Email 1: Sanjay

Dear colleagues,

I have begun to read the econophysics literature, which many of you have favourably referred to over the years, with more care recently. I have been provoked to do this by my own interest in income distribution and the hope that one might derive some insights from that literature for the development of parsimonious models thereof, as a result of my students' recent and continuing inquiries to me based upon their encounters with ideas of entropy maximization and statistical equilibrium in your classes, and of course my encounters with your related research in our department seminars and some of your writings (although I must admit to having been quite remiss in reading these).

I must also admit that when I first studied statistical mechanics as an undergraduate more than 25 years ago I also did very much have the idea in mind that it could be applied fruitfully to economics, but I did not have sufficient perspective then to see quite how. In the intervening period, I have unhappily forgotten much of the details at I once knew, but of course some exposure to the point of view still remains.

I would like to ask you a couple of questions of clarification, for which purpose it maybe helpful for me to refer to the following article by Banerjee and Yakovenko, although only because it provides a useful reference point for the queries, rather than because it contains any specific insights not otherwise available (much of Yakovenko's work in this area appears overlapping, inter alia):

http://physics.umd.edu/~yakovenk/papers/2010-NJP-v12-n7-p075032.pdf

I do apologize if my questions should seem to you to be elementary.

First of all, let me note that in the article the assumption of entropy maximization is not particularly well-motivated.

Likening an income distribution to a probability distribution, why should we make the assumption that its 'information entropy' is maximized? One answer, and that provided by the authors on p.4, following the 'Wallis derivation' of the entropy maximization principle (see https://en.wikipedia.org/wiki/Principle_of_maximum_entropy) is as follows: The assignment of probabilities across possible income bins that maximizes entropy is that which can be brought about through a larger number of possible allocations of individual persons across bins than can any other. If every such allocation of persons is deemed equally probable, by analogy to the analysis of an ensemble of possible states of a system in statistical physics, or indeed in information theoretic approaches which assume that in the absence of further information one should make an equiprobability assumption across possible outcomes, then this maximum entropy assignment of probabilities is the one which is more likely than any other. [Call this argument A*].

Fair enough, but is this a sufficient reason to assume this particular assignment to be that which is 'likely' in some sense to prevail? 'More likely' is after all not the same thing as 'likely', and there is no guarantee that the largest entropy assignmentl in any sense has a high absolute likelihood. In my undergraduate textbook (F. Reif, Fundamentals of Statistical and Thermal Physics, 1965) which I still have and which I partially reread this morning, it is said (Page 91) that, "Usually [the number of states accessible to the system] has a very pronounced maximum at some value [of one of the parameters]. In that case practically all systems in the final equilibrium situation will have values corresponding to the most probable situation, where [the parameter] is very close to [the level of the parameter that maximizes the number of states accessible to the system]". In the text, there's an accompanying indicative diagram showing the number of states accessible to the system as a function of the parameter, which has a Dirac-delta like shape. Is there any assurance in general economic systems that the corresponding diagram is not more diffuse?

Now, let me turn to the Yakovenko et al paper specifically. On p.4 he derives an exponential distribution (equation (6)) along the lines above, reflecting the conservation of total income in the way that the Boltzmann distribution in statistical physics reflects constancy of temperature. The maximization of entropy procedure undertaken there appears to be implicitly motivated by an argument along the lines of A*. Presumably, by analogy to physics, this is a case in which it is formally demonstrable that "practically all systems in the final equilibrium situation will have values corresponding to the most probable situation", in the sense that the vast majority of possible allocations of income across persons that respect the conservation of aggregate income constitute distributions that are "close to" the exponential distribution derived on the basis of entropy maximization.

Later on in the paper, however, the authors introduce what appear to be presented as a wholly independent "kinetic" approach (p.10), which involves different possible diffusion dynamics characterizing how distinct incomes, r, change by delta r over a time period delta t. They then apply the Fokker-Planck equation as an entirely general (and seemingly otherwise atheoretical) approach to characterizing the evolution of a probability distribution under an assumed diffusion process governing changes in the distribution over time, and they apply a steady-state ("stationary distribution") assumption to derive the compatible class of distributions. [Call this alternate argument as to what distribution to expect B*].

The possible further assumption of additive diffusion [call this case B*1] gives rise to the exponential distribution (previously derived by applying A*). The alternate further assumption of multiplicative diffusion [call this case B*2] gives rise to a Gibrat-Pareto power law distribution. It would appear that A* is compatible with B*1 but not with B*2. How should we interpret this? Is this because multiplicative dynamics (either alone or in conjunction with additive dynamics) necessarily entails changing, and not constant, total income of the society?

In any case, the generality of the Folker-Planck equation approach would seem to allow for many more possible specifications of the dynamics beyond B*1 and B*2 both within the steady-state framework and outside of it. Have these various further possibilities been explored in the literature and what are they? More generally, what is the relationship between A* and B*? One of our colleagues put it to me that there is an "equivalence" but is that really the case? [What would it mean to say that, in any case? Here is a possible candidate formulation of his proposition: there is a one-to-one mapping between possible specification of the diffusion dynamics of the Fokker-Planck equation and specifications of the constraints under which entropy is to maximized].

Email 2: Duncan to Sanjay

Duncan:

You raise a series of important questions, many of them not fully resolved as far as I can tell even among statistical physicists who use statistical equilibrium reasoning in a huge variety of contexts, and are very concerned about the foundational questions you raise.

One experience I had that helped me with these questions was discussions with Murray Gell-Mann in Santa Fe, where the issue of the explanatory status of statistical equilibrium is under constant (almost obsessive) discussion, usually in the form of such questions as "what is the entropy of the universe?" Gell-Mann says something like: "physicists do a very strange thing in studying complex systems with many degrees of freedom like the dynamics of a diffuse gas (where the state of the system at any instant is described by 6N numbers, where N is the number of particles the gas contains and the 6 numbers consist of 3 coordinates describing the position of the particle and 3 coordinates describing its momentum — mass times velocity). This appears to be a very difficult, perhaps practically impossible, problem, and what do physicists do? They think of what at first appears to be a much more difficult problem, which is to study not just the behavior of this particular gas, but of the collection (ensemble) of all possible systems that share certain properties with the one we are actually interested in. Remarkably, it turns out that it is much easier to draw strong and meaningful conclusions about the statistical properties of the ensemble than it is to draw conclusions about the dynamic behavior of the particular system, and, even more remarkably, those statistical conclusions agree with experimental observations in a huge range of contexts to a stunning precision." I think this is the main lesson to keep in mind methodologically.

It leads to maximum entropy reasoning in an important range of cases, but not always. (For one thing there are other entropies, functions that share mathematical properties with Shannon entropy, and for some data maximizing one of these other entropies leads to better explanations of the data.)

To my mind the fundamental issue with statistical equilibrium is pretty well summed up by considering the question of the relation of statistical equilibrium to mechanics. There are two arguments here, one a Laplacian argument based on Newtonian mechanics, and a second one that considers quantum mechanics.

If we look at the universe as Laplace did, it appears to be an ensemble of atoms each with a position and momentum, all subject to Newton's laws of motion. Laplace famously said that if you could tell him the exact state of the universe, that is, the position and momentum of all these particles, at any instant, he could in principle compute the entire past and future of the universe by applying Newton's laws. One fundamental property of Newton's laws is that they are time-reversible. Thus in a certain sense from Laplace's point of view the entropy of the universe has to be invariant over time, or, to put this another way, there is no "arrow of time", which is the fundamental concept behind the Second Law of Thermodynamics (increase of entropy).

If we look at the universe from quantum theory, it appears (to Murray Gell-Mann, for example) as a single, highly entangled, quantum "state" with very many dimensions. But quantum theory, despite the limits it puts on the simultaneous measurement of position and momentum, predicts that state transitions over time are "unitary transformations" of this complex state, and unitary transformations have the same property of time-reversibility. Gell-mann, when he is confronted on this, says that in principle there is no change in the microscopic entropy of the universe over time, and that the second law of thermodynamics is a result of "coarse graining" because we can only measure states of the universe with a certain precision, and only a limited range of the entangled state variables.

This is, as far as I know, where the matter sits, at least out in Santa Fe.

But these conversations suggest to me that the questions you raise at a fundamental level are not resolved even in the rarefied theoretical discussions of high-level statistical physics.

Some responses to your specific questions below:

Sanjay:

I would like to ask you a couple of questions of clarification, for which purpose it maybe helpful for me to refer to the following article by Banerjee and Yakovenko, although only because it provides a useful reference point for the queries, rather than because it contains any specific insights not otherwise available (much of Yakovenko's work in this area appears overlapping, inter alia):

http://physics.umd.edu/~yakovenk/papers/2010-NJP-v12-n7-p075032.pdf

Duncan:

Yakovenko is unusual among econo-physicists in his interest in the political-economic aspects of these problems. His empirical work seems to me very carefully done. He does have certain blind spots, such as the willingness to import constraints that are meaningful in physical problems into economic contexts where they are not relevant. The leading example in my mind is the assumption of "detailed balance", which is a symmetry hypothesis that two interacting particles, say, in a gas, are equally likely to transfer energy in either direction, which Yakovenko imports into models of monetary exchange in the completely implausible form of assuming that there is an equal probability of economic agent A transferring money to or from B. I've pointed this out to Yakovenko many times over the years, but without making much of an impression.

Sanjay:

First of all, let me note that in the article the assumption of entropy maximization is not particularly well-motivated.

Duncan's Response:

I'm not sure I agree with this.

All theories that try to explain observed quantitative data explain the data as a combination of central tendencies (law-like regularities, or signal) and unexplained residual variation (errors, or noise). For example, regression analysis looks for a linear relation among the observed variables and regards the deviations of the data from the linear model as "errors", and then ranks different linear models depending on properties of the errors, such as unbiasedness, or approximate Gaussian distribution. We could think of this as making a guess as to the structural relations of the variables, and then trying to evaluate the quality of the resulting model by studying the properties of the errors.

Jaynes' program inverts this methodological procedure by beginning by maximizing the randomness of the residuals according to the principle of maximum entropy. It is important to understand the discovery of statistical physics that "randomness" means "maximum entropy", not necessarily some other intuitively appealing concept like "i.i.d. Gaussian residuals". (Sometimes maximizing entropy leads to i.i.d Gaussian residuals, but often it does not. Maximizing entropy over the entire positive and negative real line subject to a 0 mean and variance constraint does yield the Gaussian normal distribution. Maximizing entropy over a bounded interval of the real line does give the intuitive uniform distribution. But a small change in constraints, like maximizing entropy over the positive half line subject to a mean constraint does not yield i.i.d., but the highly skewed exponential distribution.)

Jaynes then locates the theoretical content of a model in constraints on the maximum entropy problem. This is very characteristic of the way physicists do empirical modeling, and totally foreign to the "shoot in the dark and check the residuals" methods that dominate econometrics. Jaynes sees this as an iterative process in which you try to explain the data without constraints (usually resulting in a very poorly fitting model, signaled by the fact that the entropy of the observed data is much smaller than the maximum unconstrained entropy on the relevant state space), and then hope you can figure out simple constraints that have theoretical content and will bring the maxent model into better agreement with the data. This discovery of constraints is not a question of turning a crank on some hypothesis machine, but is where the real insight of scientific discovery resides. For example, if you maximize entropy of a gas subject only to constraints on the number of particles and the volume of the container, you get a very poor theory that is at gross odds with observation. It took decades for physicists to realize that the key to understanding these systems was the conservation of energy (proportional to the square of momentum, and therefore an inherently nonnegative quantity). When you maximize entropy subject to the number of particles, the volume of the container and the total or mean energy, voilá, you get the Maxwell-Boltzmann exponential distribution and you are in business with a theory that is in very good agreement with observation (though inevitably not with every observation, since real gases are not exactly perfect gases).

This does not, of course, mean that the particles in the gas are "trying" to maximize entropy. They are just knocking around colliding with each other and with the walls of the container. As far as each particle (if we could, as economists, speak that way) knows, the world is just Newton's laws. No individual particle ever gets to an energy level where the constraint on the energy of the whole system begins to affect it directly. But because energy is conserved, the distribution of energy among the particles does follow the maximum entropy distribution very closely, giving rise to the macro-phenomena such as pressure and temperature that characterize the behavior of the gas.

This does not work all the time. If, for example, you have a perfect billiard ball moving on a frictionless surface between completely elastic walls, and you start that ball on a trajectory exactly perpendicular to the walls with a positive velocity, it will continue to bounce back and forth on exactly the same trajectory forever. An ensemble of balls moving in parallel will do the same thing, and you will not see any increase in entropy due to their bumping into each other or veering one way or another. But this fanciful example underlines just why it is so "difficult" to avoid the maximum entropy configurations of a system with many degrees of freedom. Jaynes would say that it you manage to keep the system at a lower entropy, you are somehow imposing a constraint on it, and that the proper theory for such a system needs to incorporate that constraint. (I think this is very relevant to the problem of income, wage, and wealth distributions.)

Sanjay:

Likening an income distribution to a probability distribution, why should we make the assumption that its 'information entropy' is maximized? One answer, and that provided by the authors on p.4, following the 'Wallis derivation' of the entropy maximization principle (see https://en.wikipedia.org/wiki/Principle_of_maximum_entropy) is as follows: The assignment of probabilities across possible income bins that maximizes entropy is that which can be brought about through a larger number of possible allocations of individual persons across bins than can any other. If every such allocation of persons is deemed equally probable, by analogy to the analysis of an ensemble of possible states of a system in statistical physics, or indeed in information theoretic approaches which assume that in the absence of further information one should make an equiprobability assumption across possible outcomes, then this maximum entropy assignment of probabilities is the one which is more likely than any other. [Call this argument A*].

Duncan:

Here we are going to run across the fact that different mathematical arguments can motivate the same result. The article you link to seems to me quite good on the whole, but be careful to realize that the derivation of maximum entropy from the multinomial distribution is just one way to do it, and may be relevant in some cases and not in others. (In the case of income distribution, for example, it is clear that arguments from ergodicity, the postulate that the system visits all possible states, are irrelevant, because we know that households do not systematically visit all possible income states. Similarly the buckets argument is not very easy to interpret in that context. The irrelevance of any particular mathematical model of maximum entropy to a particular empirical problem, however, doesn't tell us much about the relevance of the maxent conception to explaining the data.)

Sanjay:

Fair enough, but is this a sufficient reason to assume this particular assignment to be that which is 'likely' in some sense to prevail? 'More likely' is after all not the same thing as 'likely', and there is no guarantee that the largest entropy assignmentI in any sense has a high absolute likelihood. In my undergraduate textbook (F. Reif, Fundamentals of Statistical and Thermal Physics, 1965) which I still have and which I partially reread this morning, it is said (Page 91) that, "Usually [the number of states accessible to the system] has a very pronounced maximum at some value [of one of the parameters]. In that case practically all systems in the final equilibrium situation will have values corresponding to the most probable situation, where [the parameter] is very close to [the level of the parameter that maximizes the number of states accessible to the system]". In the text, there's an accompanying indicative diagram showing the number of states accessible to the system as a function of the parameter, which has a Dirac-delta like shape. Is there any assurance in general economic systems that the corresponding diagram is not more diffuse?

Duncan:

This is an interesting point. In the simplest cases the probability of observing the system in a state described by a frequency distribution p[s] over the feasible states is proportional to exp[N H[p]], which, even when N is much smaller than Avogadro's Number, is indeed very sharply peaked and hard to distinguish from the Dirac Delta function. If we compute entropy with logarithms to base 2, for example, entropy is measured in bits, and each bit of entropy represents a halving (or doubling, depending which way you are going) of the probability.

Sanjay:

Now, let me turn to the Yakovenko et al paper specifically. On p.4 he derives an exponential distribution (equation (6)) along the lines above, reflecting the conservation of total income in the way that the Boltzmann distribution in statistical physics reflects constancy of temperature. The maximization of entropy procedure undertaken there appears to be implicitly motivated by an argument along the lines of A*. Presumably, by analogy to physics, this is a case in which it is formally demonstrable that "practically all systems in the final equilibrium situation will have values corresponding to the most probable situation", in the sense that the vast majority of possible allocations of income across persons that respect the conservation of aggregate income constitute distributions that are "close to" the exponential distribution derived on the basis of entropy maximization.

Duncan:

But as I argued above, this can't be literally true for income distribution data. What the observed exponential shape of the bulk of the income distribution suggests is that in addition to whatever dynamics are moving households or individuals up and down in income, there is a weak statistical effect constraining them to a given total (or average) income. This has the rather extraordinary implication, which I think we are far from understanding well, even given the pioneering work of our students Markus Schneider, Amir Ragab, and Nicholas Papanicolaou, that when one household manages to raise its income, there is a very small (given the number of households) negative effect on the incomes of everyone else.

Sanjay:

Later on in the paper, however, the authors introduce what appear to be presented as a wholly independent "kinetic" approach (p.10), which involves different possible diffusion dynamics characterizing how distinct incomes, r, change by delta r over a time period delta t. They then apply the Fokker-Planck equation as an entirely general (and seemingly otherwise atheoretical) approach to characterizing the evolution of a probability distribution under an assumed diffusion process governing changes in the distribution over time, and they apply a steady-state ("stationary distribution") assumption to derive the compatible class of distributions. [Call this alternate argument as to what distribution to expect B*].

