

Tommo So Vowel Harmony: The Math

Laura McPherson and Bruce Hayes
January 2015

I. BACKGROUND AND STARTING POINT

1. Goal of this document

- Explain and justify all of the math in our article.
- Intended audience: people who vaguely remember their math training from long ago.
- Why not in the article itself?
 - Takes a huge amount of space.
 - This isn't research, just standard math.
 - We couldn't find a textbook source that covers this completely or clearly.
 - By writing this out ourselves we can show every single step — much as one would in a traditional phonological derivation.
- Caveat to people who do math all the time:
 - The degree of detail included here is likely to be irritating! We recommend you stop reading after 2.3 and work out the results yourself as an exercise.

2. Substantive points to be established

2.1 *Our system of constraints derives sigmoids*

- If you apply the principles of maxent grammar to the system we described, with one scalar Markedness constraint and one nonscalar Faithfulness constraint, then you predict a sigmoid curve, described by the equation:

$$y = \frac{1}{1 + e^{F - Mx}}$$

y = predicted application rate of phonological process

F = weight of Faithfulness

M = weight of applicable scalar Markedness constraint

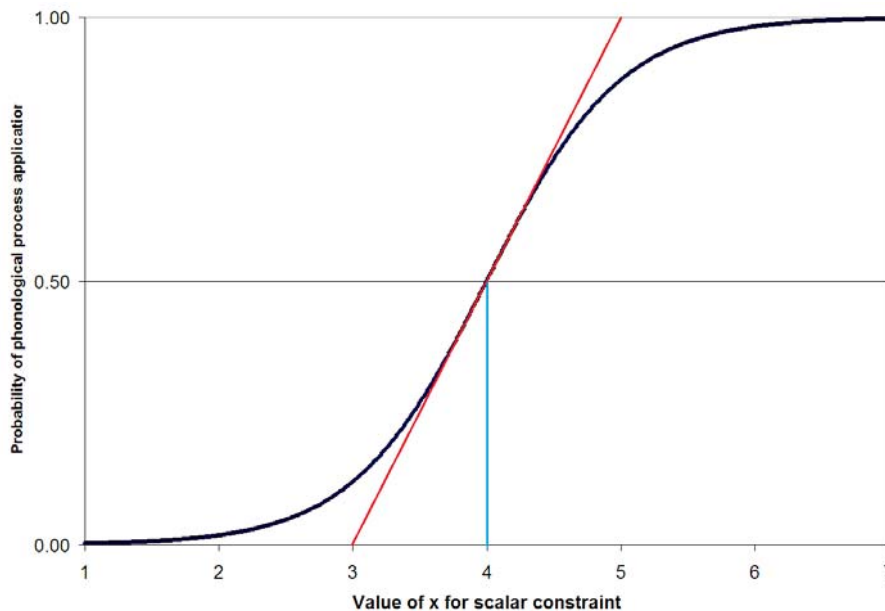
x = value on scale (for us: “distance” from root along the morphological scale we define; higher values = closer

2.2 *Properties of the sigmoid function*

- It asymptotes at 1 for large values of x
- It asymptotes at 0 for small values of x
- It crosses 50% probability at $x = M/F$
- It is symmetrical about this point.
- The maximum slope occurs at this point and is equal to $M/4$.

2.3 A schematic sigmoid showing all these properties

- Following the article, we chose a scale for x ranging from 1 to 7.
- For the graph below, we picked $F = 8$ and $M = 2$.
- So, $F/M = 4$, which is where the sigmoid curve crosses .50 (vertical blue line shows this)
- Symmetry about this point, and asymptotes at zero and one, are visually evident.
- The diagonal red line is the tangent to the sigmoid at the point of maximum slope. The slope of this line can be seen to be 0.5 (line rises by 1 in an interval of 2), which is equal to $M/4$, i.e. $2/4$.



II. MATH YOU NEED TO REMEMBER

3. Algebraic identities

- a. Multiplying exponentiated numbers is the same as adding their exponents

$$a^x * a^y = a^{(x+y)}$$

- b. Anything to the zeroth power is one.

$$a^0 = 1$$

- c. Distribution of multiplication over addition

$$a * (b + c) = (a * b) + (a * c)$$

d. Product of two sums

We can invoke the distributive property repeatedly to get this:

$$\begin{aligned}
 (a + b)(c + d) &= ((a + b) * c) + ((a + b) * d) \\
 &= (c * (a + b)) + (d * (a + b)) && \text{(commute)} \\
 &= ca + cb + da + db && \text{(distribute again, twice)} \\
 &= ac + bc + ad + bd && \text{(commute)}
 \end{aligned}$$

III. OUR SYSTEM OF CONSTRAINTS DERIVES SIGMOIDS

4. Starting point

- Let us assume a generic phonological process; in informal terms we think of the process as “applying” or “not applying.”
- We assume that the two essential conflicting constraints are a scalar Markedness constraint and an opposing nonscalar Faithfulness constraint.

