Comments on the paper

"Of Ants and Voters: Maximum entropy prediction of agent-based models with recruitment" by S. Barde

Zakaria Babutsidze
SKEMA Business School and OFCE

The paper by Sylvain Barde presents the explorations into the powers of a novel technique (to economics) called Maximum Entropy (MaxEnt hereafter). The methodology was introduced to economics by Foley (1994), but to the present day its potential is largely untapped. MaxEnt allows predicting solutions to agent-based models analytically. The previous use of methodology has been in image reconstruction, where predictions are made about the original image based on the noisy signal at hand. The approach has a great potential on reducing computational time required to run full-fledged agent-based models that are very often NP-difficult.

A particularly intriguing feature of the methodology is that time is implicitly embedded in it. This might not be important in image reconstruction but it is very important in economics as it allows to predict not only the time invariant/equilibrium solution to the model but also to describe the transitional path to it.

In previous paper (Barde 2012) the sufficient conditions for the applicability of the methodology have been derived. The same paper has applied the MaxEnt methodology to Schelling's (1969, 1971) model of segregation. It has been demonstrated that MaxEnt is powerful with respect to the models with fixed proportion of distinct populations.

In current paper the methodology is applied to two models with recruitment. These are the models of ant behavior by Kirman (1993) and that of language competition by Abrams and Strogatz (2003). The distinction with respect to the previous application is that recruitment allows the proportion between the (competing) populations to vary. The properties of the two models discussed are well known. In light of this, the performance of the methodology is tested on different time horizons. Using rigorous computational approach it is demonstrated that, similar to the previous application to Schelling's model of segregation in Barde (2012), MaxEnt performs very well in case of present two models with recruitment. This is true especially for the short-term predictions where initial conditions influence the outcome greatly (which is equivalent to noisy signal containing large chunk of undistorted information).

Let me outline a methodology to assess the powers of MaxEnt that the author follows closely with one exception on which I will concentrate below.

A researcher starts from the theoretical model which we can solve numerically using ABM. She uses general Monte-Carlo approach to generate the development paths implied by the theoretical model from numerous random initial conditions. These development paths are traced all the way to the relevant time-invariant/equilibrium distribution. This is the problem that is computationally expensive for virtually every relevant economic or social model.

In parallel to this, a researcher writes down the statistical model that is based on underlying theoretical model. Further, this statistical model is solved for the transitional path and equilibrium distribution. The solution can be analytic, however this is usually not feasible. Therefore, numerical methods are involved in solution. The distinction from the ABM approach, however, is that this does not require Monte-Carlo simulations over large set of initial conditions (that is already taken care of by the statistical model). Hence it substantially cuts down the computational time.

Further, the two equilibrium paths and resulting equilibria can be compared in order to judge upon the accuracy of MaxEnt predictions.

As mentioned earlier the author in current paper follows the methodology closely. The transitional dynamics and equilibria are derived properly though ABM. He also succeeds writing down the corresponding statistical models in case of both models. However, for solving statistical models arbitrary simplifications are made. In articular, in Kirman's (1993) model the author uses the limit density derived by Alfarano and Milakovic (2009). But, in process of solution he replaces the diffusion term in the statistical model by simple random walk. In Abrams and Strogatz's (2003) model he assumes that the probability of two agents speaking the same language is normally distributed over the distance between the agents in order to model special correlations statistically.

Both of these simplifications are necessary for numerical tractability of statistical model. However, none of them stem from respective theoretical models and, therefore, are arbitrary. In both cases the author shows that despite these simplifications the predictions derived from MaxEnt methodology are accurate. But, arbitrariness of these simplifications casts doubt on the applicability of the methodology on larger scale.

