ECON5529
Bayesian Theory

Ellis Scharfenaker

Lecture 9

Statistical Mechanics and Political Economy

- Economics, and classical political economy in particular, deals with problems of aggregation in complex economic systems with many degrees of freedom.
- Capitalist systems are characterized by decentralized interactions among and between capitalist producers and wage-laborers giving rise to a complex process of production, exchange, and the formation of prices, wages, and profit rates.
- Underlying the complex process of production and exchange is a common organizing logic of competition.

Marx on Averages

- Classical political economists often stressed that individuals interacting in complex ways through institutional structures lead to "unintended consequences."
- [The] sphere [of circulation] is the sphere of competition, which is subject to accident in each individual case; i.e. where the inner law that prevails through the accidents and governs them is visible only when these accidents are combined in large numbers, so that it remains *invisible* and *incomprehensible* to the individual agents of production themselves. (Marx, Vol.III 1981, pp. 967)
- Statistical mechanics concerns systems with a large number of degrees of freedom, for example, with a large number of particles, firms, or households, for which detailed dynamic predictions of individual trajectories are infeasible due to the size of the system.
- The statistical mechanical method substitutes statistical models (models based on probabilistic descriptions of a system) constrained by whatever is known about the system for detailed dynamic predictions.
- The enormous success in physics in dealing with similar problems of large hypothesis and state spaces that is key for rethinking equilibrium in economics is the notion of ***ensemble reasoning***.

The statistical mechanics of a gas

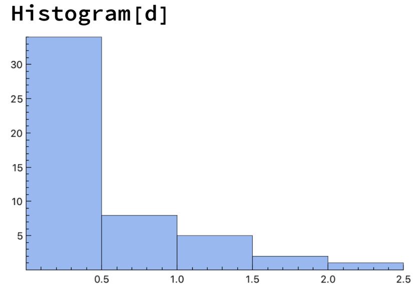
- An elementary problem in statistical physics is describing the state of an "ideal" gas, closed off from external influences, that consists of a very large number, N, of identical rapidly moving particles undergoing constant collisions with each other and the walls of their container.
- The intention of early physicists was to derive certain thermodynamic features at the ***macroscopic*** level (such as temperature or pressure) of the gas from the ***microscopic*** features of the gas, that is, from the configuration of all its particles.

- To describe the microscopic state of this system, however, requires specification of the *position* and *momentum* of each particle in 3-dimensional space. These *6N* coordinates fully describe the microstate the system is in.
- When *N* is very large, e.g. on the magnitude of Avogadro's Number $\approx 6 \times 10^{23}$ (number of atoms of atomic weight *N* in *N* grams of the substance given by one *mole*) the degrees of freedom make such a description a formidable task.

Coarse Graining the State Space

- If we "coarse grain" (i.e. partition) the state space of energy levels into discrete "bins" we can categorize individual particles corresponding to their level of energy associated with that particular bin.
- We can then describe the distribution as a ***histogram vector*** $\{n_1, n_2, \ldots, n_k\}$ where k is the number of bins, n_i is the number of particles in bin i, and $\sum_k n_k = N$, the total number of particles (degrees of freedom).

```
SeedRandom[123];
d = RandomVariate[ExponentialDistribution[2], 50];
Histogram[d]
```




```
HV = HistogramList[d]
```

$\left\{\left\{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}\right\}, \{34, 8, 5, 2, 1\}\right\}$

```
Total@HV[[2]]
```

50

```
system = Transpose@{Range[50], d};
```

- The precise location of each particle within each bin is a description of the ***microstate*** of the system.

```
TableForm[(State1 = Select[system, #[[2]] ≥ 0 && #[[2]] < 1 / 2 &]),
  TableHeadings → {None, {"Particle #", "Energy"}}] // TraditionalForm
"No. of Particles" → Length@State1
```

Particle #	Energy
1	0.392939
2	0.0112118
3	0.0292307
4	0.0192583
6	0.381025
8	0.47642
9	0.422173
10	0.125036
12	0.232395
15	0.0658598
18	0.0424471
21	0.377301
22	0.454112
23	0.0144131
27	0.250646
28	0.198925
30	0.354839
31	0.287288
32	0.29965
33	0.365311
34	0.21983
35	0.436674
36	0.175118
37	0.019718
38	0.0505695
39	0.069347
40	0.229701
41	0.467746
42	0.414727
43	0.282395
44	0.363964
46	0.018678
48	0.350325
49	0.35067

```
No. of Particles → 34
```

```
TableForm[(State2 = Select[system, #[[2]] ≥ 1 / 2 && #[[2]] < 1 &]),
  TableHeadings → {None, {"Particle #", "Energy"}}] // TraditionalForm
"No. of Particles" → Length@State2
```

Particle #	Energy
5	0.598088
13	0.64009
17	0.662585
19	0.886226
24	0.577704
26	0.650499
29	0.708581
45	0.91484

```
No. of Particles → 8
```

```
TableForm[(State2 = Select[system, #[[2]] ≥ 1 && #[[2]] < 3 / 2 &]),
  TableHeadings → {None, {"Particle #", "Energy"}}] // TraditionalForm
"No. of Particles" → Length@State2
```

Particle #	Energy
7	1.39324
14	1.20277
16	1.10771
20	1.1534
25	1.03661

```
No. of Particles → 5
```

```
TableForm[(State3 = Select[system, #[[2]] ≥ 3 / 2 && #[[2]] < 2 &]),
  TableHeadings → {None, {"Particle #", "Energy"}}] // TraditionalForm
"No. of Particles" → Length@State3
```