5. Variables

- Let the weight of the scalar Markedness constraint be M.
- Let the weight of the nonscalar Faithfulness constraint be F.
- Let there be a variable x expressing the relevant scale. In our article this is root-closeness, so
 - 7 = root-internal
 - 1 = the farthest away affix.

6. Applying maxent

- Assume we’re trying to decide the phonological outcome for an input whose value along the scale is x .

6.1 Calculate harmony

- The candidate that has undergone the phonological process violates (just) FAITH, so its harmony is: $-F$.
- The candidate that has not undergone the phonological process violates (just) the scalar Markedness constraint, so its harmony is: $-Mx$.
- We assume that all other candidates are ruled out by very strong constraints — super-high weights, and so need not be taken into account.

6.2 Negate and exponentiate

- As our article notes, the next step in maxent is to take e to the power of the negative harmony, for each candidate.

Candidate that has undergone phonology: e^{-F}
 Candidate that has not undergone phonology: e^{-Mx}
 All other candidates: $e^{-(\text{some very large value})}$

- e to some very large value is close enough to zero that we can justifiably treat these candidates as receiving the value zero.

6.3 Calculate the denominator (“Z”)

- Z is the sum of what you got for all candidates.
- As note above, most of these get vanishing small values, which we will ignore.
- So really, we just have two candidates to worry about, and their values sum to:

$$Z = e^{-F} + e^{-Mx}$$

6.4 Find the probabilities of the candidates

- In maxent, this is the result of step 6.2 above, divided by Z .

a. Probability of the “undergoing” candidate = $\frac{e^{-F}}{Z}$

b. Probability of “nonundergoing” candidate = $\frac{e^{-Mx}}{Z}$

7. Deriving the sigmoid curve

- We want to see how the probability of the “undergoing” candidate varies with x , its value along the scale.

7.1 The algebra

- Start with the probability that we just derived for the probability of the “undergoing” candidate (6.4a):

$$\text{Probability of undergoing candidate} = \frac{e^{-F}}{Z}$$

- Substitute in the formula for Z (6.3):

$$\text{Probability of undergoing candidate} = \frac{e^{-F}}{e^{-F} + e^{-Mx}}$$

- Multiply top and bottom by e^F :

$$\text{Probability of undergoing candidate} = \frac{e^F * e^{-F}}{e^F * (e^{-F} + e^{-Mx})}$$

- On the bottom, distribute multiplication over addition (3c):

$$\text{Probability of undergoing candidate} = \frac{e^F * e^{-F}}{(e^F * e^{-F}) + (e^F * e^{-Mx})}$$

- Multiplication of exponentiated terms is the same as exponentiation by the sum of the exponents (3a).

$$\text{Probability of undergoing candidate} = \frac{e^{(F + -F)}}{e^{(F + -F)} + e^{F - Mx}}$$

- Summing:

$$\text{Probability of undergoing candidate} = \frac{e^{(0)}}{e^{(0)} + e^{F - Mx}}$$

- Any positive number to the zeroth power is 1 (3b):

$$\text{Probability of undergoing candidate} = \frac{1}{1 + e^{F - Mx}}$$

7.2 Result

- We are done! What we have is a version of the standard *logistic function*, which plots as a sigmoid (see 2.3 above).
- The function derives a probability of the phonology-undergoing candidate from its value on the markedness scale, x .
- Next, let us provide mathematical demonstrations of the properties of this function laid out in §2.2 above.

IV. FIRST TWO PROPERTIES OF THE SIGMOID: LIMITS AS x IS MADE HIGH OR LOW

8. Asymptotes to 1 as x becomes large

- We consider our sigmoid function, using y to mean “probability of the undergoing candidate”:

$$y = \frac{1}{1 + e^{F - Mx}}$$

- The informal reasoning behind the asymptotes:
 - If x is very large, then Mx is very large too (we use positive weights).
 - Then $F - Mx$ is very large and negative.
 - Then $e^{F - Mx}$ is very small.
 - Then $\frac{1}{1 + e^{F - Mx}}$ approaches 1.
- Compare Tommo So, with near-obligatory root harmony.