Duncan:

This type of model is a "drift-diffusion" model and is a favorite with physicists. The basic properties of these models for one-dimensional data like income are that the dispersion of individuals in state space, which in this case is the nonnegative real line, constantly increases other things being equal, and that there is a pervasive drift term tending to push everyone toward the origin. When there is a lower bound on the state space (such as zero income) the ergodic distribution of the drift-diffusion system is an exponential distribution (which is rather intuitive) and the exponent is the ratio of the rates of drift and diffusion. This is a nice way to explain the exponential outcome (though for economists perhaps constrained maxent, which uses all of the apparatus of constrained maximization that economics students learn much more about than they do about stochastic differential equations, is pedagogically more direct) in quasi-dynamic terms. But even in the case of the statistical mechanics of a perfect gas, there is no direct derivation of the drift-diffusion model that I know of from Newton's Laws.

Sanjay:

The possible further assumption of additive diffusion [call this case B*1] gives rise to the exponential distribution (previously derived by applying A*). The alternate further assumption of multiplicative diffusion [call this case B*2] gives rise to a Gibrat-Pareto power law distribution. It would appear that A* is compatible with B*1 but not with B*2. How should we interpret this? Is this because multiplicative dynamics (either alone or in conjunction with additive dynamics) necessarily entails changing, and not constant, total income of the society?

Duncan:

Since this is a case of multiple mathematical derivations of the same formal result, the main methodological point is my caution above against identifying the principle of maxent with any particular model that implies a maxent outcome. On the other hand, the difference between additive and multiplicative diffusion makes complete sense from a political economy perspective. It is the difference already present in Adam Smith's discussion between workers pursuing maximum wages (additive) and capitalists pursuing maximum profit rates (multiplicative), and it is striking that distributions of income are mostly (approximately) exponential while distributions of wealth are mostly (approximately) power laws.

Sanjay:

In any case, the generality of the Folker-Planck equation approach would seem to allow for many more possible specifications of the dynamics beyond B*1 and B*2 both within the steady-state framework and outside of it. Have these various further possibilities been explored in the literature and what are they? More generally, what is the relationship between A* and B*? One of our colleagues put it to me that there is an "equivalence" but is that really the case? [What would it mean to say that, in any case? Here is a possible candidate formulation of his proposition: there is a one-to-one mapping between possible specification of the diffusion dynamics of the Fokker-Planck equation and specifications of the constraints under which entropy is to maximized].

Duncan:

It certainly would help if we knew a lot more about the dynamics of the income distribution, for example, from careful analysis of panel data to reconstruct Markov transition matrices (a conceptual approach suggested in the 1950s by Champernowne). In principle if you know the Markov transition matrix you can compute the ergodic distribution and compare it to the data. A couple of cautions are in order, though. Transition probabilities are basically time differentiations of state measurements, and it is notorious in social science data that differentiation greatly increases the noisiness of data. Even with pretty good measures of household income in a panel, the transitions probabilities are likely to have quite big error bars. The drift-diffusion story also highlights the sensitivity of the equilibrium distribution to small variations in the drift and diffusion parameters (since, for example, the exponential coefficient is the ratio of the two). Still, if we can get access to reasonably large-sample panels, why not get some students working on this?

I'm sure these remarks leave many questions unanswered, but they are very much worth further discussion.

Email 3: Sanjay to Duncan

Dear Duncan,

Thank you very much for this helpful response. In order to provide for as clear record-keeping as possible regarding the different points made, I will intersperse my further responses (below in italics) to which you should not at all feel obliged to respond, as we can always discuss in person.

I hope you do not mind that I am again cc-ing the colleagues originally addressed, as they may benefit from the dialogue and/or have additions and correctives to offer.

best wishes,

Sanjay

Duncan:

You raise a series of important questions, many of them not fully resolved as far as I can tell even among statistical physicists who use statistical equilibrium reasoning in a huge variety of contexts, and are very concerned about the foundational questions you raise.

One experience I had that helped me with these questions was discussions with Murray Gell-Mann in Santa Fe, where the issue of the explanatory status of statistical equilibrium is under constant (almost obsessive) discussion, usually in the form of such questions as "what is the entropy of the universe?" Gell-Mann says something like: "physicists do a very strange thing in studying complex systems with many degrees of freedom like the dynamics of a diffuse gas (where the state of the system at any instant is described by 6N numbers, where N is the number of particles the gas contains and the 6 numbers consist of 3 coordinates describing the position of the particle and 3 coordinates describing its momentum — mass times velocity). This appears to be a very difficult, perhaps practically impossible, problem, and what do physicists do? They think of what at first appears to be a much more difficult problem, which is to study not just the behavior of this particular gas, but of the collection (ensemble) of all possible systems that share certain properties with the one we are actually interested in. Remarkably, it turns out that it is much easier to draw strong and meaningful conclusions about the statistical properties of the ensemble than it is to draw conclusions about the dynamic behavior of the particular system, and, even more remarkably, those statistical conclusions agree with experimental observations in a huge range of contexts to a stunning precision." I think this is the main lesson to keep in mind methodologically.

Sanjay's Response to Duncan: *(New responses Italicized)*

How general is this lesson? I think that this is the question we must answer if we wish to generalize the approach of statistical dynamics to other (e.g. social scientific) settings. For example, it seems a reasonable thought that the improbable simplification achieved by considering the entire ensemble of systems is a feature of particular kinds of dynamic processes (e.g. those involving some version of Brownian motion or other random processes in which local evolution in all directions is equally possible, and for which it is therefore reasonable to consider all possible states to be equally probable). If so, then the lesson may not be particularly general.

Duncan:

It leads to maximum entropy reasoning in an important range of cases, but not always. (For one thing there are other entropies, functions that share mathematical properties with Shannon entropy, and for some data maximizing one of these other entropies leads to better explanations of the data.)

To my mind the fundamental issue with statistical equilibrium is pretty well summed up by considering the question of the relation of statistical equilibrium to mechanics. There are two arguments here, one a Laplacian argument based on Newtonian mechanics, and a second one that considers quantum mechanics.

If we look at the universe as Laplace did, it appears to be an ensemble of atoms each with a position and momentum, all subject to Newton's laws of motion. Laplace famously said that if you could tell him the exact state of the universe, that is, the position and momentum of all these particles, at any instant, he could in principle compute the entire past and future of the universe by applying Newton's laws. One fundamental property of Newton's laws is that they are time-reversible. Thus in a certain sense from Laplace's point of view the entropy of the universe has to be invariant over time, or, to put this another way, there is no "arrow of time", which is the fundamental concept behind the Second Law of Thermodynamics (increase of entropy).

If we look at the universe from quantum theory, it appears (to Murray Gell-Mann, for example) as a single, highly entangled, quantum "state" with very many dimensions. But quantum theory, despite the limits it puts on the simultaneous measurement of position and momentum, predicts that state transitions over time are “unitary transformations” of this complex state, and unitary transformations have the same property of time-reversibility. Gell-mann, when he is confronted on this, says that in principle there is no change in the microscopic entropy of the universe over time, and that the second law of thermodynamics is a result of “coarse graining” because we can only measure states of the universe with a certain precision, and only a limited range of the entangled state variables.

Sanjay’s Response to Duncan:

The issue of the reversibility or irreversibility of processes is indeed a deep question in physics, with far from a clear resolution, as you point out. [On this see also the recent book by by Roberto Unger and Lee Smolin, The Singular Universe and the Reality of Time]. I do not see, personally, how the corresponding problem in the social world can be approached without introducing the role of consciousness, through which "time's arrow' is subjectively as well as objectively manifested. This is part of what I find unsatisfactory about overly scientistic approaches to social enquiry, that abstract altogether from subjective self-awareness or consider it epiphenomenal. Even insofar as we are dealing with narrowly economic data in a domain such as that of income distribution it will manifest itself, for instance in the form of a desire to protect wealth and to accumulate more of it, which provides a definite gradient to economic activities, so that the evolution of individual wealth is not simply the consequence of equidirectional "brownian motion". One can view this motive in terms of "rational" action or even "maximization" if one wishes, or not. It doesn't matter deeply. What matters is that intentional action and random action are different things.

Duncan:

This is, as far as I know, where the matter sits, at least out in Santa Fe.

But these conversations suggest to me that the questions you raise at a fundamental level are not resolved even in the rarefied theoretical discussions of high-level statistical physics.

Some responses to your specific questions below:

Sanjay:

I would like to ask you a couple of questions of clarification, for which purpose it maybe helpful for me to refer to the following article by Banerjee and Yakovenko, although only because it provides a useful reference point for the queries, rather than because it contains any specific insights not otherwise available (much of Yakovenko's work in this area appears overlapping, inter alia):

http://physics.umd.edu/~yakovenk/papers/2010-NJP-v12-n7-p075032.pdf

Duncan:

Yakovenko is unusual among econo-physicists in his interest in the political-economic aspects of these problems. His empirical work seems to me very carefully done. He does have certain blind spots, such as the willingness to import constraints that are meaningful in physical problems into economic contexts where they are not relevant. The leading example in my mind is the assumption of “detailed balance”, which is a symmetry hypothesis that two interacting particles, say, in a gas, are equally likely to transfer energy in either direction, which Yakovenko imports into models of monetary exchange in the completely implausible form of assuming that there is an equal probability of economic agent A transferring money to or from B. I’ve pointed this out to Yakovenko many times over the years, but without making much of an impression.

Sanjay’s Response to Duncan’s Response:

This would appear to be precisely a specific example of the salience in modeling of the distinction between random and intentional action.

Sanjay:

First of all, let me note that in the article the assumption of entropy maximization is not particularly well-motivated.

Duncan's Response:

I'm not sure I agree with this. [Sanjay's Response: *I was in the first instance commenting on the Banerjee and Yakovenko paper I cited. Perhaps it is also not particularly well-motivated generally (and indeed, I suspect this, although I did not claim it here).*]

All theories that try to explain observed quantitative data explain the data as a combination of central tendencies (law-like regularities, or signal) and unexplained residual variation (errors, or noise). For example, regression analysis looks for a linear relation among the observed variables and regards the deviations of the data from the linear model as "errors", and then ranks different linear models depending on properties of the errors, such as unbiasedness, or approximate Gaussian distribution. We could think of this as making a guess as to the structural relations of the variables, and then trying to evaluate the quality of the resulting model by studying the properties of the errors.

Jaynes' program inverts this methodological procedure by beginning by maximizing the randomness of the residuals according to the principle of maximum entropy. It is important to understand the discovery of statistical physics that "randomness" means "maximum entropy", not necessarily some other intuitively appealing concept like "i.i.d. Gaussian residuals". (Sometimes maximizing entropy leads to i.i.d Gaussian residuals, but often it does not. Maximizing entropy over the entire positive and negative real line subject to a 0 mean and variance constraint does yield the Gaussian normal distribution. Maximizing entropy over a bounded interval of the real line does give the intuitive uniform distribution. But a small change in constraints, like maximizing entropy over the positive half line subject to a mean constraint does not yield i.i.d., but the highly skewed exponential distribution.)

Jaynes then locates the theoretical content of a model in constraints on the maximum entropy problem. This is very characteristic of the way physicists do empirical modeling, and totally foreign to the "shoot in the dark and check the residuals" methods that dominate econometrics. Jaynes sees this as an iterative process in which you try to explain the data without constraints (usually resulting in a very poorly fitting model, signaled by the fact that the entropy of the observed data is much smaller than the maximum unconstrained entropy on the relevant state space), and then hope you can figure out simple constraints that have theoretical content and will bring the maxent model into better agreement with the data. This discovery of constraints is not a question of turning a crank on some hypothesis machine, but is where the real insight of scientific discovery resides. For example, if you maximize entropy of a gas subject only to constraints on the number of particles and the volume of the container, you get a very poor theory that is at gross odds with observation. It took decades for physicists to realize that the key to understanding these systems was the conservation of energy (proportional to the square of momentum, and therefore an inherently nonnegative quantity). When you maximize entropy subject to the number of particles, the volume of the container and the total or mean energy, voilá, you get the Maxwell-Boltzmann exponential distribution and you are in business with a theory that is in very good agreement with observation (though inevitably not with every observation, since real gases are not exactly perfect gases).

This does not, of course, mean that the particles in the gas are “trying” to maximize entropy. They are just knocking around colliding with each other and with the walls of the container. As far as each particle (if we could, as economists, speak that way) knows, the world is just Newton’s laws. No individual particle ever gets to an energy level where the constraint on the energy of the whole system begins to affect it directly. But because energy is conserved, the distribution of energy among the particles does follow the maximum entropy distribution very closely, giving rise to the macro-phenomena such as pressure and temperature that characterize the behavior of the gas.

This does not work all the time. If, for example, you have a perfect billiard ball moving on a frictionless surface between completely elastic walls, and you start that ball on a trajectory exactly perpendicular to the walls with a positive velocity, it will continue to bounce back and forth on exactly the same trajectory forever. An ensemble of balls moving in parallel will do the same thing, and you will not see any increase in entropy due to their bumping into each other or veering one way or another. But this fanciful example underlines just why it is so “difficult” to avoid the maximum entropy configurations of a system with many degrees of freedom. Jaynes would say that it you manage to keep the system at a lower entropy, you are somehow imposing a constraint on it, and that the proper theory for such a system needs to incorporate that constraint. (I think this is very relevant to the problem of income, wage, and wealth distributions.)

Sanjay’s Response to Duncan:

This is certainly a powerful methodological framework, insofar as one accepts the entry maximization starting point, but it may leave more questions than it answers: which constraints, when and why?

Sanjay:

Likening an income distribution to a probability distribution, why should we make the assumption that its 'information entropy' is maximized? One answer, and that provided by the authors on p.4, following the 'Wallis derivation' of the entropy maximization principle (see https://en.wikipedia.org/wiki/Principle_of_maximum_entropy) is as follows: The assignment of probabilities across possible income bins that maximizes entropy is that which can be brought about through a larger number of possible allocations of individual persons across bins than can any other. If every such allocation of persons is deemed equally probable, by analogy to the analysis of an ensemble of possible states of a system in statistical physics, or indeed in information theoretic approaches which assume that in the absence of further information one should make an equiprobability assumption across possible outcomes, then this maximum entropy assignment of probabilities is the one which is more likely than any other. [Call this argument A*].

Duncan:

Here we are going to run across the fact that different mathematical arguments can motivate the same result. The article you link to seems to me quite good on the whole, but be careful to realize that the derivation of maximum entropy from the multinomial distribution is just one way to do it, and may be relevant in some cases and not in others. (In the case of income distribution, for example, it is clear that arguments from ergodicity, the postulate that the system visits all possible states, are irrelevant, because we know that households do not systematically visit all possible income states. Similarly the buckets argument is not very easy to interpret in that context. The irrelevance of any particular mathematical model of maximum entropy to a particular empirical problem, however, doesn’t tell us much about the relevance of the maxent conception to explaining the data.)

Sanjay's Response to Duncan:

I quite agree with you here, but this would seem precisely to illustrate the poverty of the motivation of the maximum entropy principle in the economic context.

Sanjay:

Fair enough, but is this a sufficient reason to assume this particular assignment to be that which is 'likely' in some sense to prevail? 'More likely' is after all not the same thing as 'likely', and there is no guarantee that the largest entropy assignmentI in any sense has a high absolute likelihood. In my undergraduate textbook (F. Reif, Fundamentals of Statistical and Thermal Physics, 1965) which I still have and which I partially reread this morning, it is said (Page 91) that, "Usually [the number of states accessible to the system] has a very pronounced maximum at some value [of one of the parameters]. In that case practically all systems in the final equilibrium situation will have values corresponding to the most probable situation, where [the parameter] is very close to [the level of the parameter that maximizes the number of states accessible to the system]". In the text, there's an accompanying indicative diagram showing the number of states accessible to the system as a function of the parameter, which has a Dirac-delta like shape. Is there any assurance in general economic systems that the corresponding diagram is not more diffuse?

Duncan:

This is an interesting point. In the simplest cases the probability of observing the system in a state described by a frequency distribution p[s] over the feasible states is proportional to exp[N H[p]], which, even when N is much smaller than Avogadro's Number, is indeed very sharply peaked and hard to distinguish from the Dirac Delta function. If we compute entropy with logarithms to base 2, for example, entropy is measured in bits, and each bit of entropy represents a halving (or doubling, depending which way you are going) of the probability.

Sanjay's Response to Duncan:

I will be grateful for a didactic reference. It would be interesting to examine further whether economic applications are likely to fall in the universe of 'the simplest cases'.

Sanjay:

Now, let me turn to the Yakovenko et al paper specifically. On p.4 he derives an exponential distribution (equation (6)) along the lines above, reflecting the conservation of total income in the way that the Boltzmann distribution in statistical physics reflects constancy of temperature. The maximization of entropy procedure undertaken there appears to be implicitly motivated by an argument along the lines of A*. Presumably, by analogy to physics, this is a case in which it is formally demonstrable that "practically all systems in the final equilibrium situation will have values corresponding to the most probable situation", in the sense that the vast majority of possible allocations of income across persons that respect the conservation of aggregate income constitute distributions that are "close to" the exponential distribution derived on the basis of entropy maximization.

