The function x2 has a unique global minimum at x = 0.
The function x3 has no global minima or maxima. Although the first derivative (3x2) is 0 at x = 0, this is an inflection point.
The function has a unique global maximum at x = e. (See figure at right)
The function x-x has a unique global maximum over the positive real numbers at x = 1/e.
The function x3/3 − x has first derivative x2 − 1 and second derivative 2x. Setting the first derivative to 0 and solving for x gives stationary points at −1 and +1. From the sign of the second derivative we can see that −1 is a local maximum and +1 is a local minimum.

Note that this function has no global maximum or minimum.
The function |x| has a global minimum at x = 0 that cannot be found by taking derivatives, because the derivative does not exist at x = 0.
The function cos(x) has infinitely many global maxima at 0, ±2π, ±4π, …, and infinitely many global minima at ±π, ±3π, ….
The function 2 cos(x) − x has infinitely many local maxima and minima, but no global maximum or minimum.

The function $\cos(3\pi x)/x$ with $0.1 \leq x \leq 1.1$ has a global maximum at $x = 0.1$ (a boundary), a global minimum near $x = 0.3$, a local maximum near $x = 0.6$, and a local minimum near $x = 1.0$. (See figure at top of page.)
The function x3 + 3x2 − 2x + 1 defined over the closed interval (segment) [−4,2] has a local maximum at $x = -1-\sqrt{5}/3$, a local minimum at $x = -1+\sqrt{5}/3$, a global maximum at $x = 2$ and a global minimum at $x = -4$.

Functions of more than one variable

Peano surface, a counterexample to some criteria of local maxima of the 19th century. For functions of more than one variable, similar conditions apply. For example, in the (enlargeable) figure at the right, the necessary conditions for a local maximum are similar to those of a function with only one variable. The first partial derivatives as to z (the variable to be maximized) are zero at the maximum (the glowing dot on top in the figure). The second partial derivatives are negative. These are only necessary, not sufficient, conditions for a local maximum because of the possibility of a saddle point. For use of these conditions to solve for a maximum, the function z must also be differentiable throughout. The second partial derivative test can help classify the point as a relative maximum or relative minimum.

In contrast, there are substantial differences between functions of one variable and functions of more than one variable in the identification of global extrema. For example, if a bounded differentiable function f defined on a closed interval in the real line has a single critical point, which is a local minimum, then it is also a global minimum (use the intermediate value theorem and Rolle's theorem to prove this by reductio ad absurdum). In two and more dimensions, this argument fails, as the function.

In relation to sets
Maxima and minima can also be defined for sets. In general, if an ordered set S has a greatest element m, m is a maximal element. Furthermore, if S is a subset of an ordered set T and m is the greatest element of S with respect to order induced by T, m is a least upper bound of S in T. The similar result holds for least element, minimal element and greatest lower bound.

In the case of a general partial order, the least element (smaller than all other) should not be confused with a minimal element (nothing is smaller). Likewise, a greatest element of a partially ordered set (poset) is an upper bound of the set which is contained within the set, whereas a maximal element m of a poset A is an element of A such that if m ≤ b (for any b in A) then m = b. Any least element or greatest element of a poset is unique, but a poset can have several minimal or maximal elements. If a poset has more than one maximal element, then these elements will not be mutually comparable.

In a totally ordered set, or chain, all elements are mutually comparable, so such a set can have at most one minimal element and at most one maximal element. Then, due to mutual comparability, the minimal element will also be the least element and the maximal element will also be the greatest element. Thus in a totally ordered set we can simply use the terms minimum and maximum. If a chain is finite then it will always have a maximum and a minimum. If a chain is infinite then it need not have a maximum or a minimum. For example, the set of natural numbers has no maximum, though it has a minimum. If an infinite chain S is bounded, then the closure Cl(S) of the set occasionally has a minimum and a maximum, in such case they are called the greatest lower bound and the least upper bound of the set S, respectively.

Sample maximum and minimum

In statistics, the sample maximum and sample minimum, also called the largest observation, and smallest observation, are the values of the greatest and least elements of a sample. They are basic summary statistics, used in descriptive statistics such as the five-number summary and seven-number summary and the associated box plot.

The minimum and the maximum value are the first and last order statistics (often denoted X(1) and X(n) respectively, for a sample size of n).

If there are outliers, they necessarily include the sample maximum or sample minimum, or both, depending on whether they are extremely high or low. However, the sample maximum and minimum need not be outliers, if they are not unusually far from other observations.

Robustness

The sample maximum and minimum are the *least* robust statistics: they are maximally sensitive to outliers.