9. Asymptotes to zero as x gets low

- Here, by “small” we will include even negative numbers.
- Reasoning:
 - If x is very negative, then Mx is very negative too (we use positive weights).
 - Then $F - Mx$ is very large and positive
 - Then $e^{F - Mx}$ is very large.
 - Then $\frac{1}{1 + e^{F - Mx}}$ approaches zero.
- Compare Tommo So, with zero harmony in the “outermost” inflectional levels.

V. SECOND PROPERTY OF THE SIGMOID: CROSSES 50% AT $x = F/M$

10. Demonstration

- We plug the value F/M into our formula for the sigmoid, replacing x :

$$y = \frac{1}{1 + e^{F - Mx}}$$

$$= \frac{1}{1 + e^{F - (M * F/M)}}$$

- The M terms cancel each other out and can therefore be removed:

$$= \frac{1}{1 + e^{(F - F)}}$$

- The F terms disappear:

$$= \frac{1}{1 + e^0}$$

- Anything to the zero is one (3b):

$$= \frac{1}{1 + 1}$$

- And so:

$$y = \frac{1}{2}$$

- This is 50%, the probability of undergoing phonology at the point F/M.

V. THIRD PROPERTY OF THE SIGMOID: SYMMETRICAL ABOUT $x = F/M$

11. What we want to establish

- We know that at $x = F/M$ the value of the function is .5 (see previous section).
- Let us consider two points with values of x symmetric about F/M. Call them $x + \Delta x$ and $x - \Delta x$.
- The amount that the function computed at $x + \Delta x$ *exceeds* .5 should be the same as the amount that the function at point $x - \Delta x$ *falls below* .5. This is what is meant here by symmetry.

12. The algebra that demonstrates symmetry

- Start with an equality that looks arbitrary but is chosen with forethought and is obviously true:

$$1 + e^{M\Delta x} + e^{-M\Delta x} + 1 = 1 + e^{M\Delta x} + e^{-M\Delta x} + 1$$

- Replace the last instance of 1 with its equal, e^0 ((3b)).

$$1 + e^{M\Delta x} + e^{-M\Delta x} + 1 = 1 + e^{M\Delta x} + e^{-M\Delta x} + e^0$$

- Since $M\Delta x + -M\Delta x = 0$, we can replace the zero with this expression:

$$1 + e^{M\Delta x} + e^{-M\Delta x} + 1 = 1 + e^{M\Delta x} + e^{-M\Delta x} + e^{M\Delta x + -M\Delta x}$$

- Since $e^{a+b} = e^a * e^b$ ((3a)), we can rewrite the $e^{M\Delta x + -M\Delta x}$ term like this:

$$1 + e^{M\Delta x} + e^{-M\Delta x} + 1 = 1 + e^{M\Delta x} + e^{-M\Delta x} + e^{M\Delta x} * e^{-M\Delta x}$$

- Since $(a + b)(c + d) = ac + ad + bc + bd$ (see (3d) above), we can rewrite the right side like this:

$$1 + e^{M\Delta x} + e^{-M\Delta x} + 1 = (1 + e^{M\Delta x}) * (1 + e^{-M\Delta x})$$

To check this, observe that the right side of the new version, when multiplied out, yields the right side of the old version.

- Divide both sides by $(1 + e^{M\Delta x})$:

$$\frac{1 + e^{M\Delta x} + e^{-M\Delta x} + 1}{1 + e^{M\Delta x}} = \frac{(1 + e^{M\Delta x}) * (1 + e^{-M\Delta x})}{1 + e^{M\Delta x}}$$

- Simplify on the right side by dividing top and bottom by $1 + e^{M\Delta x}$:

$$\frac{1 + e^{M\Delta x} + e^{-M\Delta x} + 1}{1 + e^{M\Delta x}} = 1 + e^{-M\Delta x}$$

- Split up the left side, giving each term the same denominator:

$$\frac{1 + e^{M\Delta x}}{1 + e^{M\Delta x}} + \frac{e^{-M\Delta x} + 1}{1 + e^{M\Delta x}} = 1 + e^{-M\Delta x}$$

- Simplify the fraction on the left side:

$$1 + \frac{e^{-M\Delta x} + 1}{1 + e^{M\Delta x}} = 1 + e^{-M\Delta x}$$

- Divide both sides by $1 + e^{-M\Delta x}$:

$$\frac{1}{1 + e^{-M\Delta x}} + \frac{1}{1 + e^{M\Delta x}} = 1$$

- Subtract .5 from both sides:

$$\frac{1}{1 + e^{-M\Delta x}} + \frac{1}{1 + e^{M\Delta x}} - .5 = .5$$

- Subtract $\frac{1}{1 + e^{-M\Delta x}}$ from both sides:

$$\frac{1}{1 + e^{M\Delta x}} - .5 = .5 - \frac{1}{1 + e^{-M\Delta x}}$$

- Zero may be added in without change:

$$\frac{1}{1 + e^{M\Delta x + 0}} - .5 = .5 - \frac{1}{1 + e^{-M\Delta x + 0}}$$

- Replace zero by $F - F$:

$$\frac{1}{1 + e^{M\Delta x + F - F}} - .5 = .5 - \frac{1}{1 + e^{-M\Delta x + F - F}}$$

