Why MAXENT is not for you.

Nicholas M Glykos <glykos@crystal2.imbb.forth.gr>

June 2001

The short answer is *"because what you need is the answer to a different question"*. The rest of this (strictly non-mathematical, strictly non-graphical) note expands, elaborates and exemplifies this thesis. If what you are looking for are some illustrated examples of the application of the program, please have a look at http://origin.imbb.-forth.gr/∼glykos/GraphEnt-html/GraphEnt.html, or the postscript documentation that comes with the program's distribution.

0.1   What is the question ?

Since you are reading this, you most probably have a map which does not show the features you expected (or hoped for). So, you are wondering whether there is way to calculate a "better" map. If you are a pragmatist, please do leave the quotes around "better" : the map you want may not be better in any reasonably justifiable way, but it is a map that will be consistent with a hypothesis (or expectation) you have already formed. Let me make this clear with an example : suppose you have collected a native and a derivative data sets and you have calculated a difference Patterson function which shows nothing more than ripples, or long-connected features, or ... The "better" map you most probably have in mind, is a Patterson function which will only show few strong peaks (on an otherwise uniform background), *with these peaks being consistent with (and fully accountable by) a heavy atom structure containing a small number of atoms.* It is even possible that you can postulate with some confidence that this heavy atom structure should only contain two atoms per asymmetric unit [because you've used ethyl-mercury phosphate, the pH is 5.5 (and so, EMP probably does not hit histidines) and you know that you only have two cysteines per crystallographic asymmetric unit]. So, the question you would like to answer is :

> *Are the observed isomorphous differences consistent with a Patterson function which only contains a number of peaks that can fully be accounted by a two-atom structure* [1] *?*

---

[1] Note that is not a question of purely academic interest. In every-day practice we actively explore answers to this question : we examine the list of largest differences for possible outliers, we attempt excluding weak data from the calculation, or we try changing the resolution limits, etc. In all cases, our criterion is whether the new maps agree better with our expectations.

Although this is a very well-informed (and valid) question, it is not the question that MAXENT answers. Actually, it is a long-sought objective of this class of methods to be able to incorporate such prior knowledge (when available) into the calculations. Unfortunately, the program that I am distributing can not help you answering such knowledgeable questions. But before discussing what is the question that MAXENT really answers, let me elaborate somewhat on this sentence about "the observed iso-morphous differences being consistent with a map containing ...".

0.2 "I thought that the data are consistent with just one map, aren't they ?"

If you measured a 100% complete, error-free data set extending to sufficiently high resolution then, yes, there is a one-to-one correspondence between the data and the map, and the way to calculate the map from the data is through a Fourier transformation. But when the data are incomplete and noisy, this one-to-one correspondence no longer holds. This point is so important, that the rest of this long section is devoted to convincing you about the following statement : "incomplete and noisy data define not a single map, but a whole set of maps, each of which is statistically consistent with the results of your experiment". In the danger of becoming repetitious : when you do an FFT to go from your data to a map, you assume that you have a 100% complete, error-free data set.

Let me give you an example : suppose that you calculate a Patterson function using a data set that is only 70% complete. All reflections (30%) that are missing from the data set (because there were never measured) enter the calculation with an amplitude of zero. The final map will reproduce exactly all these zero amplitudes, as if the data were indeed measured and found to be of zero amplitude. Indeed, if I was giving you not the data, but the map, there would be no way for you to decide whether a reflection with an amplitude of zero was measured to be zero, or was not measured at all. I hope you will agree that there is a significant difference between an unknown amplitude and an amplitude found to be zero (which, by the way, may be almost as informative as a very strong reflection).

Furthermore (and because the data are assumed to be error-free), the final map will reproduce exactly the amplitudes of all of your reflections, without taking into account their standard deviations : Suppose that you have measurements for two reflections, both of which were estimated to have an amplitude of $1000\ e^-$, but the first one is a beautifully measured datum with a standard deviation of only $1\ e^-$, while the other is a lousy measurement with a standard deviation of $500\ e^-$. In the case of the classical (conventional) map these two reflections will contribute to your density map with an equal amplitude of $1000\ e^-$. This does not sound very convincing : you could probably bet your next salary that the amplitude of the first reflection is no less than 950 and no greater than $1050\ e^-$, but would you be prepared to do the same for the second reflection ? Shouldn't the density map reflect the information content (or the trust we

place upon) the various measurements ? To make this even more clear : if at a critical point in your density map (where you would expect to find a strong density feature), these two reflections contribute with opposite signs, so that the good measurement supports the presence of density, whereas the bad measurement cancels the contribution from the good measurement, would you be prepared to trust the conventional map, and conclude that you were wrong after all, and there is no evidence for the presence of density at that region ?

