LECTURE NOTES
ECO 613/614
FALL 2007
KAREN A. KOPECKY

Numerical Differentiation

The simplest way to compute a function's derivatives numerically is to use finite difference approximations. Suppose we are interested in computing the first and second derivatives of a smooth function $f : \mathbb{R} \to \mathbb{R}$. The definition of a derivative,

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h},$$

suggests a natural approximation. Take a small number h, (more on how small latter) and

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}. \tag{1}$$

This is the easiest and most intuitive finite difference formula and it is called the forward difference. The forward difference is the most widely used way to compute numerical derivatives but often it is not the best choice as we will see. In order to compare to alternative approximations we need to derive an error bound for the forward difference. This can be done by taking a Taylor expansion of f,

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + \frac{h^3}{6} f'''(x) + \cdots \tag{2}$$

A little manipulation yields

$$f'(x) = \frac{f(x+h) - f(x)}{h} + O(h). \tag{3}$$

We say that this approximation is first-order accurate since the dominate term in the truncation error is $O(h)$.

An alternative formula to the forward difference is to use a two-sided difference or center difference. The center difference formula can be derived by taking the two second-order Taylor series

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(x) + O(h^3), \tag{4}$$

and

$$f(x-h)=f(x)-hf'(x)+\frac{h^2}{2}f''(x)+O(h^3), \tag{5}$$

and subtracting the second series from the the first and dividing by $2h$, obtaining

$$f'(x)=\frac{f(x+h)-f(x-h)}{2h}+O(h^2). \tag{6}$$

Notice that the center difference approximation is second-order accurate since the dominate term in its truncation error is $O(h^2)$. Thus the center difference is more accurate than the forward difference due to its smaller truncation error. For one-dimensional case it is also just as costly to compute (they both require two function evaluations) and therefore should almost always be chosen over the forward difference approximation. Unfortunately the usual tradeoff between accuracy and efficiency comes into play though as we increase the dimensionality of f. If $f:\mathbb{R}^n\to\mathbb{R}^m$ then computing the Jacobian matrix using the forward difference requires $m(n+1)$ function evaluations, while using the center difference requires $2mn$ function evaluations. So center differences take approximately twice as long to compute when n is large.

Three-point approximations can also be derived. Suppose we take the points h, $x+h$, and $x+\alpha h$ and write our approximate to f' as

$$f'(x)\approx af(x)+bf(x+h)+cf(x+\alpha h).$$

Using the Taylor expansions of $f(x+h)$ and $f(x+\alpha h)$ around x and applying the mean value theorem, there exists $x_1\in[x,x+h]$ and $x_2\in[x,x+\alpha h]$ such that

$$\begin{aligned} af(x)+bf(x+h)+cf(x+\alpha h) &= (a+b+c)f(x)+h(b+c\alpha)f'(x)\\ &\quad+\frac{h^2}{2}(b+c\alpha^2)f''(x)+\frac{h^3}{6}[bf'''(x_1)+c\alpha^3 f'''(x_2)]\end{aligned}$$

We now can see how to choose a, b, and c to make the right-hand-side approximately $f'(x)$. They must satisfy

$$a+b+c=0,$$

$$b+c\alpha=1/h,$$

and

$$b+c\alpha^2=0.$$

Solving yields

$$a = \frac{\alpha^2 - 1}{\alpha(1-\alpha)h},$$

$$b = -\frac{\alpha^2}{\alpha(1-\alpha)h},$$

and

$$c = \frac{1}{\alpha(1-\alpha)h}.$$

First note the in fact the center difference is the special case of the three point approximation when $\alpha = -1$. Second notice that the error from this approximation is $O(h^2)$. Not necessarily an improvement over the center difference but more costly to compute (we now have to do three function evaluations instead of two). Why would we want to use it?

One situation is when we want to approximate the derivative at the boundary of the domain. In this situation the center difference is not an option since it requires evaluating the function outside of the domain. We could choose to use the forward or equivalent of the forward but with $h < 0$ (called the backward difference) at these endpoints. But what if we want to obtain a more accurate approximation? In this case it may payoff to use the three point approximation. Often α is set to 2 yielding

$$f'(x) = \frac{1}{2h}[-3f(x) + 4f(x+h) - f(x+2h)] + O(h^2).$$

This will generate a second-order accurate approximate to the derivative at either endpoint by setting h greater than or less than 0.