- Rearranging order of addends in the exponent on both sides:

$$\frac{1}{1 + e^{-F + M\Delta x + F}} - .5 = .5 - \frac{1}{1 + e^{-F - M\Delta x + F}}$$

- Substitute in $M * -F/M$ for $-F$ (it's the same thing, since the multiplication and division cancel each other out):

$$\frac{1}{1 + e^{(M * -F/M) + M\Delta x + F}} - .5 = .5 - \frac{1}{1 + e^{(M * -F/M) - M\Delta x + F}}$$

- Extract the common factor M ((3c)):

$$\frac{1}{1 + e^{M * (-F/M + \Delta x) + F}} - .5 = .5 - \frac{1}{1 + e^{M * (-F/M - \Delta x) + F}}$$

- Put negative signs on the outside:

$$\frac{1}{1 + e^{-M * (F/M - \Delta x) + F}} - .5 = .5 - \frac{1}{1 + e^{-M * (F/M + \Delta x) + F}}$$

- Swap order:

$$\frac{1}{1 + e^{F - M * (F/M - \Delta x)}} - .5 = .5 - \frac{1}{1 + e^{F - M * (F/M + \Delta x)}}$$

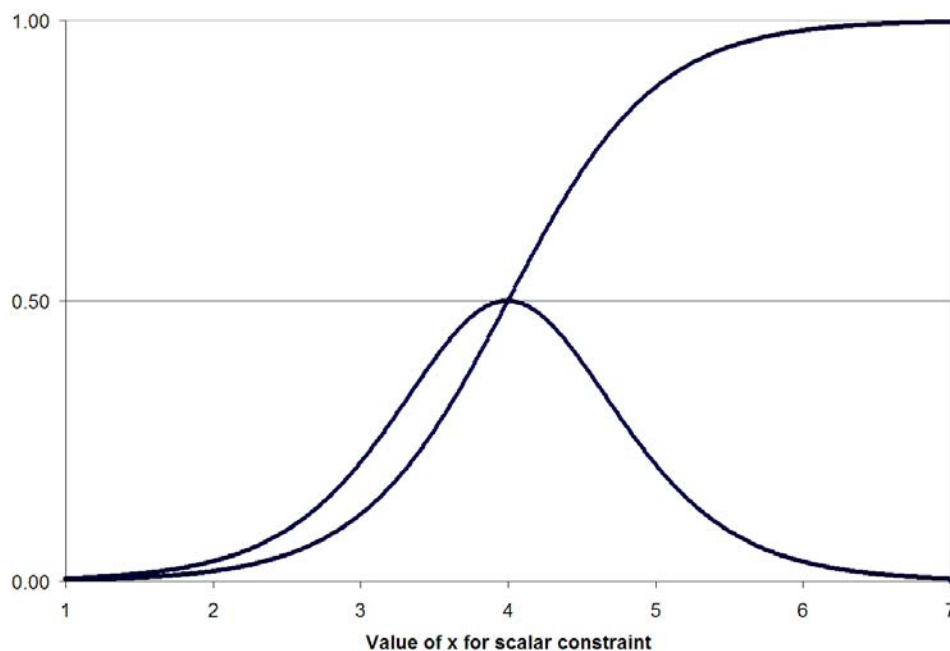
- This is just what we want. Recall that the formula deriving predicted probability of phonological application is $\frac{1}{1 + e^{F - M * x}}$. On the left side, we evaluate the function for the value $x = F/M - \Delta x$, which is Δx below the .5 crossover value of F/M ; and we compute how much this is above the crossover value of .5. On the right side, we evaluate the function for the value $x = F/M + \Delta x$, which is Δx above the crossover value; and we

compute how much this is *below* the crossover value of .5. Since these turn out always to be equal, the function is symmetrical about the value F/M .

VI. FOURTH PROPERTY OF THE SIGMOIDS: SLOPE AT $F/M = M/4$

13. Finding the slope at the symmetry point

- In calculus the steepness of a curve at any given point is expressed as its *derivative*.
- Here is the logistic curve plotted earlier on p. 2 above, this time plotted together with its derivative. The derivative has a symmetrical hump peaking at M/F (which you will remember is $8/2 = 4$), where (as we expected) it has the value 0.5, since this is $M/4$.