This leaves us with the following basic problem : if we are not to treat unobserved reflections as if having an amplitude of zero, what values should we be assigning to them ? If we are not to fit exactly the measured amplitudes, how would we chose to deviate from them in any meaningful way ? MAXENT provides a consistent (and, at least for its proponents, meaningful and objective) answer to both of these questions. Which takes us back to where we started from :

0.3   What is the question ?

MAXENT answers the following question : *from all maps that are statistically consistent with the observed data, which map should we be looking at ?* This "statistically consistent" sounds as if we are trying to hide something under the carpet, but this is not so : the consistency with the observed data is judged from the value of chi-squared calculated over the whole data set (a global statistic). If you need more details, see the original papers cited in the program's documentation, or the corresponding paper.

0.4   What is the answer ?

The MAXENT answer is : *from all maps that are consistent with the data, the map that we should be looking at, is the one for which the configurational entropy reaches a maximum.* Because the configurational entropy is a measure of the amount of information contained in the map (with a uniform map being the most uninformative and having the greatest entropy), the following definition is also valid :

> *The MAXENT map is the most uninformative (uniform, unstructured) map consistent with the data.*

Because of this property, if *there is* some structure in the MAXENT map, we can safely conclude that the data contain evidence supporting the presence of the observed features. Which means that

> *The MAXENT map only contains features for which there is evidence in the data.*

I hope you will agree that this last proposition is a very reasonable one indeed : The map that we want to look at, is the one which minimises the probability of misinterpreting it. If the map only contains features for which there is evidence in the observed

3

data (and no additional features which arise from the inversion procedure), then this is the map that we want. Which brings me back to the pragmatists : MAXENT aims for a map that minimises the probability of misinterpreting it, and in this way, *also maximises the probability of interpreting it correctly.* The point is, of course, that for most of us the word "interpretability" carries with it a rather vague (and, may I say, sensational) problem- and human-specific quality that makes us think that "interpretability" is not equivalent to "non-misinterpretability".

All this sounds very philosophical, so allow me to illustrate what I mean with an example. Suppose that you have collected anomalous difference data for one of your derivatives, but due to time limitations, you had to collect your data set fast. If the data turn out to be so weak that even a uniform map would be statistically consistent with them, MAXENT will tell you exactly this : "The data are so weak, that even a uniform (uninformative) map is consistent with them" *and it will stop* (ie. you will get no map at all, because all uniform maps are pretty much the same). Now, most of us would agree that this behaviour indeed minimises the probability of a misinterpretation. But, how many of us would call this result a "successful interpretation" ? The whole point of MAXENT is that it is indeed the correct interpretation, but, a correct interpretation of the data that have been measured [and not of the structure of the anomalous scatterers (as you had hoped)]. In plain words, MAXENT will prefer returning a "sorry, try again" message when the data are so weak that you can not confidently identify any signal, instead of attempting to give you a map showing features that are not required by the data. This —at least for the proponents of the method— is not just good science, it is common sense. Let me re-iterate that this is not to imply that any prior knowledge that we have about the problem in hand should be ignored. On the contrary : if we know that the anomalous Patterson ought to contain the origin peak plus a number of peaks expected from, say, a three-atom structure, then the correct thing to do is to incorporate this prior knowledge in the calculation. As already said, the program that I am distributing can not help you performing such a calculation.

0.5 A guide to answerable questions.

Going back to the (Patterson function) example discussed in the first section of this document, instead of an answer to your original question :

> *Are the observed isomorphous differences consistent with a Patterson function which only contains a number of peaks that can fully be accounted by a two-atom structure ?*

you will get an answer to the question :

> *Which Patterson function map only shows features for which there is evidence in the observed isomorphous differences ?*

Whether the map is consistent with our expectations is left for us to decide. Whether a map fully consistent with our expectations is also consistent with the data, we *will*

never know unless this map is also the one for which the configurational entropy reaches a maximum.

GraphEnt assumes that no prior information is available for the inversion problem in hand, and in this way (i) fails to answer knowledgeable questions, but, (ii) it preserves the one-to-one correspondence between the data and the map : the same data will produce the same map whether you are expecting a protein-like map, a 2-atom difference Patterson function, or a 20-atom anomalous Patterson function. The decision about whether you have asked the right question still rests with you.

0.6 End notes

The ideas presented in this short note are not published, have not been subjected to peer review, and for what you know, it may all be rubbish. This is not to imply that they are my ideas : most of it (if not all) is based on the written (and published work) of several people (which, nevertheless, are not responsible for my misunderstandings). If you are interested in reading more about the subject, the site entitled "Probability Theory As Extended Logic" (at the Washington University in St. Louis) is definitely worth visiting at http://bayes.wustl.edu/.

Corrections, Comments, suggestions and flames are gratefully received. A .pdf version of this document is also available *via* http://origin.imbb.forth.gr/~glykos/-GraphEnt_faq.pdf.