Duncan:

But as I argued above, this can't be literally true for income distribution data. What the observed exponential shape of the bulk of the income distribution suggests is that in addition to whatever dynamics are moving households or individuals up and down in income, there is a weak statistical effect constraining them to a given total (or average) income. This has the rather extraordinary implication, which I think we are far from understanding well, even given the pioneering work of our students Markus Schneider, Amir Ragab, and Nicholas Papanicolaou, that when one household manages to raise its income, there is a very small (given the number of households) negative effect on the incomes of everyone else.

Sanjay's Response to Duncan:

Is this not more of a tautology than an implication, let alone an extraordinary one, insofar as mandated by the conservation of total income premise? This example does raise in my mind, as others do, to what extent the framework is providing an explanation as opposed to a description.

Sanjay:

Later on in the paper, however, the authors introduce what appear to be presented as a wholly independent "kinetic" approach (p.10), which involves different possible diffusion dynamics characterizing how distinct incomes, r, change by delta r over a time period delta t. They then apply the Fokker-Planck equation as an entirely general (and seemingly otherwise atheoretical) approach to characterizing the evolution of a probability distribution under an assumed diffusion process governing changes in the distribution over time, and they apply a steady-state ("stationary distribution") assumption to derive the compatible class of distributions. [Call this alternate argument as to what distribution to expect B*].

Duncan:

This type of model is a "drift-diffusion" model and is a favorite with physicists. The basic properties of these models for one-dimensional data like income are that the dispersion of individuals in state space, which in this case is the nonnegative real line, constantly increases other things being equal, and that there is a pervasive drift term tending to push everyone toward the origin. When there is a lower bound on the state space (such as zero income) the ergodic distribution of the drift-diffusion system is an exponential distribution (which is rather intuitive) and the exponent is the ratio of the rates of drift and diffusion. This is a nice way to explain the exponential outcome (though for economists perhaps constrained maxent, which uses all of the apparatus of constrained maximization that economics students learn much more about than they do about stochastic differential equations, is pedagogically more direct) in quasi-dynamic terms. But even in the case of the statistical mechanics of a perfect gas, there is no direct derivation of the drift-diffusion model that I know of from Newton's Laws.

Sanjay's Response to Duncan:

Thank you, most helpful. I infer that the specific drift-diffusion dynamics chosen are indeed very much open to the modeler's judgment.

Sanjay:

The possible further assumption of additive diffusion [call this case B*1] gives rise to the exponential distribution (previously derived by applying A*). The alternate further assumption of multiplicative diffusion [call this case B*2] gives rise to a Gibrat-Pareto power law distribution. It would appear that A* is compatible with B*1 but not with B*2. How should we interpret this? Is this because multiplicative dynamics (either alone or in conjunction with additive dynamics) necessarily entails changing, and not constant, total income of the society?

Duncan:

Since this is a case of multiple mathematical derivations of the same formal result, the main methodological point is my caution above against identifying the principle of maxent with any particular model that implies a maxent outcome. On the other hand, the difference between additive and multiplicative diffusion makes complete sense from a political economy perspective. It is the difference already present in Adam Smith's discussion between workers pursuing maximum wages (additive) and capitalists pursuing maximum profit rates (multiplicative), and it is striking that distributions of income are mostly (approximately) exponential while distributions of wealth are mostly (approximately) power laws.

Sanjay's Response to Duncan:

Thank you. You did not specifically controvert my statement that multiplicative dynamics (either alone or in conjunction with additive dynamics) necessarily entails changing, and not constant, total income of the society. For the moment, I take this to be the case. As such, it would be incompatible with the constraint that total income is conserved, meaning that if it could be derived from maximum entropy hypothesis at all (and it is not obvious that it can be) it would be in a framework without that particular constraint.

Sanjay:

In any case, the generality of the Folker-Planck equation approach would seem to allow for many more possible specifications of the dynamics beyond B*1 and B*2 both within the steady-state framework and outside of it. Have these various further possibilities been explored in the literature and what are they? More generally, what is the relationship between A* and B*? One of our colleagues put it to me that there is an "equivalence" but is that really the case? [What would it mean to say that, in any case? Here is a possible candidate formulation of his proposition: there is a one-to-one mapping between possible specification of the diffusion dynamics of the Fokker-Planck equation and specifications of the constraints under which entropy is to maximized].

Duncan:

It certainly would help if we knew a lot more about the dynamics of the income distribution, for example, from careful analysis of panel data to reconstruct Markov transition matrices (a conceptual approach suggested in the 1950s by Champernowne). In principle if you know the Markov transition matrix you can compute the ergodic distribution and compare it to the data. A couple of cautions are in order, though. Transition probabilities are basically time differentiations of state measurements, and it is notorious in social science data that differentiation greatly increases the noisiness of data. Even with pretty good measures of household income in a panel, the transitions probabilities are likely to have quite big error bars. The drift-diffusion story also highlights the sensitivity of the equilibrium distribution to small variations in the drift and diffusion parameters (since, for example, the exponential coefficient is the ratio of the two). Still, if we can get access to reasonably large-sample panels, why not get some students working on this?

Sanjay's Response to Duncan:

This would certainly be a project of great interest and importance, which can be conducted outside of the theoretical framework of statistical equilibrium, just as it can be conducted in alliance with or indeed within it. I have always thought that such a project would be very worthwhile. I attempted a version of it for my research on country stagnation some years ago, looking at transition matrices between per-capita income categories (as well as between economic-growth and economic-stagnation) and I certainly think it is a part of the 'royal road' vis-a-vis the study of personal income distributions (another part *being attention to the different patterns of change of the distinct functional components of personal incomes)*

Duncan:
I'm sure these remarks leave many questions unanswered, but they are very much worth further discussion.
Sanjay's Response to Duncan:
Thank you, Duncan. I very much appreciate your effort to provide clarifications and engage in this dialogue. I think that our students could also benefit from our conducting it more publicly. I think that there is enormous excitement about the framework but not very much (as yet) real understanding.

Email 4: Duncan to Sanjay (New Responses in Red)

Duncan:

Hi, Sanjay,

I think the discussion is of sufficient general interest to continue to include other colleagues, and also to keep a record in the form of the email interchange, though it would certainly be useful to discuss these questions in person or in some kind of open department forum, since I think there are several students who are quite interested in the topics.

Some more responses below. I will try to avoid simply re-iterating points I think are understood, even if not agreed-on.

Duncan

Sanjay:

Dear Duncan,

Thank you very much for this helpful response. In order to provide for as clear record-keeping as possible regarding the different points made, I will intersperse my further responses (below in italics) to which you should not at all feel obliged to respond, as we can always discuss in person.

I hope you do not mind that I am again cc-ing the colleagues originally addressed, as they may benefit from the dialogue and/or have additions and correctives to offer.

best wishes,

Sanjay

Duncan:

You raise a series of important questions, many of them not fully resolved as far as I can tell even among statistical physicists who use statistical equilibrium reasoning in a huge variety of contexts, and are very concerned about the foundational questions you raise.

One experience I had that helped me with these questions was discussions with Murray Gell-Mann in Santa Fe, where the issue of the explanatory status of statistical equilibrium is under constant (almost obsessive) discussion, usually in the form of such questions as "what is the entropy of the universe?" Gell-Mann says something like: "physicists do a very strange thing in studying complex systems with many degrees of freedom like the dynamics of a diffuse gas (where the state of the system at any instant is described by 6N numbers, where N is the number of particles the gas contains and the 6 numbers consist of 3 coordinates describing the position of the particle and 3 coordinates describing its momentum — mass times velocity). This appears to be a very difficult, perhaps practically impossible, problem, and what do physicists do? They think of what at first appears to be a much more difficult problem, which is to study not just the behavior of this particular gas, but of the collection (ensemble) of all possible systems that share certain properties with the one we are actually interested in. Remarkably, it turns out that it is much easier to draw strong and meaningful conclusions about the statistical properties of the ensemble than it is to draw conclusions about the dynamic behavior of the particular system, and, even more remarkably, those statistical conclusions agree with experimental observations in a huge range of contexts to a stunning precision." I think this is the main lesson to keep in mind methodologically.

Sanjay's Response to Duncan:

How general is this lesson? I think that this is the question we must answer if we wish to generalize the approach of statistical dynamics to other (e.g. social scientific) settings. For example, it seems a reasonable thought that the improbable simplification achieved by considering the entire ensemble of systems is a feature of particular kinds of dynamic processes (e.g. those involving some version of Brownian motion or other random processes in which local evolution in all directions is equally possible, and for which it is therefore reasonable to consider all possible states to be equally probable). If so, then the lesson may not be particularly general.

Duncan's Response to Sanjay's Response to Duncan:

The question of the generalizability of the ensemble point of view to other questions than thermodynamics is central to the discussion. Of course, until someone does it successfully for each particular field or problem, the discussion remains at an abstract methodological level. I myself think methodological questions only get settled by solving specific problems.

In physics, chemistry, biology, and other "hard science" fields, there are numerous demonstrations of the usefulness of the ensemble method in areas where the particular dynamic features you mention (Brownian motion, for example) are clearly not relevant. Physicists, especially in their incarnation as "econophysicists" may go somewhat overboard on this, but there are some philosophical reasons to think that the method is in principle quite general. I have written some papers about this, but I tried to sum up the point of view, which is based on the information theory problem of compression of information, in the last lecture of my Spring class, which is available at:

https://www.dropbox.com/s/ao5vnxbp2xs71yj/FoleyGECO6219Lec9.pdf?dl=0

I don't regard these results as answering the basic methodological question, but perhaps they can help to clarify the issues.

Duncan:

It leads to maximum entropy reasoning in an important range of cases, but not always. (For one thing there are other entropies, functions that share mathematical properties with Shannon entropy, and for some data maximizing one of these other entropies leads to better explanations of the data.)

To my mind the fundamental issue with statistical equilibrium is pretty well summed up by considering the question of the relation of statistical equilibrium to mechanics. There are two arguments here, one a Laplacian argument based on Newtonian mechanics, and a second one that considers quantum mechanics.

If we look at the universe as Laplace did, it appears to be an ensemble of atoms each with a position and momentum, all subject to Newton's laws of motion. Laplace famously said that if you could tell him the exact state of the universe, that is, the position and momentum of all these particles, at any instant, he could in principle compute the entire past and future of the universe by applying Newton's laws. One fundamental property of Newton's laws is that they are time-reversible. Thus in a certain sense from Laplace's point of view the entropy of the universe has to be invariant over time, or, to put this another way, there is no "arrow of time", which is the fundamental concept behind the Second Law of Thermodynamics (increase of entropy).

If we look at the universe from quantum theory, it appears (to Murray Gell-Mann, for example) as a single, highly entangled, quantum "state" with very many dimensions. But quantum theory, despite the limits it puts on the simultaneous measurement of position and momentum, predicts that state transitions over time are "unitary transformations" of this complex state, and unitary transformations have the same property of time-reversibility. Gell-mann, when he is confronted on this, says that in principle there is no change in the microscopic entropy of the universe over time, and that the second law of thermodynamics is a result of "coarse graining" because we can only measure states of the universe with a certain precision, and only a limited range of the entangled state variables.

The issue of the reversibility or irreversibility of processes is indeed a deep question in physics, with far from a clear resolution, as you point out. [On this see also the recent book by by Roberto Unger and Lee Smolin, The Singular Universe and the Reality of Time]. I do not see, personally, how the corresponding problem in the social world can be approached without introducing the role of consciousness, through which "time's arrow' is subjectively as well as objectively manifested. This is part of what I find unsatisfactory about overly scientistic approaches to social enquiry, that abstract altogether from subjective self-awareness or consider it epiphenomenal. Even insofar as we are dealing with narrowly economic data in a domain such as that of income distribution it will manifest itself, for instance in the form of a desire to protect wealth and to accumulate more of it, which provides a definite gradient to economic activities, so that the evolution of individual wealth is not simply the consequence of equidirectional "brownian motion". One can view this motive in terms of "rational" action or even "maximization" if one wishes, or not. It doesn't matter deeply. What matters is that intentional action and random action are different things.

Duncan's response to Sanjay's response to Duncan:

I completely agree with you that social science, because it deals with conscious human beings, raises issues that can safely be ignored in many physical and chemical systems. (Particles in a gas do not keep track of the energy balance sheets: only physicists do that in order to understand the ensemble properties. Capitalist firms and households do keep track of their wealth and income and use the concepts to plan purposeful behavior.)

It is therefore very important to understand (and this took quite a lot of discussion with Gregor and Ellis for me to grasp) that purposeful or intentional behavior in social contexts has the same mathematical representation as constraints in physical systems. I've written quite a bit on this in the lectures from the Spring course, too, but one example may make the point. When Gregor and Ellis tried to explain observed profit rate distributions (on which they collected very good data) by maximizing entropy subject to constraints on total capital and total profits (they call the concept "surplus value") the resulting model fits the data very badly. IN particular it predicts way too many very small firms with very high profits and too many large firms with very small profits. Clearly what is wrong here is that there is no constraint corresponding to the pursuit of higher profit rates by capitalists along the lines of Adam Smith's idea of capitalist competition which Marx later developed. When one introduces that in the mathematical form of a constraint, the maximum entropy model fit to the data improves dramatically, in fact, astonishingly. The point is that movement of capital is not a Brownian motion, but is constrained by the purposeful actions of capitalists in allocating allocating capital to maximize profit rates. (Note, too, the crucial substantive difference induced by recognizing that capitalists aim for the highest profit rate, which relates to the point we have already discussed about the difference between multiplicative and additive constraints.

Duncan:

This is, as far as I know, where the matter sits, at least out in Santa Fe.

But these conversations suggest to me that the questions you raise at a fundamental level are not resolved even in the rarefied theoretical discussions of high-level statistical physics.

Some responses to your specific questions below:

Sanjay:

I would like to ask you a couple of questions of clarification, for which purpose it maybe helpful for me to refer to the following article by Banerjee and Yakovenko, although only because it provides a useful reference point for the queries, rather than because it contains any specific insights not otherwise available (much of Yakovenko's work in this area appears overlapping, inter alia):

http://physics.umd.edu/~yakovenk/papers/2010-NJP-v12-n7-p075032.pdf

Duncan:

Yakovenko is unusual among econo-physicists in his interest in the political-economic aspects of these problems. His empirical work seems to me very carefully done. He does have certain blind spots, such as the willingness to import constraints that are meaningful in physical problems into economic contexts where they are not relevant. The leading example in my mind is the assumption of "detailed balance", which is a symmetry hypothesis that two interacting particles, say, in a gas, are equally likely to transfer energy in either direction, which Yakovenko imports into models of monetary exchange in the completely implausible form of assuming that there is an equal probability of economic agent A transferring money to or from B. I've pointed this out to Yakovenko many times over the years, but without making much of an impression.

Sanjay's Response to Duncan's Response:

This would appear to be precisely a specific example of the salience in modeling of the distinction between random and intentional action.

Sanjay:

First of all, let me note that in the article the assumption of entropy maximization is not particularly well-motivated.

Duncan's Response:

I'm not sure I agree with this. [Sanjay's Response: *I was in the first instance commenting on the Banerjee and Yakovenko paper I cited. Perhaps it is also not particularly well-motivated generally (and indeed, I suspect this, although I did not claim it here).*]

All theories that try to explain observed quantitative data explain the data as a combination of central tendencies (law-like regularities, or signal) and unexplained residual variation (errors, or noise). For example, regression analysis looks for a linear relation among the observed variables and regards the deviations of the data from the linear model as "errors", and then ranks different linear models depending on properties of the errors, such as unbiasedness, or approximate Gaussian distribution. We could think of this as making a guess as to the structural relations of the variables, and then trying to evaluate the quality of the resulting model by studying the properties of the errors.

Jaynes' program inverts this methodological procedure by beginning by maximizing the randomness of the residuals according to the principle of maximum entropy. It is important to understand the discovery of statistical physics that "randomness" means "maximum entropy", not necessarily some other intuitively appealing concept like "i.i.d. Gaussian residuals". (Sometimes maximizing entropy leads to i.i.d Gaussian residuals, but often it does not. Maximizing entropy over the entire positive and negative real line subject to a 0 mean and variance constraint does yield the Gaussian normal distribution. Maximizing entropy over a bounded interval of the real line does give the intuitive uniform distribution. But a small change in constraints, like maximizing entropy over the positive half line subject to a mean constraint does not yield i.i.d., but the highly skewed exponential distribution.)

Jaynes then locates the theoretical content of a model in constraints on the maximum entropy problem. This is very characteristic of the way physicists do empirical modeling, and totally foreign to the "shoot in the dark and check the residuals" methods that dominate econometrics. Jaynes sees this as an iterative process in which you try to explain the data without constraints (usually resulting in a very poorly fitting model, signaled by the fact that the entropy of the observed data is much smaller than the maximum unconstrained entropy on the relevant state space), and then hope you can figure out simple constraints that have theoretical content and will bring the maxent model into better agreement with the data. This discovery of constraints is not a question of turning a crank on some hypothesis machine, but is where the real insight of scientific discovery resides. For example, if you maximize entropy of a gas subject only to constraints on the number of particles and the volume of the container, you get a very poor theory that is at gross odds with observation. It took decades for physicists to realize that the key to understanding these systems was the conservation of energy (proportional to the square of momentum, and therefore an inherently nonnegative quantity). When you maximize entropy subject to the number of particles, the volume of the container and the total or mean energy, voilá, you get the

Maxwell-Boltzmann exponential distribution and you are in business with a theory that is in very good agreement with observation (though inevitably not with every observation, since real gases are not exactly perfect gases).