- Calculating the derivative of $y = \frac{1}{1 + e^{F-Mx}}$ (see Appendix for how this can be done), it turns out to be:

$$\frac{dy}{dx} = \frac{Me^{F-Mx}}{(1+e^{F-Mx})^2}$$

This is indeed the function that was used in plotting the derivative curve in the figure above.

- We can now profitably ask what is the slope at the symmetry point, which it will be recalled, is at $x = F/M$. We substitute F/M into the formula for the derivative, and obtain:

$$\frac{dy}{dx}(F/M) = \frac{Me^{F-M \cdot F/M}}{(1+e^{F-M \cdot F/M})^2}$$

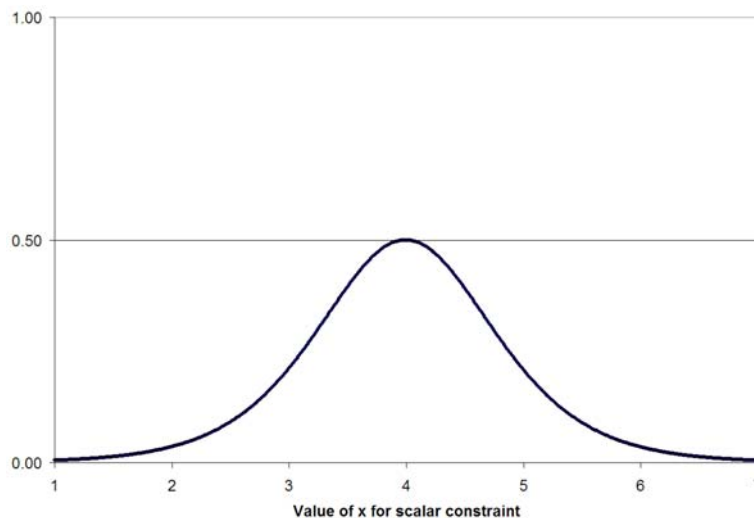
$$\begin{aligned}
 &= \frac{Me^{F-F}}{(1+e^{F-F})^2} \\
 &= \frac{Me^0}{(1+e^0)^2} \\
 &= \frac{M*1}{(1+1)^2} \\
 &= \frac{M}{(2)^2} \\
 &= \frac{M}{4}
 \end{aligned}$$

- So, we've established that the slope at the symmetry point is the Markedness weight divided by 4.

VII. FIFTH PROPERTY OF THE SIGMOIDS: M/4 IS THE MAXIMUM SLOPE

14. The last detail

- We've found that the slope at F/M is indeed M/4, but we haven't shown that that is the *steepest* slope of the logistic curve.
- In calculus, the way you find a maximum of a function is to find the spot where its derivative levels out. For example, if we look again at (a pristine version of) the derivative of the logistic function that we plotted earlier, it appears like this:



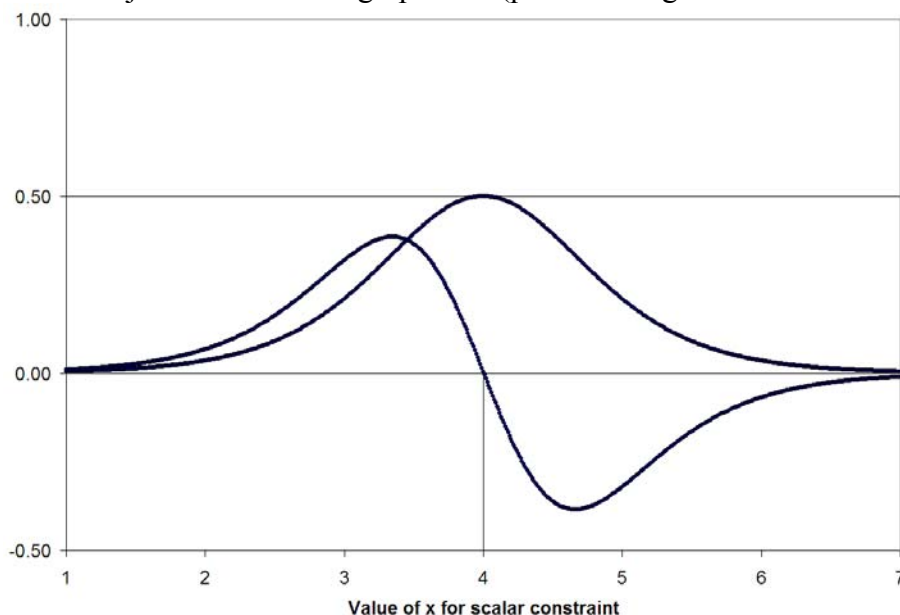
It would appear (and we'll nail this down shortly) that the maximum is indeed at $x = 4$, with the value 0.5. To verify this, we would want to know that a line tangent to the

derivation function at this point would be level; i.e. would have slope zero.
(Conveniently, the vertical scale line at $y = .5$ actually happens to be this tangent line.)