We can find finite difference approximations for second derivatives and other higher order derivatives using a similar approach. For example, a centered finite difference approximation to the second derivative can be derived by adding together the two Taylor series

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(x) + O(h^4), \tag{7}$$

and

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(x) + O(h^4), \tag{8}$$

and solving for $f''(x)$. This yields

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} + O(h^2).$$

In addition we can derive general second-order accurate approximations to f'' using weighted sums of f evaluated at various points, only now we would need 4 points instead of 3. In an analogous way to the f' case we can use these (usually more costly) general formulas to handle special situations like approximating the second derivative at the boundary of the domain.

How to Choose h

Notice that the truncation error that arises from the finite difference approximations is increasing in h suggesting that h should be as small as possible to reduce this error. But also notice that the computation of the finite differences involves the computer taking differences and sums of floating-point numbers and dividing the result by a small number, h. This is equivalent to multiplying the result by a large number. The multiplication will magnify any round-off errors in the numerator. The smaller h is the larger is this magnification. Hence in fact h should be chosen to minimize total error which includes both the error that arises from truncation (and is decreasing in h) and the round-off error that arises from doing floating-point arithmetic with floating-point numbers (and is increasing in h).

Let us first look at the case of the forward difference. We want to derive an upper bound on the roundoff error incurred when computing the approximation. Let $\widetilde{f}(x)$ be the numerical representation of $f(x)$, and assume that the only error arising in the computation of f is round-off error. Denote machine epsilon by u and

$$L_f = \sup_{x \in [x, x+h]} |f(x)|$$

so that

$$|\widetilde{f(x+h)} - f(x+h)| \leq uL_f, \tag{9}$$

and

$$|\widetilde{f(x)} - f(x)| \leq uL_f, \tag{10}$$

and the

$$|fl(\widetilde{f(x+h)} - \widetilde{f(x)}) - (\widetilde{f(x+h)} - \widetilde{f(x)})| \leq 2uL_f, \tag{11}$$

Then the roundoff error is bounded by

$$\left| \frac{\widetilde{f(x+h)} - \widetilde{f(x)}}{h} - \frac{f(x+h) - f(x)}{h} \right| \leq \frac{4uL_f}{h}. \tag{12}$$

Notice that if h is too small relative to u, the error can easily be huge. If you pick $h = u$ your error is $4L_f$. The total roundoff error increases as we decrease the bandwith.

Now we need an upper bound on the truncation error. We can show, using the mean value theorem, that there exists $c \in [x, x+h]$ such that

$$f'(x) = \frac{f(x+h) - f(x)}{h} - \frac{f''(c)}{2}h. \tag{13}$$

Let

$$M = \sup_{x \in [x, x+h]} |f''(x)|.$$

Then the truncation error is bounded by

$$\left| \frac{f(x+h) - f(x)}{h} - f'(x) \right| \leq \frac{Mh}{2}. \tag{14}$$

Notice that the truncation error decreases as we decrease h. Smaller values of h improve our approximation.

As seen above, choosing h requires a trade-off between more round-off error, on the one hand, and more truncation error, on the other. Thus the optimal h is the one which minimizes the sum of these two errors. Combining the roundoff and truncation error we get that the total error incurred in the computation of $f'(x)$ using the forward difference is bounded by

$$\left| \frac{\widetilde{f(x+h)} - \widetilde{f(x)}}{h} - f'(x) \right| \leq \frac{4uL_f}{h} + \frac{Mh}{2} \tag{15}$$

Minimizing this bound with respect to h leads to an optimal bandwidth of

$$h^* = \left(\frac{8L_f u}{M} \right)^{1/2}. \tag{16}$$

In practice, we will usually not be able to compute L_f or M but a good approximation is to set the optimal bandwidth to

$$h^\star \approx \sqrt{\frac{L_f}{M}} \sqrt{u} \approx |x| \sqrt{u}$$

This rule works well as long as x is not to small. Small x will lead us to bandwidth's that are too small for floating point operations. Thus the general rule of thumb is to set

$$h^\star = \max(|x|, 1)\sqrt{u}.$$

For double precision computing, $\sqrt{u} \approx 10^{-8}$. Finally we can compute the upper bound on the total absolute error incurred in the computation,

$$\frac{4uL_f M^{1/2}}{(8uL_f)^{1/2}} + \frac{M\,(8uL_f)^{1/2}}{2M^{1/2}} = 2\sqrt{2L_f M u}$$

and the relative error is

$$\frac{2\sqrt{2L_f M u}}{|f'(x)|} \approx \frac{\sqrt{L_f M}}{|f'(x)|}\sqrt{u} \approx \sqrt{u}.$$

So using the forward difference with the optimal bandwidth, you shouldn't expect your derivatives to accurate to more than about 8 significant digits.