This does not, of course, mean that the particles in the gas are “trying” to maximize entropy. They are just knocking around colliding with each other and with the walls of the container. As far as each particle (if we could, as economists, speak that way) knows, the world is just Newton’s laws. No individual particle ever gets to an energy level where the constraint on the energy of the whole system begins to affect it directly. But because energy is conserved, the distribution of energy among the particles does follow the maximum entropy distribution very closely, giving rise to the macro-phenomena such as pressure and temperature that characterize the behavior of the gas.

This does not work all the time. If, for example, you have a perfect billiard ball moving on a frictionless surface between completely elastic walls, and you start that ball on a trajectory exactly perpendicular to the walls with a positive velocity, it will continue to bounce back and forth on exactly the same trajectory forever. An ensemble of balls moving in parallel will do the same thing, and you will not see any increase in entropy due to their bumping into each other or veering one way or another. But this fanciful example underlines just why it is so “difficult” to avoid the maximum entropy configurations of a system with many degrees of freedom. Jaynes would say that it you manage to keep the system at a lower entropy, you are somehow imposing a constraint on it, and that the proper theory for such a system needs to incorporate that constraint. (I think this is very relevant to the problem of income, wage, and wealth distributions.)

Sanjay’s Response to Duncan:

This is certainly a powerful methodological framework, insofar as one accepts the entry maximization starting point, but it may leave more questions than it answers: which constraints, when and why?

Sanjay:

Likening an income distribution to a probability distribution, why should we make the assumption that its 'information entropy' is maximized? One answer, and that provided by the authors on p.4, following the 'Wallis derivation' of the entropy maximization principle (see https://en.wikipedia.org/wiki/Principle_of_maximum_entropy) is as follows: The assignment of probabilities across possible income bins that maximizes entropy is that which can be brought about through a larger number of possible allocations of individual persons across bins than can any other. If every such allocation of persons is deemed equally probable, by analogy to the analysis of an ensemble of possible states of a system in statistical physics, or indeed in information theoretic approaches which assume that in the absence of further information one should make an equiprobability assumption across possible outcomes, then this maximum entropy assignment of probabilities is the one which is more likely than any other. [Call this argument A*].

Duncan:

Here we are going to run across the fact that different mathematical arguments can motivate the same result. The article you link to seems to me quite good on the whole, but be careful to realize that the derivation of maximum entropy from the multinomial distribution is just one way to do it, and may be relevant in some cases and not in others. (In the case of income distribution, for example, it is clear that arguments from ergodicity, the postulate that the system visits all possible states, are irrelevant, because we know that households do not systematically visit all possible income states. Similarly the buckets argument is not very easy to interpret in that context. The irrelevance of any particular mathematical model of maximum entropy to a particular empirical problem, however, doesn't tell us much about the relevance of the maxent conception to explaining the data.)

Sanjay's Response to Duncan:

I quite agree with you here, but this would seem precisely to illustrate the poverty of the motivation of the maximum entropy principle in the economic context.

Sanjay:

Fair enough, but is this a sufficient reason to assume this particular assignment to be that which is 'likely' in some sense to prevail? 'More likely' is after all not the same thing as 'likely', and there is no guarantee that the largest entropy assignmentI in any sense has a high absolute likelihood. In my undergraduate textbook (F. Reif, Fundamentals of Statistical and Thermal Physics, 1965) which I still have and which I partially reread this morning, it is said (Page 91) that, "Usually [the number of states accessible to the system] has a very pronounced maximum at some value [of one of the parameters]. In that case practically all systems in the final equilibrium situation will have values corresponding to the most probable situation, where [the parameter] is very close to [the level of the parameter that maximizes the number of states accessible to the system]". In the text, there's an accompanying indicative diagram showing the number of states accessible to the system as a function of the parameter, which has a Dirac-delta like shape. Is there any assurance in general economic systems that the corresponding diagram is not more diffuse?

Duncan:

This is an interesting point. In the simplest cases the probability of observing the system in a state described by a frequency distribution p[s] over the feasible states is proportional to exp[N H[p]], which, even when N is much smaller than Avogadro's Number, is indeed very sharply peaked and hard to distinguish from the Dirac Delta function. If we compute entropy with logarithms to base 2, for example, entropy is measured in bits, and each bit of entropy represents a halving (or doubling, depending which way you are going) of the probability.

Sanjay's Response to Duncan:

I will be grateful for a didactic reference. It would be interesting to examine further whether economic applications are likely to fall in the universe of 'the simplest cases'.

Sanjay:

Now, let me turn to the Yakovenko et al paper specifically. On p.4 he derives an exponential distribution (equation (6)) along the lines above, reflecting the conservation of total income in the way that the Boltzmann distribution in statistical physics reflects constancy of temperature. The maximization of entropy procedure undertaken there appears to be implicitly motivated by an argument along the lines of A*. Presumably, by analogy to physics, this is a case in which it is formally demonstrable that "practically all systems in the final equilibrium situation will have values corresponding to the most probable situation", in the sense that the vast majority of possible allocations of income across persons that respect the conservation of aggregate income constitute distributions that are "close to" the exponential distribution derived on the basis of entropy maximization.

Duncan:

But as I argued above, this can't be literally true for income distribution data. What the observed exponential shape of the bulk of the income distribution suggests is that in addition to whatever dynamics are moving households or individuals up and down in income, there is a weak statistical effect constraining them to a given total (or average) income. This has the rather extraordinary implication, which I think we are far from understanding well, even given the pioneering work of our students Markus Schneider, Amir Ragab, and Nicholas Papanicolaou, that when one household manages to raise its income, there is a very small (given the number of households) negative effect on the incomes of everyone else.

Sanjay's Response to Duncan:

Is this not more of a tautology than an implication, let alone an extraordinary one, insofar as mandated by the conservation of total income premise? This example does raise in my mind, as others do, to what extent the framework is providing an explanation as opposed to a description.

Duncan's Response to Sanjay's Response to Duncan:

It is striking that neither neoclassical nor Marxian wage theory imply a constraint on total wage income.

Sanjay:

Later on in the paper, however, the authors introduce what appear to be presented as a wholly independent "kinetic" approach (p.10), which involves different possible diffusion dynamics characterizing how distinct incomes, r, change by delta r over a time period delta t. They then apply the Fokker-Planck equation as an entirely general (and seemingly otherwise atheoretical) approach to characterizing the evolution of a probability distribution under an assumed diffusion process governing changes in the distribution over time, and they apply a steady-state ("stationary distribution") assumption to derive the compatible class of distributions. [Call this alternate argument as to what distribution to expect B*].

Duncan:

This type of model is a "drift-diffusion" model and is a favorite with physicists. The basic properties of these models for one-dimensional data like income are that the dispersion of individuals in state space, which in this case is the nonnegative real line, constantly increases other things being equal, and that there is a pervasive drift term tending to push everyone toward the origin. When there is a lower bound on the state space (such as zero income) the ergodic distribution of the drift-diffusion system is an exponential distribution (which is rather intuitive) and the exponent is the ratio of the rates of drift and diffusion. This is a nice way to explain the exponential outcome (though for economists perhaps constrained maxent, which uses all of the apparatus of constrained maximization that economics students learn much more about than they do about stochastic differential equations, is pedagogically more direct) in quasi-dynamic terms. But even in the case of the statistical mechanics of a perfect gas, there is no direct derivation of the drift-diffusion model that I know of from Newton's Laws.

Sanjay's Response to Duncan:

Thank you, most helpful. I infer that the specific drift-diffusion dynamics chosen are indeed very much open to the modeler's judgment.

Sanjay:

The possible further assumption of additive diffusion [call this case B*1] gives rise to the exponential distribution (previously derived by applying A*). The alternate further assumption of multiplicative diffusion [call this case B*2] gives rise to a Gibrat-Pareto power law distribution. It would appear that A* is compatible with B*1 but not with B*2. How should we interpret this? Is this because multiplicative dynamics (either alone or in conjunction with additive dynamics) necessarily entails changing, and not constant, total income of the society?

Duncan:

Since this is a case of multiple mathematical derivations of the same formal result, the main methodological point is my caution above against identifying the principle of maxent with any particular model that implies a maxent outcome. On the other hand, the difference between additive and multiplicative diffusion makes complete sense from a political economy perspective. It is the difference already present in Adam Smith's discussion between workers pursuing maximum wages (additive) and capitalists pursuing maximum profit rates (multiplicative), and it is striking that distributions of income are mostly (approximately) exponential while distributions of wealth are mostly (approximately) power laws.

Sanjay's Response to Duncan:

Thank you. You did not specifically controvert my statement that multiplicative dynamics (either alone or in conjunction with additive dynamics) necessarily entails changing, and not constant, total income of the society. For the moment, I take this to be the case. As such, it would be incompatible with the constraint that total income is conserved, meaning that if it could be derived from maximum entropy hypothesis at all (and it is not obvious that it can be) it would be in a framework without that particular constraint.

Sanjay:

In any case, the generality of the Folker-Planck equation approach would seem to allow for many more possible specifications of the dynamics beyond B*1 and B*2 both within the steady-state framework and outside of it. Have these various further possibilities been explored in the literature and what are they? More generally, what is the relationship between A* and B*? One of our colleagues put it to me that there is an "equivalence" but is that really the case? [What would it mean to say that, in any case? Here is a possible candidate formulation of his proposition: there is a one-to-one mapping between possible specification of the diffusion dynamics of the Fokker-Planck equation and specifications of the constraints under which entropy is to maximized].

Duncan:

It certainly would help if we knew a lot more about the dynamics of the income distribution, for example, from careful analysis of panel data to reconstruct Markov transition matrices (a conceptual approach suggested in the 1950s by Champernowne). In principle if you know the Markov transition matrix you can compute the ergodic distribution and compare it to the data. A couple of cautions are in order, though. Transition probabilities are basically time differentiations of state measurements, and it is notorious in social science data that differentiation greatly increases the noisiness of data. Even with pretty good measures of household income in a panel, the transitions probabilities are likely to have quite big error bars. The drift-diffusion story also highlights the sensitivity of the equilibrium distribution to small variations in the drift and diffusion parameters (since, for example, the exponential coefficient is the ratio of the two). Still, if we can get access to reasonably large-sample panels, why not get some students working on this?

Sanjay's Response to Duncan:

This would certainly be a project of great interest and importance, which can be conducted outside of the theoretical framework of statistical equilibrium, just as it can be conducted in alliance with or indeed within it. I have always thought that such a project would be very worthwhile. I attempted a version of it for my research on country stagnation some years ago, looking at transition matrices between per-capita income categories (as well as between economic-growth and economic-stagnation) and I certainly think it is a part of the 'royal road' vis-a-vis the study of personal income distributions (another part being attention to the different patterns of change of the distinct functional components of personal incomes)

Duncan:
I'm sure these remarks leave many questions unanswered, but they are very much worth further discussion.

Sanjay's Response to Duncan:
Thank you, Duncan. I very much appreciate your effort to provide clarifications and engage in this dialogue. I think that our students could also benefit from our conducting it more publicly. I think that there is enormous excitement about the framework but not very much (as yet) real understanding.

Email 5: Sanjay to Duncan (New Responses in Blue)

Sanjay:

Dear Duncan,

Thank you so much for your further responses. I am responding to your latest remarks below, this time using underlined text, to make my own latest additions clear. I do agree that we could fruitfully have a more public discussion of the issues, perhaps in the context of a department seminar or an informal chat during one of our Wednesday meeting slots, or indeed in a written dialogue in an appropriate setting, for the benefit of others in the community who have an interest in understanding the possible role of the methodology.

best wishes,

Sanjay

P.S. I am personally also open to making our correspondence, now or as it might develop further, public, e.g. via your website or mine, should that be thought a helpful contribution to a more public discussion of the possibilities and limits of the approach under discussion.

Duncan:

Hi, Sanjay,

I think the discussion is of sufficient general interest to continue to include other colleagues, and also to keep a record in the form of the email interchange, though it would certainly be useful to discuss these questions in person or in some kind of open department forum, since I think there are several students who are quite interested in the topics.

Some more responses below. I will try to avoid simply re-iterating points I think are understood, even if not agreed-on.

Duncan

Sanjay:

Dear Duncan,

Thank you very much for this helpful response. In order to provide for as clear record-keeping as possible regarding the different points made, I will intersperse my further responses (below in italics) to which you should not at all feel obliged to respond, as we can always discuss in person.

I hope you do not mind that I am again cc-ing the colleagues originally addressed, as they may benefit from the dialogue and/or have additions and correctives to offer.

best wishes,

Sanjay

Duncan:

You raise a series of important questions, many of them not fully resolved as far as I can tell even among statistical physicists who use statistical equilibrium reasoning in a huge variety of contexts, and are very concerned about the foundational questions you raise.

One experience I had that helped me with these questions was discussions with Murray Gell-Mann in Santa Fe, where the issue of the explanatory status of statistical equilibrium is under constant (almost obsessive) discussion, usually in the form of such questions as "what is the entropy of the universe?" Gell-Mann says something like: "physicists do a very strange thing in studying complex systems with many degrees of freedom like the dynamics of a diffuse gas (where the state of the system at any instant is described by 6N numbers, where N is the number of particles the gas contains and the 6 numbers consist of 3 coordinates describing the position of the particle and 3 coordinates describing its momentum — mass times velocity). This appears to be a very difficult, perhaps practically impossible, problem, and what do physicists do? They think of what at first appears to be a much more difficult problem, which is to study not just the behavior of this particular gas, but of the collection (ensemble) of all possible systems that share certain properties with the one we are actually interested in. Remarkably, it turns out that it is much easier to draw strong and meaningful conclusions about the statistical properties of the ensemble than it is to draw conclusions about the dynamic behavior of the particular system, and, even more remarkably, those statistical conclusions agree with experimental observations in a huge range of contexts to a stunning precision." I think this is the main lesson to keep in mind methodologically.

Sanjay's Response to Duncan:

How general is this lesson? I think that this is the question we must answer if we wish to generalize the approach of statistical dynamics to other (e.g. social scientific) settings. For example, it seems a reasonable thought that the improbable simplification achieved by considering the entire ensemble of *systems is a feature of particular kinds of dynamic processes (e.g. those involving some version of Brownian motion or other random processes in which local evolution in all directions is equally possible, and for which it is therefore reasonable to consider all possible states to be equally probable). If so, then the lesson may not be particularly general.*

Duncan's Response to Sanjay's Response to Duncan:

The question of the generalizability of the ensemble point of view to other questions than thermodynamics is central to the discussion. Of course, until someone does it successfully for each particular field or problem, the discussion remains at an abstract methodological level. I myself think methodological questions only get settled by solving specific problems.

In physics, chemistry, biology, and other "hard science" fields, there are numerous demonstrations of the usefulness of the ensemble method in areas where the particular dynamic features you mention (Brownian motion, for example) are clearly not relevant. Physicists, especially in their incarnation as "econophysicists" may go somewhat overboard on this, but there are some philosophical reasons to think that the method is in principle quite general. I have written some papers about this, but I tried to sum up the point of view, which is based on the information theory problem of compression of information, in the last lecture of my Spring class, which is available at:

https://www.dropbox.com/s/ao5vnxbp2xs71yj/FoleyGECO6219Lec9.pdf?dl=0

I don't regard these results as answering the basic methodological question, but perhaps they can help to clarify the issues.

Sanjay's Response to Duncan's Response to Sanjay's Response to Duncan:

I have tried to read through your slides carefully. Although a deep seated and unifying perspective is evident there, especially in particular in terms of the interpretation you put forward of the assignment of posterior probabilities over alternative "ensemble theories" in terms of the activity of a Universal Turing Machine, I must admit that I am left unsatisfied as far as the fundamental question of why to choose the posterior probability maximization over ensemble theories (or naively, entropy maximization) method over other methods or to unduly privilege its particular 'way of seeing'.

You are clear in stating that judgment plays a crucial role in the application of the method, as it is involved in basic steps such as the definition of the 'bins' for probability assignment, and the definition of the constraints that are deemed (by the analyst) to apply. You are also clear in stating that the application of the abstract point of view presented in the slides to individual cases may thus not be terribly useful. From this perspective, what you are doing appears to be to construct a sort of meta-theory, that you hope will make sense of more specific successful exercises in prediction or explanation. However, here I feel that there is a tension. You hold that 'methodological questions get solved by settling specific problems' but on your own account the conceptual frame of entropy maximization cannot provide, in and of itself, very much concrete guidance on many of these specific problems.

In any event, if the proof of the pudding must be in the eating then we will have to see specific demonstrations of the effectiveness of the method. My own predisposition is to be much more insistent that a proposed methodology must be justified in terms of its intrinsic conceptual merits, as well as that the tests of external correspondence which we place upon it should have to do not only with efficacy in prediction but with consistency with our (subjective and objective) understanding of the *processes* at work. Judgment is inextricably involved in assessing the performance of a theory according to both such intrinsic and extrinsic considerations, and that is no embarassment. Explanatory, and in particular predictive, efficacy can bolster the case for a method -- but it cannot make it.