- So what we do is compute the *second derivative*; i.e. the derivative of the derivative, and find where it goes to zero.

15. The second derivative

- We'll examine the equation for the second derivative in a moment. But first, it would be useful just to look at the graph of it (plotted along with the first derivative):



- The second derivative has an intriguing symmetrical pattern: just a little bit above zero coming up from negative infinity, peaking, then plunging rapidly to its lowest value, and lastly edging upward to remain just a little bit below zero out to positive infinity. The point at which the plunge reaches zero looks to be at $x = 4$; let us confirm.
- The second derivative of our logistic function $\frac{1}{1 + e^{F-Mx}}$ turns out to be:

$$\frac{d^2y}{dx^2} = \frac{M^2 e^{F-Mx} (e^{F-Mx} - 1)}{(1 + e^{F-Mx})^3}$$

Again, see Appendix for how this can be calculated.

- The last step is to plug the symmetry point, $x = F/M$, into our formula for the second derivative. If F/M is a maximum¹ this ought to yield zero.

$$\frac{d^2y}{dx^2}(F/M) = \frac{M^2 e^{F-M \cdot F/M} (e^{F-M \cdot F/M} - 1)}{(1 + e^{F-M \cdot F/M})^3}$$

¹ Or minimum, really, but as you can see it turned out to be a maximum.

$$\begin{aligned}
&= \frac{M^2 e^{F-F} (e^{F-F} - 1)}{(1+e^{F-F})^3} \\
&= \frac{M^2 e^0 (e^0 - 1)}{(1+e^0)^3} \\
&= \frac{M^2 1 (1 - 1)}{(1+1)^3} \\
&= \frac{M^2 1 (0)}{(2)^3} \\
&= 0
\end{aligned}$$

Sure enough.

- Summarizing: the second derivative of the logistic function $y = \frac{1}{1 + e^{F-Mx}}$ is $\frac{M^2 e^{F-Mx} (e^{F-Mx} - 1)}{(1+e^{F-Mx})^3}$. This reaches zero at M/F , the symmetry point of the logistic function. This means that the first derivative, representing the slope of the logistic function, reaches an extreme at M/F , which by inspection we see is the maximum. The maximum slope occurs as M/F .

Appendix: Finding the derivatives

XIII. REVIEWING THE RULES FOR DIFFERENTIATION IN CALCULUS

16. Derivative of a constant

If $y(x) = c$, where c is some constant, then

$$\frac{dy}{dx} = 0$$

17. Derivative of a linear equation

If $y(x) = mx + b$, where m and b are constants, then

$$\frac{dy}{dx} = m$$

18. Differentiating a power

If

$$y = x^n$$

then

$$\frac{dy}{dx} = nx^{n-1}$$

It's fine for n to be negative, since $1/x^n$ is the same as x^{-n} . We will see this below.

19. Differentiating exponentials

If

$$y = e^x$$

then, amazingly,

$$\frac{dy}{dx} = e^x$$

20. Sum Rule

The derivative of the sum of two functions is the sum of their derivatives.

21. Product Rule

If

$$y = f(x) * g(x)$$

then

$$\frac{dy}{dx} = (f(x) * \frac{dg}{dx}) + (g(x) * \frac{df}{dx})$$

22. Chain Rule

if

$$y = f(g(x))$$

then

$$\frac{dy}{dx} = df/dg * dg/dx$$

This can be applied more than once, so if

$$y = f(g(h(x)))$$

then

$$\frac{dy}{dx} = df/dg * dg/dh * dh/dx$$

IX. FINDING THE FIRST DERIVATIVE OF THE LOGISTIC FUNCTION

23. Setting up the problem for solution with the Chain Rule

- We seek to differentiate our logistic function, which is:

$$y = \frac{1}{1 + e^{F-Mx}}$$

- This is facilitated if we rewrite it as a “function of a function of a function”, with intermediate steps.

$$y = f(g) = \frac{1}{g}$$

$$g(h) = 1 + e^h$$

$$h(x) = F - Mx$$

- In other words, we work through the formula and re-express its content step-by-step. The functions $h(x)$, $g(h)$, and $f(g)$ reexpress the formula going “from the inside out”. We can put the functions back together as follows:

$$y = f(g(h(x)))$$

- The Chain Rule (22e) tells us to take the derivative of each “subfunction”, then multiply the derivatives out to get the whole derivative.