The merit of the methodology is that it allows a researcher to derive the approximation of the solution in considerable shorter time. This is only useful in cases where ABM formulation of the problem is NP-difficult and solving it in real time is not feasible. In contrast to the evaluation exercises that the author has performed in present paper, when a researcher really needs to use MaxEnt she will not have the actual ABM solution to check the accuracy of MaxEnt.[1] Then if she would have to make arbitrary simplifications in the statistical model in order to derive MaxEnt predictions she will have absolutely no guarantee that the prediction at hand has theoretical validity.[2]

In light of this shortcoming it would be very useful if we would have some kind of taxonomy that would match each class of models with types of simplifications that a researcher can make in the process of solving a statistical model without undermining the validity of the MaxEnt predictions. This clearly involves immense amount of work and the methodology of creating such taxonomy is not clear for me at the present moment. However, I am afraid, without such a reference the applicability of the MaxEnt methodology is restricted to the class of models for whom the statistical models can be solved at least numerically *without* simplifications. And, again, based on the simplicity of the two models that we have seen MaxEnt applied to in current paper, I believe this class does not include all that many models.

1. If she did, she would have no need to run MaxEnt in first place.
2. On the other hand, if the statistical model can be solved without simplifications the researcher is on the safe side. However, given the extreme simplicity of two models discussed in this paper, I doubt that any relevant theoretical model would generate the statistical model that would be (at least numerically) tractable.

References

Abrams, D. M. and S. H. Strogatz, 2003. "Modelling the dynamics of language death." *Nature,* 424:900.

Alfarano, S. and M. Milakovic, 2009. Network structure and n- dependence in agent-based herding models." *Journal of Economic Dynamics and Control,* 33:78–92.

Barde, S., 2012. Back to the future: economic rationality and maximum entropy prediction." *University of Kent School of Economics Discussion Paper,* 12(02).

Foley, D. K., 1994. "A statistical equilibrium theory of markets." *Journal of Economic Theory,* 62:321–345.

Kirman, A., 1993. "Ants, rationality and recruitment." *Quarterly Journal of Economics,* 108:137–156.

Schelling, T. C., 1969. "Models of segregation." *American Economic Review,* 59:488–493.

Schelling, T. C., 1971. "Dynamic models of segregation." *Journal of Mathematical Sociology,* 1:143–186.

Reply to Comments

Sylvain Barde

School of Economics, Keynes College, University of Kent, Canterbury, OFCE

The main aim of the paper is to apply the image processing interpretation of the Maximum Entropy (MaxEnt) method to the Kirman (1993) model and the Abrams and Strogatz (20003) voter model as implemented by Stauffer *et al.* (2007). This follows the initial work in Barde 2012 which showed that the Schelling (1969) model of segregation can be predicted with the methodology. The discussant does point out some of the major issues that are associated with the methodology, many of which I agree with. The most important comment is probably the fact that more exploratory work is needed to establish a taxonomy of valid assumptions for corresponding statistical properties. Having said this, I feel that two important clarifications are needed.

My first comment relates to the claim that the assumptions or simplifications required to obtain the MaxEnt solution are arbitrary. Given some data d (the initial condition in agent-based models), the basic formulation for obtaining the prediction μ the maximum entropy problem is given by:

$$\max_{\mu}\left[\alpha S\left(\mu \mid m\right)+\ell\left(d \mid \mu\right)\right]$$

The first part of the expression, $S(\mu \mid m)$ is the relative entropy of with respect to a model m and $\ell(d \mid \mu)$ is the likelihood that the initial condition d is a noisy version of the prediction μ. For any given problem, two terms need to be specified: the model term m and log likelihood $\ell(d \mid \mu)$. While there is an element of 'educated guessing' in specifying these terms, this is not as arbitrary as the discussant claims.

— The model term m is a diffusion term which specifies how far the prediction can stray from initial condition, and this is the term that controls for time in the system. Intuitively, if very little time has elapsed, one should used a very peaked m, as μ will be very close to d. Conversely, long time horizons are represented with a flatter m. It is also important to note that m can have several dimensions, depending on the nature of the problem: one dimensional for the ants model, two dimensions for the Schelling and voter models.