If, instead of using the forward difference, we use the center difference formula we have a different optimal bandwidth. The derivation is identical to that for the forward difference. By a similar argument to before we can derive an upper bound on the roundoff error

$$\left| \frac{\widetilde{f(x+h)} - \widetilde{f(x-h)}}{2h} - \frac{f(x+h) - f(x-h)}{2h} \right| \leq \frac{4uL_f}{2h}. \tag{17}$$

By using the center difference formula, we do not modify the order of the rounding error. This should not be surprising since the motivation for the center difference formula is to reduce errors due to curvature whereas the rounding error is due to the computer accuracy.

Now we need an upper bound on the truncation error. Again we can show, using the mean value theorem, that there exists $c \in [x, x+h]$ such that

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \frac{f'''(c)}{6}h^2. \tag{18}$$

Let

$$S = \sup_{x \in [x-h, x+h]} |f'''(x)|.$$

Then the truncation error is bounded by

$$\left| \frac{f(x+h) - f(x-h)}{2h} - f'(x) \right| \leq \frac{Sh^2}{6}. \tag{19}$$

The total error in the computation of $f'(x)$ is now bounded by:

$$\left| \frac{\widetilde{f(x+h)} - \widetilde{f(x-h)}}{h} - f'(x) \right| \leq \frac{2uL_f}{h} + \frac{Sh^2}{6}. \tag{20}$$

Minimizing with respect to h leads to the optimal bandwidth:

$$h^* = \left(\frac{12uL_f}{S} \right)^{1/3}. \tag{21}$$

Since, in practice, we cannot compute L_f or S notice that a good approximation to h^* is

$$h^* \approx \left(\frac{L_f}{S}\right)^{1/3} u^{1/3} \approx |x|u^{1/3}.$$

As in the case of the forward difference it is best to rule out h becoming too small so the rule of thumb is to set

$$h^* = \max(|x|, 1)u^{1/3}.$$

For double precision, $u^{1/3}$ is approximately 10^{-6}. Finally we can compute the upper bound on the total error incurred in the computation,

$$4\left(\frac{SL_f^2}{12}\right)^{1/3} u^{2/3} \tag{22}$$

which is on the order of 10^{-11} in double precision. So we can expect are center difference approximation to be accurate to approximately the first 11 digits.

There is one further problem that can arise. If the floating point value of $x + h$ does not equal $x + h$ but $x + h + e$ then we are actually using the approximation

$$\frac{f(x+h+e) - f(x)}{h} = \frac{f(x+h+e) - f(x+h)}{e}\frac{e}{h} + \frac{f(x+h) - f(x)}{h}$$

Notice that this is approximately

$$f'(x+h)\frac{e}{h} + f'(x) \approx \left(1 + \frac{e}{h}\right) f'(x).$$

If the rounding error e is on the order of magnitude of u and h is on the order of magnitude of $\sqrt{u}$ we have now introduced an relative error on the order of $\sqrt{u}$ into the calculation. To avoid this problem it is recommended that you define h as follows. For one-sided (forward or backward) difference first define $\tilde{h}$ as

$$\tilde{h} = \sqrt{u} * \max(|x|, 1)$$

then set xh as

$$xh = x + \tilde{h}$$

and finally set

$$h = xh - x.$$

For two-sided (or center-difference) define h as

$$h = u^{1/3} \max(|x|, 1)$$

then set $xh1$ as

$$xh1 = x + h$$

and

$$xh0 = x - h$$

and finally

$$hh = xh1 - xh0.$$

Then hh is $2h$ without the roundoff error. For the second derivative approximation, the same reasoning can be used to derive a reasonable value for h. The value is

$$h = u^{1/4} \max(|x|, 1).$$

A similar procedure for removing roundoff error in the computation of $x+h$ and $x-h$ should be used. All of this reasoning also extends to any other finite-difference approximations used such as the other three point approximations discussed above.

References

- **Miranda, Mario J. and Paul L. Fackler.** 2002. *Applied Computational Economics and Finance.* Cambridge, MA: MIT Press.

- **Nocedal, Jorge and Stephen J. wright.** 1999. *Numerical Optimization.* Springer-Verlag New York, Inc.