24. Differentiating $f(g)$, $g(h)$, and $h(x)$

- If $f(g) = \frac{1}{g}$, then

$$\frac{df}{dg} = \frac{-1}{g^2}$$

This applies (18), treating $\frac{1}{g}$ as its equivalent g^{-1} .

- If $g(h) = 1 + e^h$, then we have a sum of two functions; a trivial one (the constant 1) and e^h .
 - We differentiate them separately, then add ((20)).
 - The derivative of the constant one is zero ((16)).
 - The derivative of e^h is itself ((19)).
 - so:

$$\frac{dg}{dh} = e^h$$

- If $h(x) = F - Mx$ then by (17) we have

$$\frac{dh}{dx} = -M$$

25. Applying the Chain Rule

- Now that we have our three derivatives, we can substitute them back into the formula of the Chain Rule:

$$\frac{dy}{dx} = df/dg * dg/dh * dh/dx$$

$$= \frac{-1}{g^2} * e^h * -M$$

- And we know what g and h are (second bullet point of this section), so we can substitute their definitions back in to the formula:

$$\frac{dy}{dx} = \frac{-1}{(1 + e^h)^2} * e^{F-Mx} * -M$$

And once again, filling in definition of $h(x)$:

$$= \frac{-1}{(1 + e^{F-Mx})^2} * e^{F-Mx} * -M$$

- Cleanup: the two minus signs cancel each other, and we rearrange prettily into a single fraction:

$$\frac{dy}{dx} = \frac{Me^{F-Mx}}{(1+e^{F-Mx})^2}$$

- This is the correct answer, and is indeed the derivative assumed in (13), second bullet point, above.

26. Checking the answer

- Are we sure we're right? A check comes from the process of making the actual graphs displayed above. These are made (in Excel) by plotting one dot at every interval of .01 along the x axis. The fraction $\frac{y(x_{n+1}) - y(x_n)}{x_{n+1} - x_n}$, where x_n is the n th value along the x axis, is a close approximation to the derivative when the interval between dots is small. We find that the values thus obtained closely match those calculated with the true derivative $\frac{Me^{F-Mx}}{(1+e^{F-Mx})^2}$, thus confirming that we haven't made a mistake.

X. FINDING THE SECOND DERIVATIVE OF THE LOGISTIC FUNCTION

27. Defining the task

- To obtain the second derivative, we need to differentiate the first derivative we just obtained. I.e. we want $\frac{dy}{dx}$, where $y(x)$ is now redefined as

$$y(x) = \frac{Me^{F-Mx}}{(1+e^{F-Mx})^2}$$

- This derivative has two appearances of x in it, so we can't follow the same method we used for the first derivative. Rather, our strategy is to treat $\frac{Me^{F-Mx}}{(1+e^{F-Mx})^2}$ as a product, and use the Product Rule (21).

28. Setting up for the Product Rule

- The product is this:

$$y(x) = Me^{F-Mx} * \frac{1}{(1+e^{F-Mx})^2}$$

- We'll give the multiplicands names, so we can refer to them.

$$f(x) = Me^{F-Mx}$$

$$g(x) = \frac{1}{(1+e^{F-Mx})^2}$$

- The Product Rule tells us that we will get our derivative if we compute the derivatives of $f(x)$ and $g(x)$ and plug them into the formula

$$\frac{dy}{dx} = (f(x) * \frac{dg}{dx}) + (g(x) * \frac{df}{dx})$$

- This requires that we differentiate both $f(x)$ and $g(x)$.

29. Differentiating $f(x)$

- We have

$$f(x) = Me^{F-Mx}$$

- This can be set up for application of the Chain Rule again, with functions $h(k)$, $k(m)$, $m(x)$

$$h(k) = Mk$$

$$k(m) = e^m$$

$$m(x) = F - Mx$$

- The derivatives are:

$$\frac{dh}{dk} = M \quad (\text{by (17), where the } b \text{ of (17) is taken to be zero})$$

$$\frac{dk}{dm} = e^m \quad (\text{by (19)})$$

$$\frac{dm}{dx} = -M \quad (\text{by (17)})$$

- Multiplying it out, following the Chain Rule:

$$\frac{df}{dx} = \frac{dh}{dk} * \frac{dk}{dm} * \frac{dm}{dx}$$

$$= M * e^m * -M$$

$$= -M^2 e^m$$

- Substituting in definition of $m(x)$:

$$= -M^2 e^{F-Mx}$$

This is the derivative of $f(x)$.