Duncan:

It leads to maximum entropy reasoning in an important range of cases, but not always. (For one thing there are other entropies, functions that share mathematical properties with Shannon entropy, and for some data maximizing one of these other entropies leads to better explanations of the data.)

To my mind the fundamental issue with statistical equilibrium is pretty well summed up by considering the question of the relation of statistical equilibrium to mechanics. There are two arguments here, one a Laplacian argument based on Newtonian mechanics, and a second one that considers quantum mechanics.

If we look at the universe as Laplace did, it appears to be an ensemble of atoms each with a position and momentum, all subject to Newton's laws of motion. Laplace famously said that if you could tell him the exact state of the universe, that is, the position and momentum of all these particles, at any instant, he could in principle compute the entire past and future of the universe by applying Newton's laws. One fundamental property of Newton's laws is that they are time-reversible. Thus in a certain sense from Laplace's point of view the entropy of the universe has to be invariant over time, or, to put this another way, there is no "arrow of time", which is the fundamental concept behind the Second Law of Thermodynamics (increase of entropy).

If we look at the universe from quantum theory, it appears (to Murray Gell-Mann, for example) as a single, highly entangled, quantum "state" with very many dimensions. But quantum theory, despite the limits it puts on the simultaneous measurement of position and momentum, predicts that state transitions over time are "unitary transformations" of this complex state, and unitary transformations have the same property of time-reversibility. Gell-mann, when he is confronted on this, says that in principle there is no change in the microscopic entropy of the universe over time, and that the second law of thermodynamics is a result of "coarse graining" because we can only measure states of the universe with a certain precision, and only a limited range of the entangled state variables.

The issue of the reversibility or irreversibility of processes is indeed a deep question in physics, with far from a clear resolution, as you point out. [On this see also the recent book by by Roberto Unger and Lee Smolin, The Singular Universe and the Reality of Time]. I do not see, personally, how the corresponding problem in the social world can be approached without introducing the role of consciousness, through which "time's arrow' is subjectively as well as objectively manifested. This is part of what I find unsatisfactory about overly scientistic approaches to social enquiry, that abstract altogether from subjective self-awareness or consider it epiphenomenal. Even insofar as we are dealing with narrowly economic data in a domain such as that of income distribution it will manifest itself, for instance in the form of a desire to protect wealth and to accumulate more of it, which provides a definite gradient to economic activities, so that the evolution of individual wealth is not simply the consequence of equidirectional "brownian motion". One can view this motive in terms of "rational" action or even "maximization" if one wishes, or not. It doesn't matter deeply. What matters is that intentional action and random action are different things.

Duncan's response to Sanjay's response to Duncan:

I completely agree with you that social science, because it deals with conscious human beings, raises issues that can safely be ignored in many physical and chemical systems. (Particles in a gas do not keep track of the energy balance sheets: only physicists do that in order to understand the ensemble properties. Capitalist firms and households do keep track of their wealth and income and use the concepts to plan purposeful behavior.)

It is therefore very important to understand (and this took quite a lot of discussion with Gregor and Ellis for me to grasp) that purposeful or intentional behavior in social contexts has the same mathematical representation as constraints in physical systems. I've written quite a bit on this in the lectures from the Spring course, too, but one example may make the point. When Gregor and Ellis tried to explain observed profit rate distributions (on which they collected very good data) by maximizing entropy subject to constraints on total capital and total profits (they call the concept "surplus value") the resulting model fits the data very badly. IN particular it predicts way too many very small firms with very high profits and too many large firms with very small profits. Clearly what is wrong here is that there is no constraint corresponding to the pursuit of higher profit rates by capitalists along the lines of Adam Smith's idea of capitalist competition which Marx later developed. When one introduces that in the mathematical form of a constraint, the maximum entropy model fit to the data improves dramatically, in fact, astonishingly. The point is that movement of capital is not a Brownian motion, but is constrained by the purposeful actions of capitalists in allocating allocating capital to maximize profit rates. (Note, too, the crucial substantive difference induced by recognizing that capitalists aim for the highest profit rate, which relates to the point we have already discussed about the difference between multiplicative and additive constraints.

Sanjay's Response to Duncan's Response to Sanjay's Response to Duncan:

Thanks for this helpful reference. I do not see that as a a general matter "purposeful or intentional behavior in social contexts has the same mathematical representation as constraints in physical systems" . I do see that one might *attempt* to capture such behavior by using the language of constraints, though if that were possible it might be only be at the cost of a rather artificial and stretched understanding of what is a constraint (as in Bellmanian optimization when specifying the relationship between inter-temporal problem and sub-problems, where the constraint from one perspective is the optimizand from another).

I have just looked at the paper by Gregor and Ellis that is the one to which believe you refer (http://www.economicpolicyresearch.org/images/docs/research/political_economy/WP_2015-5_Statistical_Equilibrium_Profit_Rates.pdf). The authors are very straightforward in recognizing (citing Jaynes), as you are in your slides, the role of judgment (although neither they nor you use that word): "innovation in the distribution can be accounted for by additional constraints, that would have to be motivated by economic theory". I do not, however, see clearly the place in which they introduce the "constraint corresponding to the pursuit of higher profit rates by capitalists" to which you refer in citing them, unless you mean the reference at the top of p.18 and above equation (8) to the idea that "the problem in Eq. 6 is additionally constrained to have a constant average deviation (not absolute) from the mean," The motivation provided by the authors for that additional constraint is that, "Intuitively, depending on whether the deviation is positive or negative, a larger or smaller share than half the firms have above average profit rates." This looks to me like an *ad hoc* way of bringing the result closer to that which is desired rather than a principled or even transparent way of incorporating a behavioral postulate. Please correct me if I have misunderstood the place in the argument in which the idea of directional behavior as a constraint appears or how it plays a role. I have gone into this specific example precisely because I have my doubts about the idea that there is a general possibility, let alone a necessity (as you seem to suggest when you use the words "has the same mathematical representation") of being able to describe behavioral postulates *meaningfully* in terms of constraints (of the posterior probability assignment or entropy maximization problem).

I look forward to further discussion of these and related issues.

Duncan:

This is, as far as I know, where the matter sits, at least out in Santa Fe.

Sanjay: One of many places in our wide, wide world!

But these conversations suggest to me that the questions you raise at a fundamental level are not resolved even in the rarefied theoretical discussions of high-level statistical physics.

Some responses to your specific questions below:

Sanjay:

I would like to ask you a couple of questions of clarification, for which purpose it maybe helpful for me to refer to the following article by Banerjee and Yakovenko, although only because it provides a useful reference point for the queries, rather than because it contains any specific insights not otherwise available (much of Yakovenko's work in this area appears overlapping, inter alia):

http://physics.umd.edu/~yakovenk/papers/2010-NJP-v12-n7-p075032.pdf

Duncan:

Yakovenko is unusual among econo-physicists in his interest in the political-economic aspects of these problems. His empirical work seems to me very carefully done. He does have certain blind spots, such as the willingness to import constraints that are meaningful in physical problems into economic contexts where they are not relevant. The leading example in my mind is the assumption of "detailed balance", which is a symmetry hypothesis that two interacting particles, say, in a gas, are equally likely to transfer energy in either direction, which Yakovenko imports into models of monetary exchange in the completely implausible form of assuming that there is an equal probability of economic agent A transferring money to or from B. I've pointed this out to Yakovenko many times over the years, but without making much of an impression.

Sanjay's Response to Duncan's Response:

This would appear to be precisely a specific example of the salience in modeling of the distinction between random and intentional action.

Sanjay:

First of all, let me note that in the article the assumption of entropy maximization is not particularly well-motivated.

Duncan's Response:

I'm not sure I agree with this. [Sanjay's Response: *I was in the first instance commenting on the Banerjee and Yakovenko paper I cited. Perhaps it is also not particularly well-motivated generally (and indeed, I suspect this, although I did not claim it here).*]

All theories that try to explain observed quantitative data explain the data as a combination of central tendencies (law-like regularities, or signal) and unexplained residual variation (errors, or noise). For example, regression analysis looks for a linear relation among the observed variables and regards the deviations of the data from the linear model as "errors", and then ranks different linear models depending on properties of the errors, such as unbiasedness, or approximate Gaussian distribution. We could think of this as making a guess as to the structural relations of the variables, and then trying to evaluate the quality of the resulting model by studying the properties of the errors.

Jaynes' program inverts this methodological procedure by beginning by maximizing the randomness of the residuals according to the principle of maximum entropy. It is important to understand the discovery of statistical physics that "randomness" means "maximum entropy", not necessarily some other intuitively appealing concept like "i.i.d. Gaussian residuals". (Sometimes maximizing entropy leads to i.i.d Gaussian residuals, but often it does not. Maximizing entropy over the entire positive and negative real line subject to a 0 mean and variance constraint does yield the Gaussian normal distribution. Maximizing entropy over a bounded interval of the real line does give the intuitive uniform distribution. But a small change in constraints, like maximizing entropy over the positive half line subject to a mean constraint does not yield i.i.d., but the highly skewed exponential distribution.)

Jaynes then locates the theoretical content of a model in constraints on the maximum entropy problem. This is very characteristic of the way physicists do empirical modeling, and totally foreign to the "shoot in the dark and check the residuals" methods that dominate econometrics. Jaynes sees this as an iterative process in which you try to explain the data without constraints (usually resulting in a very poorly fitting model, signaled by the fact that the entropy of the observed data is much smaller than the maximum unconstrained entropy on the relevant state space), and then hope you can figure out simple constraints that have theoretical content and will bring the maxent model into better agreement with the data. This discovery of constraints is not a question of turning a crank on some hypothesis machine, but is where the real insight of scientific discovery resides. For example, if you maximize entropy of a gas subject only to constraints on the number of particles and the volume of the container, you get a very poor theory that is at gross odds with observation. It took decades for physicists to realize that the key to understanding these systems was the conservation of energy (proportional to the square of momentum, and therefore an inherently nonnegative quantity). When you maximize entropy subject to the number of particles, the volume of the container and the total or mean energy, voilá, you get the Maxwell-Boltzmann exponential distribution and you are in business with a theory that is in very good agreement with observation (though inevitably not with every observation, since real gases are not exactly perfect gases).

This does not, of course, mean that the particles in the gas are "trying" to maximize entropy. They are just knocking around colliding with each other and with the walls of the container. As far as each particle (if we could, as economists, speak that way) knows, the world is just Newton's laws. No individual particle ever gets to an energy level where the constraint on the energy of the whole system begins to affect it directly. But because energy is conserved, the distribution of energy among the particles does follow the maximum entropy distribution very closely, giving rise to the macro-phenomena such as pressure and temperature that characterize the behavior of the gas.

This does not work all the time. If, for example, you have a perfect billiard ball moving on a frictionless surface between completely elastic walls, and you start that ball on a trajectory exactly perpendicular to the walls with a positive velocity, it will continue to bounce back and forth on exactly the same trajectory forever. An ensemble of balls moving in parallel will do the same thing, and you will not see any increase in entropy due to their bumping into each other or veering one way or another. But this fanciful example underlines just why it is so “difficult” to avoid the maximum entropy configurations of a system with many degrees of freedom. Jaynes would say that it you manage to keep the system at a lower entropy, you are somehow imposing a constraint on it, and that the proper theory for such a system needs to incorporate that constraint. (I think this is very relevant to the problem of income, wage, and wealth distributions.)

Sanjay’s Response to Duncan:

This is certainly a powerful methodological framework, insofar as one accepts the entry maximization starting point, but it may leave more questions than it answers: which constraints, when and why?

Sanjay:

Likening an income distribution to a probability distribution, why should we make the assumption that its 'information entropy' is maximized? One answer, and that provided by the authors on p.4, following the 'Wallis derivation' of the entropy maximization principle (see https://en.wikipedia.org/wiki/Principle_of_maximum_entropy) is as follows: The assignment of probabilities across possible income bins that maximizes entropy is that which can be brought about through a larger number of possible allocations of individual persons across bins than can any other. If every such allocation of persons is deemed equally probable, by analogy to the analysis of an ensemble of possible states of a system in statistical physics, or indeed in information theoretic approaches which assume that in the absence of further information one should make an equiprobability assumption across possible outcomes, then this maximum entropy assignment of probabilities is the one which is more likely than any other. [Call this argument A*].

Duncan:

Here we are going to run across the fact that different mathematical arguments can motivate the same result. The article you link to seems to me quite good on the whole, but be careful to realize that the derivation of maximum entropy from the multinomial distribution is just one way to do it, and may be relevant in some cases and not in others. (In the case of income distribution, for example, it is clear that arguments from ergodicity, the postulate that the system visits all possible states, are irrelevant, because we know that households do not systematically visit all possible income states. Similarly the buckets argument is not very easy to interpret in that context. The irrelevance of any particular mathematical model of maximum entropy to a particular empirical problem, however, doesn’t tell us much about the relevance of the maxent conception to explaining the data.)

Sanjay’s Response to Duncan:

I quite agree with you here, but this would seem precisely to illustrate the poverty of the motivation of the maximum entropy principle in the economic context.

Sanjay:

Fair enough, but is this a sufficient reason to assume this particular assignment to be that which is 'likely' in some sense to prevail? 'More likely' is after all not the same thing as 'likely', and there is no guarantee that the largest entropy assignmentI in any sense has a high absolute likelihood. In my undergraduate textbook (F. Reif, Fundamentals of Statistical and Thermal Physics, 1965) which I still have and which I partially reread this morning, it is said (Page 91) that, "Usually [the number of states accessible to the system] has a very pronounced maximum at some value [of one of the parameters]. In that case practically all systems in the final equilibrium situation will have values corresponding to the most probable situation, where [the parameter] is very close to [the level of the parameter that maximizes the number of states accessible to the system]". In the text, there's an accompanying indicative diagram showing the number of states accessible to the system as a function of the parameter, which has a Dirac-delta like shape. Is there any assurance in general economic systems that the corresponding diagram is not more diffuse?

Duncan:

This is an interesting point. In the simplest cases the probability of observing the system in a state described by a frequency distribution p[s] over the feasible states is proportional to exp[N H[p]], which, even when N is much smaller than Avogadro's Number, is indeed very sharply peaked and hard to distinguish from the Dirac Delta function. If we compute entropy with logarithms to base 2, for example, entropy is measured in bits, and each bit of entropy represents a halving (or doubling, depending which way you are going) of the probability.

Sanjay's Response to Duncan:

I will be grateful for a didactic reference. It would be interesting to examine further whether economic applications are likely to fall in the universe of 'the simplest cases'.

Sanjay:

Now, let me turn to the Yakovenko et al paper specifically. On p.4 he derives an exponential distribution (equation (6)) along the lines above, reflecting the conservation of total income in the way that the Boltzmann distribution in statistical physics reflects constancy of temperature. The maximization of entropy procedure undertaken there appears to be implicitly motivated by an argument along the lines of A*. Presumably, by analogy to physics, this is a case in which it is formally demonstrable that "practically all systems in the final equilibrium situation will have values corresponding to the most probable situation", in the sense that the vast majority of possible allocations of income across persons that respect the conservation of aggregate income constitute distributions that are "close to" the exponential distribution derived on the basis of entropy maximization.

Duncan:

But as I argued above, this can't be literally true for income distribution data. What the observed exponential shape of the bulk of the income distribution suggests is that in addition to whatever dynamics are moving households or individuals up and down in income, there is a weak statistical effect constraining them to a given total (or average) income. This has the rather extraordinary implication, which I think we are far from understanding well, even given the pioneering work of our students Markus Schneider, Amir Ragab, and Nicholas Papanicolaou, that when one household manages to raise its income, there is a very small (given the number of households) negative effect on the incomes of everyone else.

Sanjay's Response to Duncan:

Is this not more of a tautology than an implication, let alone an extraordinary one, insofar as mandated by the conservation of total income premise? This example does raise in my mind, as others do, to what extent the framework is providing an explanation as opposed to a description.

Duncan's Response to Sanjay's Response to Duncan:

It is striking that neither neoclassical nor Marxian wage theory imply a constraint on total wage income.

Sanjay:

Later on in the paper, however, the authors introduce what appear to be presented as a wholly independent "kinetic" approach (p.10), which involves different possible diffusion dynamics characterizing how distinct incomes, r, change by delta r over a time period delta t. They then apply the Fokker-Planck equation as an entirely general (and seemingly otherwise atheoretical) approach to characterizing the evolution of a probability distribution under an assumed diffusion process governing changes in the distribution over time, and they apply a steady-state ("stationary distribution") assumption to derive the compatible class of distributions. [Call this alternate argument as to what distribution to expect B*].

Duncan:

This type of model is a "drift-diffusion" model and is a favorite with physicists. The basic properties of these models for one-dimensional data like income are that the dispersion of individuals in state space, which in this case is the nonnegative real line, constantly increases other things being equal, and that there is a pervasive drift term tending to push everyone toward the origin. When there is a lower bound on the state space (such as zero income) the ergodic distribution of the drift-diffusion system is an exponential distribution (which is rather intuitive) and the exponent is the ratio of the rates of drift and diffusion. This is a nice way to explain the exponential outcome (though for economists perhaps constrained maxent, which uses all of the apparatus of constrained maximization that economics students learn much more about than they do about stochastic differential equations, is pedagogically more direct) in quasi-dynamic terms. But even in the case of the statistical mechanics of a perfect gas, there is no direct derivation of the drift-diffusion model that I know of from Newton's Laws.