— The likelihood term ℓ depends on the nature of the path linking the initial condition to the predicted state of the system. The image-reconstruction algorithm treats μ as the true image to be discovered and d as a noisy version of μ. This time-reversed path is conditioned on the fact that if the sequence of actions taking the system from its initial condition to its equilibrium distribution is best-response (a common assumption in economics), then the reverse path is effectively a noise process. The likelihood term is therefore determined by knowledge of the updating process, which determines the implicit noise process in the reversed path.

Both these terms are determined from the updating rules of the system, and are therefore not as arbitrary as it may seem. It is true that if little information is available (for instance if the exact transition probabilities are unknown), they must be approximated. For instance, in the generic version used for the voter model, both a gaussian likelihood $\ell(d \mid \mu)$, *i.e.* a gaussian noise process, and gaussian correlations over two-dimensional space for the model term m are assumed as an approximation. However this can be refined if more information is available from the updating process. This is the case in the ants model, where the transition probabilities are well known. In this case the model term is the diffusion of a stopped random walk rather than a gaussian diffusion and the likelihood is designed directly from a path integral of the transition probabilities.

Clearly, MaxEnt is no miracle solution: if the researcher has no information about the dynamic updating process of a system, then there is no way that knowledge of the initial condition alone can lead to a decent prediction of future states. In the Kirman ant model, for instance, the initial condition at $t = 0$ is simply a value $x \in [0,1]$ representing the share of ants of a certain colour. If the researcher is ignorant of the recruitment mechanisms, then x alone does not provide much information on the stable distribution of the system at a later time $t = n$. The central argument for using MaxEnt in the context of agent-based models is precisely that the updating rules of the system are known *ex ante*, as they are provided by the researcher.

My second comment is would be that the aim of the methodology is not to replace the traditional Monte-Carlo methods used in agent-based models but instead to provide a complement. The methodology is analytical in so far as the derivation of the maximum entropy problem is obtained from a rigourous Bayesian approach however, as mentioned by the discussant, in most cases a numerical methodology is required to solve for the solution of the problem. Furthermore, as pointed out by the discussant, the three simple models analysed so far with MaxEnt are a far cry from the complex systems routinely used in the agent-based literature. So given this, what is the usefulness or purpose of the proposed methodology?

An important application in my opinion is to provide a tool for categorising types of agent-based models according to the strength of their convergence to a stable distribution. A key finding of the paper, as well as the companion work on the Schelling model is that while the three models are clearly stochastic, the fact that they are amenable to MaxEnt prediction reveals that they are much more predictable that one might think. In technical terms, this is related to the fact that the image reconstruction MaxEnt algorithm works only if one is able to treat the reversed time-evolution of the system as a noise process, indicating that the time-evolution is in fact a finite improvement path. I agree with the discussant that more work is needed

In the future, rather than providing a direct solution tool for large agent-based model, a potentially important application for MaxEnt is the prediction of those component modules of the larger model that are amenable to MaxEnt. In interesting possibility in this regard is to take advantage of the faster execution speed of the methodology compared to Monte-Carlo to directly provide agents in the model with expectations, by using MaxEnt on the current state to obtain predicted future values for key state variables. Similarly, it could be used to speed-up large agent-based models by using the faster MaxEnt method on those components that are known to be amenable to the methodology.

References

Abrams D. M. and S. H. Strogatz, 2003. "Modelling the dynamics of language death." *Nature*, 424:900.

Barde S., 2012. "Back to the future: economic rationality and maximum entropy prediction." *University of Kent School of Economics Discussion Paper* 12(02).

Kirman A., 1993. "Ants, rationality and recruitment." *Quarterly Journal of Economics*, 108:137–156.

Schelling T. C., 1969. "Models of segregation." *American Economic Review*, 59:488–493.

Stauffer D., X. Castelló, V. M. Eguíluz, and M. San Miguel, 2007. "Microscopic Abrams-Strogatz model of language competition." *Physica A*, 374:835–842.