30. Differentiating $g(x)$

- We have:

$$g(x) = \frac{1}{(1+e^{F-Mx})^2}$$

- This can be set up for application of the Chain Rule again, with functions $h(k)$, $k(m)$, $m(x)$ (we are recycling some letters here).

$$h(k) = \frac{1}{k^2}$$

$$k(m) = 1 + e^m$$

$$m(x) = F - Mx$$

- The derivatives:

$$\frac{dh}{dk} = \frac{-2}{k^3} \quad (\text{by (18), where the power in question is } -2)$$

$$\frac{dk}{dm} = e^m \quad (\text{sum of the derivatives of the constant 1 and of } e^m; \text{ see (20), (16), and (19)})$$

$$\frac{dm}{dx} = -M \quad (\text{by (17)})$$

- Multiplying it out, following the Chain Rule:

$$\frac{df}{dx} = \frac{dh}{dk} * \frac{dk}{dm} * \frac{dm}{dx}$$

$$= \frac{-2}{k^3} * e^m * -M$$

$$= \frac{2Me^m}{k^3}$$

- Substituting in definition of $k(x)$:

$$= \frac{2Me^m}{(1 + e^m)^3}$$

- Substituting in definition of $m(x)$:

$$\frac{dg}{dx} = \frac{2Me^{F-Mx}}{(1 + e^{F-Mx})^3}$$

This is the derivative of $g(x)$.

31. Applying the Product Rule

- We now have the functions and derivatives we need to plug into the Product Rule formula:

$$\frac{dy}{dx} = (f(x) * \frac{dg}{dx}) + (g(x) * \frac{df}{dx})$$

- All four expressions we need appear on the last three pages. Hunting them down and plugging them in, we get:

$$\frac{dy}{dx} = (Me^{F-Mx} * \frac{2Me^{F-Mx}}{(1 + e^{F-Mx})^3}) + (\frac{1}{(1+e^{F-Mx})^2} * -M^2e^{F-Mx})$$

This is indeed the derivative, but it looks like a mess.

- Let us clean it up a bit. Factor out M^2 (3c):

$$\frac{dy}{dx} = M^2 * ((e^{F-Mx} * \frac{2e^{F-Mx}}{(1 + e^{F-Mx})^3}) + (\frac{1}{(1+e^{F-Mx})^2} * -e^{F-Mx}))$$

- Factor out $\frac{1}{(1+e^{F-Mx})^2}$:

$$\frac{dy}{dx} = M^2 * \frac{1}{(1+e^{F-Mx})^2} * ((e^{F-Mx} * \frac{2e^{F-Mx}}{1 + e^{F-Mx}}) + -e^{F-Mx})$$

- Make a pretty fraction:

$$\frac{dy}{dx} = \frac{M^2}{(1+e^{F-Mx})^2} * ((e^{F-Mx} * \frac{2e^{F-Mx}}{1 + e^{F-Mx}}) + -e^{F-Mx})$$

- Factor out e^{F-Mx} :

$$\frac{dy}{dx} = \frac{M^2}{(1+e^{F-Mx})^2} * e^{F-Mx} * (\frac{2e^{F-Mx}}{1 + e^{F-Mx}} - 1)$$

- Make a pretty fraction:

$$\frac{dy}{dx} = \frac{M^2 e^{F-Mx}}{(1+e^{F-Mx})^2} * \left(\frac{2e^{F-Mx}}{1+e^{F-Mx}} - 1 \right)$$

- Replace the 1 by the equivalent $\frac{1+e^{F-Mx}}{1+e^{F-Mx}}$:

$$\frac{dy}{dx} = \frac{M^2 e^{F-Mx}}{(1+e^{F-Mx})^2} * \left(\frac{2e^{F-Mx}}{1+e^{F-Mx}} - \frac{1+e^{F-Mx}}{1+e^{F-Mx}} \right)$$

- Gather the terms with the same denominator:

$$\frac{dy}{dx} = \frac{M^2 e^{F-Mx}}{(1+e^{F-Mx})^2} * \left(\frac{2e^{F-Mx} - (1+e^{F-Mx})}{1+e^{F-Mx}} \right)$$

- Subtract:

$$\frac{dy}{dx} = \frac{M^2 e^{F-Mx}}{(1+e^{F-Mx})^2} * \frac{e^{F-Mx} - 1}{1+e^{F-Mx}}$$

- Multiply out the whole thing so we get a single fraction:

$$\frac{dy}{dx} = \frac{M^2 e^{F-Mx} (e^{F-Mx} - 1)}{(1+e^{F-Mx})^3}$$

- This is about as clean as we can make it. This is the second derivative of our logistic function, as employed in (15) above to verify the location of maximum slope.
- It checks by the same method we used for the first derivative.