This can either be an advantage or a drawback: if extreme values are real (not measurement errors), and of real consequence, as in applications of extreme value theory such as building dikes or financial loss, then outliers (as reflected in sample extrema) are important. On the other hand, if outliers have little or no impact on actual outcomes, then using non-robust statistics such as the sample extrema simply cloud the statistics, and robust alternatives should be used, such as other quantiles: the 10th and 90th percentiles (first and last decile) are more robust alternatives.

Derived statistics

Other than being a component of every statistic that uses all samples, the sample extrema are important parts of the range, a measure of dispersion, and mid-range, a measure of location. They also realize the maximum absolute deviation: they are the *furthest* points from any given point, particularly a measure of center such as the median or mean.

Applications

Summary statistics

Firstly, the sample maximum and minimum are basic summary statistics, showing the most extreme observations, and are used in the five-number summary and seven-number summary and the associated box plot.

Prediction interval

The sample maximum and minimum provide a non-parametric prediction interval: in a sample set from a population, or more generally an exchangeable sequence of random variables, each sample is equally likely to be the maximum or minimum.

Thus if one has a sample set $\{X_1, \dots, X_n\}$, and one picks another sample X_{n+1}, then this has $1/(n+1)$ probability of being the largest value seen so far, $1/(n+1)$ probability of being the smallest value seen so far, and thus the other $(n-1)/(n+1)$ of the time, X_{n+1} falls between the sample maximum and sample minimum of $\{X_1, \dots, X_n\}$. Thus, denoting the sample maximum and minimum by *M* and *m*, this yields an $(n-1)/(n+1)$ prediction interval of [*m*,*M*].

For example, if *n*=19, then [*m*,*M*] gives an 18/20 = 90% prediction interval – 90% of the time, the 20th observation falls between the smallest and largest observation seen heretofore. Likewise, *n*=39 gives a 95% prediction interval, and *n*=199 gives a 99% prediction interval.

Estimation

Due to their sensitivity to outliers, the sample extrema cannot reliably be used as estimators unless data is clean – robust alternatives include the first and last deciles.

However, with clean data or in theoretical settings, they can sometimes prove very good estimators, particularly for platykurtic distributions, where for small data sets the mid-range is the mostefficient estimator.

They are inefficient estimators of location for mesokurtic distributions, such as the normal distribution, and leptokurtic distributions, however.

Uniform distribution

For sampling without replacement from a uniform distribution with one or two unknown endpoints (so $1, 2, \ldots, N$ with *N* unknown, or $M, M+1, \ldots, N$ with both *M* and *N* unknown), the sample maximum, or respectively the sample maximum and sample minimum, are sufficient and complete statistics for the unknown endpoints; thus an unbiased estimator derived from these will be UMVU estimator.

If only the top endpoint is unknown, the sample maximum is a biased estimator for the population maximum, but the unbiased estimator $\frac{k+1}{k}m - 1$ (where *m* is the sample maximum and *k*is the sample size) is the UMVU estimator; see German tank problem for details.

If both endpoints are unknown, then the sample range is a biased estimator for the population range, but correcting as for maximum above yields the UMVU estimator.

If both endpoints are unknown, then the mid-range is an unbiased (and hence UMVU) estimator of the midpoint of the interval (here equivalently the population median, average, or mid-range).

The reason the sample extrema are sufficient statistics is that the conditional distribution of the non-extreme samples is just the distribution for the uniform interval between the sample maximum and minimum – once the endpoints are fixed, the values of the interior points add no additional information.

Normality testing

Sample extrema can be used for normality testing, as events beyond the 3σ range are very rare. The sample extrema can be used for a simple normality test, specifically of kurtosis: one computes the t-statistic of the sample maximum and minimum (subtracts sample mean and divides by the sample standard deviation), and if they are unusually large for the sample size (as per the three sigma rule and table therein, or more precisely a Student's t-distribution), then the kurtosis of the sample distribution deviates significantly from that of the normal distribution.

For instance, a daily process should expect a 3σ event once per year (of calendar days; once every year and a half of business days), while a 4σ event happens on average every 40 years of calendar days, 60 years of business days (once in a lifetime), 5σ events happen every 5,000 years (once in recorded history), and 6σ events happen every 1.5 million years (essentially never). Thus if the sample extrema are 6 sigmas from the mean, one has a significant failure of normality.

Further, this test is very easy to communicate without involved statistics.

These tests of normality can be applied if one faces kurtosis risk, for instance.

Extreme value theory

Events can be more extreme than any previously observed, as in the 1755 Lisbon earthquake. See also: Extreme value theory

Sample extrema play two main roles in extreme value theory:

firstly, they give a lower bound on extreme events – events can be at least this extreme, and for this size sample; secondly, they can sometimes be used in estimators of probability of more extreme events. However, caution must be used in using sample extrema as guidelines: in heavy-tailed distributions or for non-stationary processes, extreme events can be significantly more extreme than any previously observed event. This is elaborated in black swan theory.