Sanjay's Response to Duncan:

Thank you, most helpful. I infer that the specific drift-diffusion dynamics chosen are indeed very much open to the modeler's judgment.

Sanjay:

The possible further assumption of additive diffusion [call this case B*1] gives rise to the exponential distribution (previously derived by applying A*). The alternate further assumption of multiplicative diffusion [call this case B*2] gives rise to a Gibrat-Pareto power law distribution. It would appear that A* is compatible with B*1 but not with B*2. How should we interpret this? Is this because multiplicative dynamics (either alone or in conjunction with additive dynamics) necessarily entails changing, and not constant, total income of the society?

Duncan:

Since this is a case of multiple mathematical derivations of the same formal result, the main methodological point is my caution above against identifying the principle of maxent with any particular model that implies a maxent outcome. On the other hand, the difference between additive and multiplicative diffusion makes complete sense from a political economy perspective. It is the difference already present in Adam Smith's discussion between workers pursuing maximum wages (additive) and capitalists pursuing maximum profit rates (multiplicative), and it is striking that distributions of income are mostly (approximately) exponential while distributions of wealth are mostly (approximately) power laws.

Sanjay's Response to Duncan:

Thank you. You did not specifically controvert my statement that multiplicative dynamics (either alone or in conjunction with additive dynamics) necessarily entails changing, and not constant, total income of the society. For the moment, I take this to be the case. As such, it would be incompatible with the constraint that total income is conserved, meaning that if it could be derived from maximum entropy hypothesis at all (and it is not obvious that it can be) it would be in a framework without that particular constraint.

Sanjay:

In any case, the generality of the Folker-Planck equation approach would seem to allow for many more possible specifications of the dynamics beyond B*1 and B*2 both within the steady-state framework and outside of it. Have these various further possibilities been explored in the literature and what are they? More generally, what is the relationship between A* and B*? One of our colleagues put it to me that there is an "equivalence" but is that really the case? [What would it mean to say that, in any case? Here is a possible candidate formulation of his proposition: there is a one-to-one mapping between possible specification of the diffusion dynamics of the Fokker-Planck equation and specifications of the constraints under which entropy is to maximized].

Duncan:

It certainly would help if we knew a lot more about the dynamics of the income distribution, for example, from careful analysis of panel data to reconstruct Markov transition matrices (a conceptual approach suggested in the 1950s by Champernowne). In principle if you know the Markov transition matrix you can compute the ergodic distribution and compare it to the data. A couple of cautions are in order, though. Transition probabilities are basically time differentiations of state measurements, and it is notorious in social science data that differentiation greatly increases the noisiness of data. Even with pretty good measures of household income in a panel, the transitions probabilities are likely to have quite big error bars. The drift-diffusion story also highlights the sensitivity of the equilibrium distribution to small variations in the drift and diffusion parameters (since, for example, the exponential coefficient is the ratio of the two). Still, if we can get access to reasonably large-sample panels, why not get some students working on this?

Sanjay's Response to Duncan:

This would certainly be a project of great interest and importance, which can be conducted outside of the theoretical framework of statistical equilibrium, just as it can be conducted in alliance with or indeed within it. I have always thought that such a project would be very worthwhile. I attempted a version of it for my research on country stagnation some years ago, looking at transition matrices between per-capita income categories (as well as between economic-growth and economic-stagnation) and I certainly think it is a part of the 'royal road' vis-a-vis the study of personal income distributions (another part being attention to the different patterns of change of the distinct functional components of personal incomes)

Duncan:

I'm sure these remarks leave many questions unanswered, but they are very much worth further discussion.

Sanjay's Response to Duncan:
Thank you, Duncan. I very much appreciate your effort to provide clarifications and engage in this dialogue. I think that our students could also benefit from our conducting it more publicly. I think that there is enormous excitement about the framework but not very much (as yet) real understanding.

Email 6: Duncan and Sanjay

Duncan to Sanjay:

Hi Sanjay,

I had to put our discussion on the back burner due to the exigencies of travel and other demands on my time, but I've now had time to look carefully at your further questions, and have a couple of things to say about them.

It is fine with me if we can find a good way to make this discussion more widely available. (It might be useful to go through what is now a complex and lengthy email thread and distinguish your and my contributions, say, by using colors, to make it easier for a reader to follow.)

At this point I think the most important issues you have raised are the "motivation" for maximizing entropy and the sense in which purposeful behavior (say, Smith's notion that capitalist seek the highest profit rate in allocating their capital) can be incorporated into statistical equilibrium theory as a constraint. The discussion of this latter point in Ellis and Gregor's paper is, as you discovered, tentative and incomplete. There are sharper formulations of this point in my lecture slides from my Spring course, and Ellis and I are working on a paper discussing this in the context of social science problems like the distribution of profit rates where aggregate variables like realized profit rates interact with individual decisions like entry and exit from specific market niches interact to produce the regularities we observe in the data.

These points are connected, and in my mind related to the "Occam's razor" approach I tried to develop formally in the last lecture of my Spring course. This approach is far from uncontroversial. The idea, however, is formally straightforward, which is to prefer the theory that allows for the greatest informational compression of the observed data. This aspect of theory selection is a constant theme in the history of science under the guise of "simplicity", "elegance", "parsimony" and other similar categories of theory evaluation.

One famous example that illustrates this type of argument is the Kepler-Newton analysis of planetary orbits as ellipses. In this case the raw data that needs to be compressed is the series of planetary observations of Tycho Brahe (and others stretching back to the Babylonian astronomers). The Kepler system compresses this data by describing it in terms of elliptical orbits which require only two parameters to describe, and deviations of the actual observations from the theoretically predicted positions of the planets. Kepler's ellipses do indeed compress the data dramatically.

In my attempt to reconstruct this line of meta-analysis (as you correctly diagnose it) within a Bayesian framework, which I summarized in the last lecture of my Spring course, I follow the "Kolmogorov" tradition of measuring the complexity of a theory in terms of the length of the Turing machine program that computes the theory. (This idea, which was first articulated by Ray Solomonov, is often called the "universal prior". Before we have any data at all, we prefer theories that can be encoded as short and therefore simple programs. There is an open question as to whether Turing machines are the only or even the most relevant, model of complexity in this sense, since there are other formal computation models, such as analog computation, Turing machines supplemented by "oracles", and neural networks, which may be more relevant as the basis of a "universal prior".)

Note that this "universal prior" is not a maximum entropy concept. The entropy seems to come into the problem at another level. The computer program sorts possible data strings (or data sets) into classes on the basis of some relevant properties (which are the content of the theory and cannot be deduced from the prior directly), and assigns a frequency to each class. (This elaboration of Kolmogorov's theory is Murray Gell-Mann and Seth Lloyd's theory of "effective complexity".) It is in this stage that the entropy issues arise as well as constraints. The entropy, as far as I can see, comes into the story because it provides a system of ranking and encoding frequency distributions, which is part of the length of the Turing machine program that instantiates the theory. When we prefer maximum entropy distributions it is, in this formal framework, because they will be described by short code words and therefore reduce the length of the computer program.

The simplest concrete representation of this whole setup is Gibbs "canonical ensemble" representing a particle system in contact with a heat bath. The exact configuration of the system in the state space of positions and momenta of its particles determines a particular energy, and is the basis of the sorting the Turing machine does (essentially computing the energy of each configuration and classifying it into one of a finite number of bins — "coarse graining".) The probabilities assigned to the bins have to meet the constraint that the mean energy of the system is given (which is equivalent to assuming that it is in equilibrium with the heat bath of given temperature). The statistical equilibrium Gibbs derived is the maximum entropy assignment of probabilities subject to this constraint.

There are other physical (and presumably social) systems where the maximum compression of the data is achieved by minimizing Shannon entropy, or maximizing some other entropy-like function.

This story about theory choice is problematic because it appears to depend on purely formal considerations, and many people as a result do not see why the theory that achieves maximum compression of the known data ought to be regarded as "true". My Swarthmore instructor in the seminar on social philosophy, Larry Sklar, who taught for many years at Michigan, and I had vigorous discussions on this point when I was an undergraduate (and only very dimly saw the whole implications of the information approach). We can discuss this philosophical point at greater length, since it touches on some central issues in political economy such as the F(riedman)-twist argument that theories stand or fall not by the "realism" (truth value?) of their assumptions (such as market clearing perfect competition) but by the accuracy of their predictions.

But I think the more immediate issue you raise is the question of introducing purposeful behavior into this general ensemble framework, which is clearly of crucial importance if we are going to get anything out of this approach in social science. My thinking on this question has also been shaped by Anwar's questioning of the need for any payoff- or utility-bases principles of explanation in economics. At a general level, I addressed this question in my Spring lectures in terms of the "entropy-constrained behavior model", from which it is straightforward to derive the quantal response empirical model that pops up in so many applied areas in economics and other social sciences. This model is just the von Neumann-Morgentstern maximum expected payoff (utility) model but with the addition of a constraint requiring the mixed strategy of the actor to have a minimum entropy. In a shorthand version the actor is supposed to choose a frequency distribution f[a] over available actions so as to maximize E_f[U[a]] subject to the constraint S[f] ≥ Sbar, where E_f is the expectation with respect to the mixed strategy, U[a] are the payoffs associated with each action, S[f] is the informational entropy of the mixed strategy, and Sbar is the minimum entropy relevant to the decision problem. The Lagrangian for this maximization is E_f[U[a]] - T S[f]. Given the linearity of the objective function in the frequencies and the strict concavity of the entropy, a saddle point of this Lagrangian is necessary and sufficient to solve the original constrained optimization program, (By analogy with the concept of "free energy" in thermodynamics, some people call the Lagrangian the "free utility" of the system.) This model has a very good track record in explaining observed behavior in a wide range of experimental and observational contexts in social science.

But suppose we set the problem up as maximizing entropy subject to a constraint on expected payoff (utility). Then the Lagrangian would be S[f] - \beta E_f[U[a]], which has exactly the same saddle points as the free utility define above, taking \beta = 1/T. It is in this formal sense that purposeful pursuit of goals such as wealth acquisition, profit rate maximization, and so forth can be transparently represented in the maximum entropy framework as constraints on maximization of entropy. (In general we do not have very good ways of predicting either the behavior temperature T, or, equivalently, the satisficing level of expected payoff, but no one has been paying much attention to these dimensions of social and human phenomena, either, and there are some ways of using controls and comparisons to get a handle on this question.)

All the best,
Duncan

Dear Duncan,

Thanks so much for your extended and thoughtful reply. I am sorry that I in turn have taken so long to respond to you. I am grateful to you for engaging in this dialogue, which I have as previously found very illuminating. In the interest of further insight, let me take up now some of the issues that you have raised or to which you have responded, though not in the same order. I also do apologize if I have not understood correctly your points. I am taking the liberty of now cc-ing a few current or former students with whom I have shared versions of the correspondence.

First, let us address the question of the representation of an objective (of agents) as a constraint (of the system). You point out that two different problem statements can be viewed as dual, viz. (I) the choice of a probability distribution that maximizes the expected utility of the outcomes realized subject to the constraint that the "informational entropy of the mixed strategy" must exceed the "minimum entropy relevant to the decision problem", and (II) the maximization of entropy subject to a constraint on expected payoff (presumably that the payoff must be at least the level attained when maximization is achieved under (I). Now, I would make two remarks in response to this admittedly very interesting point. First, the presence of such a duality is not specific to the entropy formulation but rather is wholly generic. Max $f(x)$ st. $g(x) <= 0$ requires us to find the critical points of the Lagrangian $f(x)-k1g(x)$, where k1 is the relevant constraint multiplier. Similarly, the problem Min $g(x)$ s.t. $f(x)>=a$ requires us to find the critical points of the Lagrangian $g(x)-k2((f(x)-a)$ where k2 is the relevant constraint multiplier. It is readily observed that the FOC of the first problem is that $f'(x)=k1g'(x)=0$ and that of the second problem is $g'(x)=k2f'(x)=0$. These are the same statement (as k2 can just as well be rewritten as 1/k1 and thus the values of x satisfying the FOC are identical. This dual structure is part of the bread and butter of standard microeconomics, as you know well. [Here I have of course abstracted from niceties involving the establishment of the existence of interior solutions to the maximization problems and such like]. From this point of view I might be tempted to say, as you have said in relation to entropy maximization problems, that the 'natural' representation of an objective is in terms of a constraint, but it seems to me that this would strictly speaking be an error, since to say that the maximization of utility subject to an income constraint has a dual problem which is that of the minimization of expenditure subject to a utility constraint is to say that two problems have a relationship, and in particular that they possess a common solution set, but not that they are the same problem. In any case, it would seem that the existence of such a dual relationship is an intrinsic semantic feature of the concept of maximization subject to constraints, and to that degree not especially insight generating. However, we might reasonably ask whether some 'traction' can be generated by using the dual formulation instead. I am afraid that this is where I am really unconvinced, since the premise of the argument you present is that a good basis for social scientific problem solving is to describe situations in terms maximization of a (probabilistic) objective function subject to minimum entropy. You write that "This model has a very good track record in explaining observed behavior in a wide range of experimental and observational contexts in social science". I must admit that I do not know what you are referring to here, both because I do not know what the 'subject to minimum entropy' constraint would mean concretely and what is the basis of the claim that it has a 'very good track record'. Putting it even more pointedly, to the extent that the 'subject to minimum entropy' formulation is the same as recognizing conventionally recognized microeconomic constraints, there is nothing added by the entropy approach, and to the extent that it is different there is a large amount of work to do to convince sceptics either that this is a sensible (let alone, natural) formulation of the problem, and still more work to do to demonstrate that it has a good 'track record'. I remain concerned that the application of maximum entropy reasoning to cases such as dice throwing does not analogize to the social world for a variety of reasons, turning on the fact of human agency [I have heard it said for instance that human beings in a crowded elevator normally position themselves in such a way as to maximize mutual distance, i.e. to maximize entropy. However, it would seem plausible that this is most likely true when they are strangers. Does it change when the persons see each other in the elevator everyday and become friends, or when a calamitous event influences their propensity to seek solace in one another? For now, this is a speculative remark. This argument will have to be developed at greater length subsequently]. Perhaps I have misunderstood your argument, in which case it may help to discuss specific examples.

Second, let us address the idea that entropy maximization may provide the key to identifying theories that satisfy Occam's razor. Following the Kolmogorov-Solomonov perspective as you describe it, the complexity of a theory might be measured in terms of the length of the Turing machine program that computes the theory. In that case, Occam's razor would constitute a requirement to identify a program length that is as short as possible but still suffices to 'explain' the observed data. However, this deserves some deliberation since what constitutes an explanation is very much what is at issue. Even if we take a view (the F-twist as you note is of this kind) that explanation has no independent content or interest in theory selection beyond prediction (or more generally ability to produce or reproduce the observed data) there is a question of what constitutes a good predictive record, and this determination can be vexed in the social sciences, unlike in the physical sciences (in part because of the way in which observations and the contexts of observations and indeed the observers themselves co-evolve, but also because of the role of human agency in generating future outcomes).

Let us, however, suppose that it had been conceded that a particular theory is predictively successful. It would not necessarily follow that the theory offered a good *explanation*, since whether it did so might depend on, for instance, the extent to which we could identify principles that both accounted for the observations, and that carried over to other cases. The Kepler-Newton analysis of planetary orbits, which you cite, was successful in these terms, because it could explain Brahe's observations not only in terms of a highly successful predictive theory that was enormously 'compressed' but which also appeared to have broader 'warrant' in terms of the applicability of Newton's laws to problems beyond the domain of planetary observations, and in terms of the 'sense and sensibility' that they expressed. This is not to say the predictive efficacy was not essential, which it of course was. This is a crucial point for the human sciences, in which predictive efficacy is less often realized, but it seems wholly reasonably to require that in order to be deemed adequate a theory must bear a consonant relationship with our understanding of ourselves and others, in addition to possessing a suitable correspondence with 'objective' external observations. As Newtonian physics came to be understood to be a special case of a more general theory it became necessary to qualify the domain of its successful application, not because it is any less predictive in its original domain than it was but because the explanation that it provides is now understood to have a more limited domain of application. Newtonian mechanics remains the favoured theory for very large areas of application because of its predictive power, even if its 'truth' as a free-standing body of ideas is now understood to be qualified by its embedding in a still more general theory. In any case, as you note, to adopt the 'universal prior' formulation does not necessarily have to take us toward the maximum entropy approach to identifying the form such a prior should take. [Here, it would seem that, using the language of our earlier correspondence, maximum entropy is meta-theory and universal prior is meta-meta theory. This reminds me of a famous one-liner of G A Cohen: 'For anything you can do, I can do meta'!]

Maximum entropy maximization is one methodological choice that could undergird substantive theory development (at the meta-theory level) and the case for or against it would rest not only on its predictive power -- in my view, yet to be adequately established for the human sciences -- but on the explanatory sense of its postulates, once shorn of technical armature and asked to provide contextually convincing justifications. Entropy maximization, like everything else, will sail the seas or run aground on the shoals of the concrete.

best wishes, as always,

Sanjay

Dear Sanjay,

Some responses to your questions.

The duality properties of mathematical programming problems are, I agree, well-known, and in fact were first developed as far as I know in the context of thermodynamic theories in physics. When economists, particularly Paul Samuelson in his Foundations, imported these ideas into economics, however, they curiously failed to import one part of thermodynamic theory, namely its success in integrating theories of central tendencies and fluctuations. I commented on this in my INET talk in Cambridge in 2011, which is downloadable from:

https://www.dropbox.com/s/9va0lsapxzus3k6/FoleyINETCambridgeRev.pdf?dl=0

Nonetheless, it took Ellis, Gregor, and myself quite a bit of thinking to understand the implications of this duality for the equivalence of introducing goal-oriented behavior such as Smith's theory of competition guided by profit rate differentials and the imposition of constraints in statistical equilibrium models. As far as I can tell, this point is still not widely understood, even among the handful of people working on statistical equilibrium in economics and other social sciences.

The "traction" available from systematically employing this type of constraint depends, as you would expect, on the details of the particular problem under consideration, but appears to be very helpful in a wide range of problems including profit rate distributions, induced technical change, stock pricing, housing price distributions, and labor markets with migration, in all of which our students have discovered the characteristic signs of statistical equilibrium in the relevant distributions.

The "good track record" I referred to is the widespread empirical use of the logit quantal response and related models to study data involving individual or household or firm differences in categorical outcomes, such as home ownership vs renting, as well as a huge range of experimental data, starting with Luce and Suppes' work on expected utility theory and going up to the quantal response solutions to non-cooperative games. The connection to statistical equilibrium is that maximization of expected utility (or payoff) subject to an entropy constraint results in the logit quantal response function, though admittedly there are other derivations as well.

Every data-based empirical analysis has the goal of separating data fluctuations into central law-like tendencies and "noise". The main contribution statistical physics makes to this broad effort is the discovery that the correct definition of noise is in terms of informational entropy. Many econometric and related methods depend on one or another intuitively plausible notion of randomness, such as i.i.d. disturbances, or Gaussian normal distributions based on central limit theorem-type considerations. In specific contexts these models do maximize informational entropy, but not in others, particularly contexts where there are limits on the domain of the data such as non-negativity. Maximizing entropy under constraints from this point of view is just being cautious in the sense of not inadvertently introducing assumptions into the theory that are not explicitly supported by evidence, which is E. T. Jaynes' fundamental message.

The question of whether Occam's razor and data compression carries the day in terms of theory choice is certainly controversial. My impression, though, is that the data compression view is a fairly widely accepted consensus among practicing scientists. The controversy seems to be located more in the discussion of the philosophy of science, which as we know, tends to lag behind actual scientific practice. In my view economic research suffers from a very widespread epidemic of over-fitting, which increasing use of rules of thumb like the AIC and BIC criteria have done something to mitigate, but still seem to crop up in poor out-of-sample performance of models. I personally think it makes more sense for working scholars and scientists to see themselves as refining and improving theoretical understanding, without assuming that there is a "true model" somewhere at the end of the line.

Duncan

Email 7: Sanjay and Paulo

Dear Paulo,

Do you also have a methodological paper on these issues regarding the salience of econophysics and statistical mechanical methods in the social sciences? If so I would be grateful if you might share with me the latest version. (I believe that you had signalled earlier that you have written such a paper.)

I intend to put up our correspondence on these issues on a webpage, as discussed earlier with Duncan, if there is still no objection by him, so that it can be a contribution to a more public engagement with these issues. I may try to edit the text to avoid exact repetition of text resulting from the multiple rounds of responses in the email correspondence, but otherwise do not expect to make any substantive amendments.

If you have also written such a paper, I will group my response to the paper by Scharfenaker and Foley - sent to me in December, but to which I have not yet responded - and my response to your paper together. (Duncan, if there is a more recent version of your paper with Ellis, please also do send that if you can find the time).

best regards,

Sanjay

Hi, Sanjay,

Great to hear from you on this again. I'm still working on a more thorough note discussing several aspects of the application of methods form Statistical Mechanics to economic analysis. It's the first thing on my list for after the semester ends.

In the meantime, I am attaching the most recent version of my paper on social scaling. It does offer a brutally brief version of most of the central methodological points I've been mulling over. It's also a very brief paper. It's doing the rounds now, and will hopefully appear out there soon.

I'll keep you posted about all other relevant materials. We should chat in person about all of this soon too.

Welcome back,

P

Dear Duncan, Paulo and colleagues,

Please find below a further contribution to our discussion. Please forgive me if I have misunderstood anything, which I may well have, or if my critique appears too frontal. I have thought it best to dispense of politesse in the interest of clarification of these important issues. As mentioned earlier, assuming that you still agree, I propose now to circulate this correspondence publicly in order to generate a broader discussion.

I do hope that all is well with you.

With best wishes, as always,

Sanjay

Does the approach of statistical mechanics / information entropy / econophysics provide a unifying approach that can help satisfactorily to address a very broad range of problems central to social science, by applying a common scientific lens? That may not be the claim being made by its advocates, in which case there is no need for me to address it. However, I understood Duncan Foley to be making precisely a broad claim of this sort on behalf of the approach. If that *is* the claim being made, I would offer a response in the form of two propositions and a conclusion that derives from them (having now attempted to read the methodologically reflective papers offered by Scharfenaker and Foley (https://www.dropbox.com/s/qsbcy3r95gmeov1/QuantalResponseEquilibrium20161208.pdf?dl=0) and by dos Santos (attached)):

1. Causality, Description and Explanation: To make an inference of (or a judgment about) the probability of one kind of occurrence conditional upon another is not in and of itself to establish the causal relationship between the two occurrences. Foley and Scharfeneker write that, "we regard conditional distributions as rep- resenting causal relationships and thereby conveying the explanatory content of the model". Whether or not this is, as they state, "general statistical practice", it is a particular understanding of causal relations. To take an example, on their own account, the causal impact understood in this sense of firm entry and exit decisions on the profit rate or vice versa (to use their example) is determined by observed relationships. However, the impact of *potential* entry on profits could not by definition be captured by observations of, or inferences about, the relationship between actual entry and profits.

One might respond that a different kind of observation, that captures the extent or ease of potential entry, is required. That is indeed a way of restoring the interpretation of causes as conditional probabilities, but to concede this is to recognize the need to be attuned to unobserved, and perhaps unobservable, factors, as well as observed ones, in identifying the central descriptive features of a model (and not merely the variation around those features).

This is only an example. The deep-seated and general question that arises here concerns the requirements of understanding. If I am able to identify all of the conditional probabilities of a riot taken place given relevant prior observed or unobserved facts (such as those concerning the relations between members of different groups) does this offer an *explanation*? On the statistical view advanced by the authors, a sufficiently fine-grained account of such probabilities is itself an adequate explanation.

On another view, a causal account of a riot is a narrative concerning the structured relationship between distinct phenomena over time and space, and across persons. It requires a substantive understanding of the actions of individuals, the reactions of other individuals, the responses or non-responses of the police or others, and so on, each of which may contain internal (subjective) as well as external (objective) elements. Understanding these relevant facts about the riot in light of broader contexts (social, political, economic, cultural) is moreover necessary to anchor the effort at causal explanation. If one were to assert that this task is one of mere identification of conditional probabilities, even if this were not formally speaking wrong it would involve a kind of entropic imperialism, fitting a variable statistical crown to whatever heads can be found.

Sharfenaker and Foley themselves concede that, "In general different constraints correspond to different substantive assumptions about the behavior of individuals and their interaction through social institutions. These substantive questions cannot be avoided purely through the invocation of the maxent formalism." Precisely for this reason, the question of what value is added by the maxent formalism looms large.

The information entropy approach might provide a formal description of certain problems -- in particular those in which law-like behavior of actors (up to some informational constraint) can be presumed, as embodied in the specific Lagrangian maximization program thought to be descriptively relevant (such as maximizing entropy subject to utility being above a certain threshold or maximizing utility subject to entropy being above a certain threshold). But can such description be said to rise to the level of explanation? In my view this is unlikely since supplemental assumptions appear necessary to develop a theory that goes beyond the mere 'what' of description (e..g of a pattern of profit seeking actors entering or exiting industries in response to profit rates while making errors in the process) to encompass the 'how' and the 'why' of explanation (as would for instance an account of the mechanism by which profit rates adjust to reflect such entry or exit).

2. Descriptive Choice (or Description vs. Description): To show that a certain description of a relationship between two kinds of occurrences is possible, or even parsimonious (i.e. that it captures a lot with a little) does not demonstrate that it is a superior description. In providing a description of a social phenomenon involving entropy maximization, the choice of presumed constraints (equivalent to the choice of narratival frame) remains crucial, thus raising a question about whether the supposed generality of the method in fact lies in its being an empty vessel, or indeed whether it is science or art. Emptiness of a vessel can indeed be a 'virtue', but not one on account of which it is superior to a vessel containing fine wine (or for that matter, orange juice) or indeed to some other empty vessel that is available alongside. Since the choice of specific assumptions regarding the constraints (e.g. regarding the relevant properties of the distribution, such as sums, means, or summed deviations) that are to be maintained while maximizing entropy is crucial to the resulting predicted distribution, the exercise as a whole might be more convincingly described as one of retrospective curve fitting than of prediction. (The impression that this is so is not reduced by a detailed reading of the technicalities involved in applying the approach, which involve confronting statements by Sharfenaker and Foley such as that "δ can be estimated by minimizing the Kullback-Leibler divergence between the maximum entropy marginal distribution f[x] and the observed frequency distribution f[x]" which appears in effect to mean that it is precisely retrospective curve fitting that is at the heart of the applied exercise). The 'unreasonable effectiveness' of the method is hardly a surprise from this point of view ("we calculate the informational performance of the fitted distribution, f[x], as capturing approximately 98 percent of the informational content of the observed frequency distribution f[x]" (ibid)).

The idea of a social interaction that preserves a central moment, as advanced by Dos Santos, is consistent with but more general than that of conservation of a total quantity (since the latter entails the former but not vice versa), and can provide a description of the constraint that can give structure to an entropy maximization approach. However, in my view this is not a deep discovery about the world. Rather it is a discovery about the possibility of offering a certain (non-unique) description of the world.

[Incidentally, although the joint distribution being sought is defined by Foley and Scharfeneker as a statistical equilibrium, it is not clear to me why it deserves the moniker of an equilibrium, if by the use of that word we mean to invoke the mutually reinforcing character of a relationship over time. It is true that given the joint distribution we can derive the conditional distributions. However, the sense in which an equilibrium exists between these is obscure to me and not illuminated by the prior fact. This appears to be an instance of a form of descriptive narrowness that is not entailed by the formalism itself, but I may well have failed to appreciate something)].

Conclusion:

The ability of the information entropy approach to facilitate progress in answering the central questions of social science appears to me, at this point in my understanding of it, to be limited. The approach seems at most to provide a contribution to the armamentarium of social science. Among the problems faced by the information entropy approach are the plurality of possible descriptions (internal to the approach) making of it an empty vessel; the difficult of identifying laws (identified constraints) that are invariant over the domain of observation - in the face of ever disruptive social facts such as that of human agency – making it depend on an unwarranted scientistic conceit; and the need to characterize the dynamics that bring about outcomes (mechanisms) in terms of substantive motivations, intentions and processes of interaction in order to arrive at satisfactory causal explanations – a necessity the method might accommodate but not inform. It remains to be seen whether, when allied with other techniques of analysis, it could begin to address these difficulties in a manner that justifies its claim to be more than a dignifying cloak.

The approach of statistical mechanics / information entropy / econophysics may provide a useful tool, in particular for retrospective summary descriptions fitted to data at a high level of abstraction, but does not appear, at least to me, here and now, to offer a way out of our basic conundrum: that of how to choose descriptions and develop explanations that have the *substantive* content that would enable them to appear credible and persuasive in characterizing the causal dynamics that give rise to observations, both when applied to individual cases and to diverse cases considered together.

Email 8: Duncan's reply to Sanjay

Dear Sanjay,

This dialog has been very helpful in clarifying my thinking on the methodological issues around statistical equilibrium methods in social science. At some point we will need to separate the discussions into threads to make the different issues easier for other interested readers to follow.

One set of your questions, such as whether or in what way statistical equilibrium methods provide "explanations", whether the method is an "empty vessel", the significance of maximizing Shannon entropy under constraints, and what the relation of statistical equilibrium models to dynamical models is, are applicable to all statistical equilibrium theories, including the widely used methods of statistical mechanics in statistical physics.

A second class of your questions raise somewhat narrower issues, in particular, whether and how methods that work well in physical science can legitimately be imported into social science and used to model social data.

A final set of questions are specific to the statistical equilibrium model Ellis Scharfenaker and I proposed to explain profit rate convergence.

In organizing this dialog for other readers, or in setting up a possible conference to discuss these issues, it would probably be a good idea to separate these threads. I'm responding to your points below as they come up in your comments. I think I have answers to some of your questions, while others raise issues that are controversial among statistical physicists and philosophers of science on which I have my own ideas, but feel I have a lot learn about.

As a prologue, let me retell an interaction I had with Murray Gell-mann in discussing these methods. Gell-mann, in an effort to instruct my limited understanding of these issues said something like: "Statistical mechanics is a very strange enterprise. It addresses the problem of describing and analyzing a system, for example a gas, with an enormous number of particles (Avogradro's Number is on the order of 10^23) and a corresponding enormous possible number of configurations. This is an impossibly difficult problem for dynamics, since to solve it would require measuring an astronomical number of initial conditions and solving an astronomical number of non-linear differential equations. How does statistical mechanics approach this problem? By transforming it into an apparently even harder problem, seeking to analyze not just this particular system, but the whole set (ensemble) of systems that share certain properties with this particular system, such as number of particles and energy. The astonishing thing is that it turns out that this apparently much harder problem is more tractable than the original one, and that applying statistical reasoning to the ensemble we can successfully recover from theory precisely the features of gas behavior that we can actually observe and measure, such as temperature and pressure, as emergent self-organizing properties that all of the members of the ensemble share."

(I think there is a somewhat eerie echo here of von Mises' critique of the feasibility of socialism on the ground that successful socialist allocation of resources would require the solution of an impossibly large number of equations expressing marginal equality conditions.)

Please feel free to post this as a response on your blog.

Duncan

Green indicates Duncan's reply to Sanjay's statement

Sanjay:

Does the approach of statistical mechanics / information entropy / econophysics provide a unifying approach that can help satisfactorily to address a very broad range of problems central to social science, by applying a common scientific lens? That may not be the claim being made by its advocates, in which case there is no need for me to address it. However, I understood Duncan Foley to be making precisely a broad claim of this sort on behalf of the approach. If that *is* the claim being made, I would offer a response in the form of two propositions and a conclusion that derives from them (having now attempted to read the methodologically reflective papers offered by Scharfenaker and Foley (https://www.dropbox.com/s/qsbcy3r95gmeov1/QuantalResponseEquilibrium20161208.pdf?dl=0) and by dos Santos (attached)):

Duncan:

I have not actually made this broad claim, since the success of any method in any context depends on demonstrating its success with data generated by that particular phenomenon. I do think that there is a broad class of problems in which there is good reason to believe that observed social outcomes are the result of the interaction of a large number of individual actions, which includes profit rate equalization, Tobin's q, wage- and price setting and inflation, and possibly even more exotic cases such as elections, where it would be worth exploring the effectiveness of these methods.

1. Causality, Description and Explanation: To make an inference of (or a judgment about) the probability of one kind of occurrence conditional upon another is not in and of itself to establish the causal relationship between the two occurrences. Foley and Scharfeneker write that, "we regard conditional distributions as rep- resenting causal relationships and thereby conveying the explanatory content of the model". Whether or not this is, as they state, "general statistical practice", it is a particular understanding of causal relations. To take an example, on their own account, the causal impact understood in this sense of firm entry and exit decisions on the profit rate or vice versa (to use their example) is determined by observed relationships. However, the impact of *potential* entry on profits could not by definition be captured by observations of, or inferences about, the relationship between actual entry and profits.

This seems to me to be true, but I don't see it as a criticism. The effectiveness of statistical equilibrium methods is inseparable from the problem of noisy and limited data. If you can observe potential entry somehow, there is no reason not to use that information to supplement and extend the statistical equilibrium model. In some cases, however, it appears that the usefulness of further data might be quite limited if a model that ignores it successfully explains what we can observe and exhausts all the information in available data. It might be possible to measure the average velocity of molecules in a gas, in addition to their average energy, but a) it would turn out to be close to zero, and b) it would not actually add anything to the explanatory power of the model for the data we can observe.

One might respond that a different kind of observation, that captures the extent or ease of potential entry, is required. That is indeed a way of restoring the interpretation of causes as conditional probabilities, but to concede this is to recognize the need to be attuned to unobserved, and perhaps unobservable, factors, as well as observed ones, in identifying the central descriptive features of a model (and not merely the variation around those features).

This is only an example. The deep-seated and general question that arises here concerns the requirements of understanding. If I am able to identify all of the conditional probabilities of a riot taken place given relevant prior observed or unobserved facts (such as those concerning the relations between members of different groups) does this offer an *explanation*? On the statistical view advanced by the authors, a sufficiently fine-grained account of such probabilities is itself an adequate explanation.

The explanatory power of ensemble reasoning lies in the degree to which it expresses meaningful constraints on the range of alternative possible hypotheses. To my thinking, this means it has content (not everything is possible) even if the ensemble model inevitably cannot tell us exactly which of the consistent possibilities we might have made our incomplete and noisy measurements on.

On another view, a causal account of a riot is a narrative concerning the structured relationship between distinct phenomena over time and space, and across persons. It requires a substantive understanding of the actions of individuals, the reactions of other individuals, the responses or non-responses of the police or others, and so on, each of which may contain internal (subjective) as well as external (objective) elements. Understanding these relevant facts about the riot in light of broader contexts (social, political, economic, cultural) is moreover necessary to anchor the effort at causal explanation. If one were to assert that this task is one of mere identification of conditional probabilities, even if this were not formally speaking wrong it would involve a kind of entropic imperialism, fitting a variable statistical crown to whatever heads can be found.

The riot example is a propos, and in fact in my view quite helpful. We are unlikely ever to recover all the details of the dynamics and interactions that occurred in any particular riot, or near-riot, or non-riot, but there certainly seems to be explanatory power in a theory that identifies, even statistically, the factors that predispose a situation to turn into a riot. Again, if you have more information, and it contributes to this project, so much the better. But we often just don't have anything remotely close to complete information about complex social interactions.

Maximum entropy methods do in principle work in a wide variety of situations, but their empire is limited in the first instance to equilibrium or close-to equilibrium phenomena. When they do not work, they really don't work, and their failure is easy to see.

Sharfenaker and Foley themselves concede that, "In general different constraints correspond to different substantive assumptions about the behavior of individuals and their interaction through social institutions. These substantive questions cannot be avoided purely through the invocation of the maxent formalism." Precisely for this reason, the question of what value is added by the maxent formalism looms large.

This is a somewhat different question because it asks what the significance of maximum entropy principles is. This is quite controversial even in statistical physics, and some physicists are uncomfortable with Jaynes' general program, and prefer other methods (which turn out to be mathematically equivalent) such as drift-diffusion models or ergodic reasoning to reach the same conclusions. There are also alternative entropies to Shannon entropy which some argue are more effective in modeling certain kinds of phenomena that exhibit, for example, long range correlations in time and space. But there are others who prefer to introduce those correlations in constraints.

When the maxent formalism does add value it is because, as Jaynes insists, it insures that the resulting theory incorporates all of the information actually at hand and no more.

The information entropy approach might provide a formal description of certain problems -- in particular those in which law-like behavior of actors (up to some informational constraint) can be presumed, as embodied in the specific Lagrangian maximization program thought to be descriptively relevant (such as maximizing entropy subject to utility being above a certain threshold or maximizing utility subject to entropy being above a certain threshold). But can such description be said to rise to the level of explanation? In my view this is unlikely since supplemental assumptions appear necessary to develop a theory that goes beyond the mere 'what' of description (e..g of a pattern of profit seeking actors entering or exiting industries in response to profit rates while making errors in the process) to encompass the 'how' and the 'why' of explanation (as would for instance an account of the mechanism by which profit rates adjust to reflect such entry or exit).

I think it is a fair question as to whether entropy-constrained payoff maximization, for example, constitutes an "explanation" of the logit quantal response behavior we see in many contexts. Presumably a better explanation would be to identify the sources of the entropy constraints in, for example, the allocation of attention or computing power to decisions. Again, if you have direct observations of these factors, which are often difficult to come by, one would want to use them. We can be pretty confident, however, that if they are consistent with the observations underlying the statistical equilibrium, they will be consistent with the entropy constraints in the statistical equilibrium model.

2. Descriptive Choice (or Description vs. Description): To show that a certain description of a relationship between two kinds of occurrences is possible, or even parsimonious (i.e. that it captures a lot with a little) does not demonstrate that it is a superior description.

Duncan: I agree with this, though I might quibble that parsimony is worth a good deal in scientific explanation.

In providing a description of a social phenomenon involving entropy maximization, the choice of presumed constraints (equivalent to the choice of narratival frame) remains crucial, thus raising a question about whether the supposed generality of the method in fact lies in its being an empty vessel, or indeed whether it is science or art. Emptiness of a vessel can indeed be a 'virtue', but not one on account of which it is superior to a vessel containing fine wine (or for that matter, orange juice) or indeed to some other empty vessel that is available alongside. Since the choice of specific assumptions regarding the constraints (e.g. regarding the relevant properties of the distribution, such as sums, means, or summed deviations) that are to be maintained while maximizing entropy is crucial to the resulting predicted distribution, the exercise as a whole might be more convincingly described as one of retrospective curve fitting than of prediction. (The impression that this is so is not reduced by a detailed reading of the technicalities involved in applying the approach, which involve confronting statements by Sharfenaker and Foley such as that "δ can be estimated by minimizing the Kullback-Leibler divergence between the maximum entropy marginal distribution f[x] and the observed frequency distribution f[x]" which appears in effect to mean that it is precisely retrospective curve fitting that is at the heart of the applied exercise). The 'unreasonable effectiveness' of the method is hardly a surprise from this point of view ("we calculate the informational performance of the fitted distribution, f[x], as capturing approximately 98 percent of the informational content of the observed frequency distribution f[x]" (ibid)).

Here I think we get to a very important point that is not widely understood. I agree that maximizing entropy per se has no content, and is perhaps an "empty vessel". Gell-mann sternly reminded me of this at one point in our discussions when I outlined the conclusion of my reasoning about his paper with Seth Lloyd and he said “but you are neglecting the constraints!”

But there are constraints and constraints. For example, the profit rate distributions (like firm growth distributions) can empirically be modeled to a high degree of fit by “Subbotin” distributions that are the result of maximizing entropy subject to constraints on the mean and mean absolute deviation. The problem is that so far no one has been able to come up with any economically relevant way to interpret the mean absolute deviation constraint. The capitalists have no motive to limit mean absolute deviation. They do have a plausible reason to maximize their own profit rate, and the problem is to show how this motive can lead to equilibrium distributions that have many of the properties of the Subbotin distribution. The model of Smithian competition we propose does in fact solve this problem by positing two measurable relations: a) the tendency of capitalists to enter submarkets with high profit rates and exit submarkets with low profit rates according to a logit quantal response and b) a finite impact of entry and exit decisions on the profit rates in the submarkets. There may, of course, be other ways to address this problem, but this is at least one.

The idea of a social interaction that preserves a central moment, as advanced by Dos Santos, is consistent with but more general than that of conservation of a total quantity (since the latter entails the former but not vice versa), and can provide a description of the constraint that can give structure to an entropy maximization approach. However, in my view this is not a deep discovery about the world. Rather it is a discovery about the possibility of offering a certain (non-unique) description of the world.

Methodological insights are not discoveries about the world. They are heuristics to guide research. But such heuristics are very important in the history of science. Paulo can speak for himself on this, but it seems to me he is developing an important line of thinking in the social theory tradition, which emphasizes the emergence of norms that do not necessarily reflect external physical constraints from social interaction itself.

[Incidentally, although the joint distribution being sought is defined by Foley and Scharfeneker as a statistical equilibrium, it is not clear to me why it deserves the moniker of an equilibrium, if by the use of that word we mean to invoke the mutually reinforcing character of a relationship over time. It is true that given the joint distribution we can derive the conditional distributions. However, the sense in which an equilibrium exists between these is obscure to me and not illuminated by the prior fact. This appears to be an instance of a form of descriptive narrowness that is not entailed by the formalism itself, but I may well have failed to appreciate something)].

I think there are good reasons for regarding the joint distribution as a stable equilibrium. If you unpack the joint distribution into the conditional distributions, it portrays a scenario where a departure of profit rates from equilibrium in any submarket set off a negative feedback effect, first by attracting more entry (or spurring more exit) and second by the effects of entry and exit on the profit rate in that submarket. That looks like an equilibrium stabilized by negative feedback to me.

This is not very different from the usual way economists have taught themselves to think about equilibrium. Supply-demand equilibrium, for example, in theory specifies a unique price-quantity combination. (Because market clearing supply-demand is based on price-taking behavior of both buyers and sellers, the usual development of the theory cannot actually say anything very convincing about stability and wanders off into just-so stories of “auctioneers” or “tatonnement” processes, but that is another issue.) When someone tries to apply the theory to real data, they notice immediately that even when the data have a central tendency that confirms the theory, there are residual deviations of the data from the theoretical predictions. Economists then tend to posit an arbitrary secondary purely statistical process (an econometric specification of "errors") to account for these deviations. Statistical equilibrium in contrast incorporates deviations into the core economic theory itself, and therefore provides a theory of both the central tendency and the distribution of deviations, both of which are stable. In the profit rate case the deviations can be traced to the quantal response of the typical capitalist in the face of profit rate differentials. Operating at a given decision temperature the capitalist pursues higher profit rates, but not with perfect efficiency, and the equilibrium distribution of profit rates reflects this point. If capitalists operated at a decision temperature of zero, which might be one way of thinking of what received economic theory implicitly assumes, then they would instantly move capital in the face of the smallest deviation of profit rates, so the profit rate distribution would have to be degenerate, with all submarkets having the average profit rate.

Conclusion:
The ability of the information entropy approach to facilitate progress in answering the central questions of social science appears to me, at this point in my understanding of it, to be limited.

Duncan: I couldn't really disagree with this formulation. Unlimited improvements are pretty rare in human life.

The approach seems at most to provide a contribution to the armamentarium of social science. Among the problems faced by the information entropy approach are the plurality of possible descriptions (internal to the approach) making of it an empty vessel;

As I've tried to clarify in these remarks, the statistical equilibrium method inherently regards the actual data as reflecting one possible state of the system among an ensemble of states that share certain characteristics. The statistical equilibrium inherently cannot avoid this feature, because of its statistical nature: for example it can't distinguish real-world states in which the statistics of individual actions are the same, but the actions of specific individuals are interchanged. In many circumstances this is a feature, not a bug. Because the statistical equilibrium also rules out some possible configurations of the system, it has falsifiable content.

the difficult of identifying laws (identified constraints) that are invariant over the domain of observation - in the face of ever disruptive social facts such as that of human agency – making it depend on an unwarranted scientistic conceit;

Here I think the statistical equilibrium is on firmer ground. If you accept the "limitation" that the statistical equilibrium can specify only the average or "typical" behavior (which might arise from quite different distributions of the behavior of specific identified individual actors), it does convey important information about dynamics and causality in the form of constraints on the class of models with which the statistical equilibrium is consistent. Since statistical equilibrium incorporates variability in actor behavior in the form of the logit quantal response, which could be replaced by some other description of individual behavior if that is called for in particular problems, it seems to me it does much better than received methods in accommodating "human agency".

and the need to characterize the dynamics that bring about outcomes (mechanisms) in terms of substantive motivations, intentions and processes of interaction in order to arrive at satisfactory causal explanations – a necessity the method might accommodate but not inform.

In the profit rate case, the analysis strongly confirms Smith's conjecture that capitalists at this high level of abstraction seek higher profit rates (not absolute profits or market share, for example). This seems to me to be quite a bit of information about "substantive motivation". It is instructive to compute the statistical equilibrium in the profit rate case on the assumption that capitalists seek to minimize the profit rate, which can be expressed simply by interchanging the entry and exit quantal responses. The equilibrium outcome (there still is one) is dramatically different.

It remains to be seen whether, when allied with other techniques of analysis, it could begin to address these difficulties in a manner that justifies its claim to be more than a dignifying cloak.
The approach of statistical mechanics / information entropy / econophysics may provide a useful tool, in particular for retrospective summary descriptions fitted to data at a high level of abstraction, but does not appear, at least to me, here and now, to offer a way out of our basic conundrum: that of how to choose descriptions and develop explanations that have the *substantive* content that would enable them to appear credible and persuasive in characterizing the causal dynamics that give rise to observations, both when applied to individual cases and to diverse cases considered together.

I don't like to claim the last word, and of course you are right that we have much to learn and understand about statistical equilibrium models in social science. No one is claiming that this method provides complete information about complex social interactions, only that it provides more information than received methods which (usually unwittingly) implicitly assume zero-entropy behavior. The method is not a universal curve-fitting technique, either. Validating specific models on real data does involve comparison of model predictions to observations, as with any scientific theory, but the fact that there is falsifiable content in the statistical models is assured on several grounds, including a) they don't work for some kinds of data on systems that are not near equilibrium and b) the method highlights the importance of the economic meaningfulness of the constraints incorporated.

Email 9: Paulo's reply to Sanjay

Dear Sanjay,

Thanks for your ongoing interventions and critical questions on the application of the Principle of Maximum Entropy (PME) to economic analysis. I agree with you entirely that this application requires very deliberate consideration, especially in light of many rather mechanistic applications of methods from Statistical Mechanics and Information Theory to Economics. As I noted earlier, I am working on a longer, positive statement of how, why and where exactly the PME can help us make some progress in economic analysis—primarily by helping the development of observationally grounded characterizations of the macroscopic functioning of some important economic systems. Its central punchlines are very quickly summarized in the most recent version of my paper on the "Principle of Social Scaling", which I would recommend to all colleagues wishing to delve and perhaps weigh into the important discussions you have initiated.

I do want to take some space here to underscore—in some haste—a few points from the paper concerning the content and significance of the Principle of Social Scaling.

The paper is motivated by the need to offer economically meaningful and plausible accounts for the first-moment constraints implicit in the persistently well-organized cross-sectional distributions of several important economic variables. The physicists who first drew our attention to the presence of these constraints—in distributions of wage income—have argued that underlying "conservation principles" account for those constraints. Most of us would agree this is not at all convincing.

The contention sustaining the Principle of Social Scaling is that in many settings, complex competitive interactions between individuals in markets effectively boil down to *socio-referential* processes. The resulting relationships between individual and social measures of certain variables impose emergent, systemic regulations on individual market outcomes. Under fairly plausible conditions, these regulations can take the form of first-moment constraints on the distribution of variables generated by competitive processes. This suggests first-moment constraints implicit in the patterns of organization we observe for many economic variables may express not "conservation principles," but the emergent social content of certain patterns of economic competition.

Now, I'm not going to speculate on the profundity of this recognition. But I am certain of its usefulness, and of the fact it opens interesting lines of further theoretical and observationally grounded inquiry.

For one, it suggests that what we observe in some economic data is the result of competitive interdependences that are seldom considered explicitly by contemporary economic analysis. This includes the effective process of competition between wage earners that is created by the competitive mobility of capital; the fact that Tobin's q is in effect a measure of the *excess* rates of total return investors expect on the assets of individual corporations, relative to average measures of the rates they expect on the assets of all corporations; or the fact that the monetary outcomes of product-market competition reflect, *inter alia*, a scaling of measures of physical productivity of individual enterprises.

Recognition of these interdependences and the emergent patterns of organization they impose on economic systems casts light on the irreducible social, systemic character of economic competition and outcomes. In this the Principle of Social Scaling embodies and helps vindicate empirically a central contention of Classical Political Economy: That analysis should look beyond immediate appearances, fetishes, and individualist prejudices, to focus on the unintended and oft perverse social outcomes of competition between self-regarding individuals. Here I believe applications of the Principle of Social Scaling can give us new bases to motivate the limitations of methodological individualism, and to demonstrate the purchase of observationally grounded macroscopic or aggregative approaches to economic analysis.

A third indication of the usefulness of the Principle of Social Scaling is given by the promising practical conclusions it sustains in individual applications. This includes, for instance, the development of new systemic diagnostics for the presence of speculative behavior in capital markets (see the attached paper developing an account for the observed distribution of Tobin's q in U.S. markets for a detailed discussion).

Lastly, I wanted to raise a few points about the comparative purchase of social scaling. As you know better than most, empirical inquiry in Economics and Political Economy is almost exclusively observational. As a result, we typically face inverse and ill-posed problems. This means that even more so than in fields where *real* laboratory controls are possible, all our explanations are subject to improvement or refutation. While inductive application of the PME can in some instances help us identify the explanatory burden theories must meet, it offers no bases to choose among accounts offering explanations for the aggregate constraints we infer from observation.

This should not bother us too much, as it is in the very nature of our lines of inquiry. Neither should it in itself take away from any given account offering economic or social insights into the processes generating the patterns we observe. My sense is that we should be flexible and practical in discussing what the "better" or "most accurate" accounts of what we observe really are. Here there is indeed an irreducible element of "art" in our science, as you have suggested.

To my mind, social scaling is a significant explanatory improvement on conservation principles. It is in line with Classical theories of competition, it is grounded on plausible economic mechanisms, and, more practically, it opens the way for interesting new avenues of inquiry, including on the purchase of the suppositions on which it rests. Now, further work may well lead us to reject social scaling in favor of principles or processes offering far deeper explanations of what we are observing. But until social scaling is shown to be implausible on broader economic grounds, or less convincing than alternative mechanisms capable of accounting for the patterns of organization we observe, I don't think we should stop inquiring into its possible empirical purchase and consequences for the functioning of economic systems. Not because I've put the idea forward, but because it is helping us develop and motivate a better understanding of economic processes.

That's all I have, for now. More soon.

Paulo