Particle #	Energy
11	1.51216
47	1.53797

```
No. of Particles → 2
```

```
TableForm[(State4 = Select[system, #[[2]] ≥ 2 &]),
  TableHeadings → {None, {"Particle #", "Energy"}}] // TraditionalForm
"No. of Particles" → Length@State3
```

Particle #	Energy
50	2.197

```
No. of Particles → 2
```

- The histogram describes the ***macrostate*** which is the distribution of energy over the N particles but tells us nothing about the exact microscopic state of the system (location of each particular particle).

```
HV = HistogramList[d]
```

$$\left\{\left\{0, \frac{1}{2}, 1, \frac{3}{2}, 2, \frac{5}{2}\right\}, \{34, 8, 5, 2, 1\}\right\}$$

- Or, for example, if we had only four particles, call them $\{a, b, c, d\}$, and three bins $\{b_1, b_2, b_3\}$, to say "particle a and b are in bin b_1, c is in b_2, and d is in b_3" is a description of the microstate, where as "two particles are in b_1, one particle is in b_2, and one particle is in b_3" is a description of the macrostate.

- Notice *this implies many combinations of particles can lead to the same distribution of particles over bins* so that any histogram will correspond to many microstates.

Multiplicity and Combinatorics

- The number of microstates corresponding to any particular macrostate is the "multiplicity" of that macrostate.
- Continuing with the above example we can calculate multiplicity of the macrostate where two particles are in one bin and the remaining bins each have one particle. The number of ways this macrostate can be realized is:

  (1) $b_1 = \{a, b\}, b_2 = \{c\}, b_3 = \{d\}$,
  (2) $b_1 = \{a, b\}, b_2 = \{d\}, b_3 = \{c\}$,
  (3) $b_1 = \{a, c\}, b_2 = \{d\}, b_3 = \{b\}$,
  (4) $b_1 = \{a, c\}, b_2 = \{b\}, b_3 = \{d\}$,
  (5) $b_1 = \{a, d\}, b_2 = \{b\}, b_3 = \{c\}$,
  (6) $b_1 = \{a, d\}, b_2 = \{c\}, b_3 = \{b\}$,
  (7) $b_1 = \{b, c\}, b_2 = \{a\}, b_3 = \{d\}$,
  (8) $b_1 = \{b, c\}, b_2 = \{d\}, b_3 = \{a\}$,
  (9) $b_1 = \{b, d\}, b_2 = \{c\}, b_3 = \{a\}$,
  (10) $b_1 = \{b, d\}, b_2 = \{a\}, b_3 = \{c\}$,
  (11) $b_1 = \{c, d\}, b_2 = \{a\}, b_3 = \{b\}$,
  (12) $b_1 = \{c, d\}, b_2 = \{b\}, b_3 = \{a\}$

- The combinatorial equation corresponding to such problems is the *multinomial coefficient:*

$$\binom{N}{n_1, n_2, \ldots, n_k} = \frac{N!}{n_1!\, n_2! \ldots n_k!} = \frac{N!}{\left(\left(N \frac{n_1!}{N}\right)\left(N \frac{n_2!}{N}\right) \ldots \left(N \frac{n_k!}{N}\right)\right)} = \frac{N!}{(N\, p_1!\, N\, p_2! \ldots N\, p_k!)} \quad (1)$$

- The numerator is the number of permutations of N and the denominator factors out all permutations with the same particles in a bin to arrive at combinations only.
- For this example we see $\frac{4!}{2!\ \ 1!\ \ 1!} = 12$
- The insight of statistical mechanics is that macrostates of the gas that can be achieved by a large number of microstates, i.e. have higher multiplicity, are more likely to be observed because they can be realized in a greater number of ways (recall the binomial prior from problem set 1).
- For the above example other macrostates might be:
  - all particles in one bin, in which case the multiplicity is $\frac{4!}{4!} = 1$, or

    (1) $b_1 = \{a, b, c, d\}$ or $b_2 = \{a, b, c, d\}$ or $b_3 = \{a, b, c, d\}$

  - two particles in b_1, two in b_2, and zero in b_3 such that $\frac{4!}{2!\ \ 2!} = 6$.

    (1) $b_1 = \{a, b\},\ b_2 = \{c, d\},\ b_3 = \{\}$
    (2) $b_1 = \{a, c\},\ b_2 = \{b, d\},\ b_3 = \{\}$
    (3) $b_1 = \{a, d\},\ b_2 = \{b, c\},\ b_3 = \{\}$
    (4) $b_1 = \{c, d\},\ b_2 = \{a, b\},\ b_3 = \{\}$
    (5) $b_1 = \{c, b\},\ b_2 = \{a, d\},\ b_3 = \{\}$
    (6) $b_1 = \{b, d\},\ b_2 = \{a, c\},\ b_3 = \{\}$

- We can easily see that $\frac{4!}{2!\ \ 1!\ \ 1!}$ = 12 has maximum multiplicity of all macrostates.
- What would need to be true of the system to achieve a multiplicity greater than 12?

Entropy

- For large N, the Stirling approximation Log[N !] ≈ N logN − N is a good approximation to the logarithm of the multinomial coefficient.

$$\mathrm{Log}\left[\frac{N!}{n_1!\,n_2!\ldots n_k!}\right] = \mathrm{Log}\left[N! \Big/ \left(\left(N\,\frac{n_1}{N}\right)!\left(N\,\frac{n_2}{N}\right)!\ldots\left(N\,\frac{n_k}{N}\right)!\right)\right] = \tag{2}$$
$$\mathrm{Log}[N!\,/\,(N\,p_1!\,N\,p_2!\ldots N\,p_k!)] \approx N\,\mathrm{Log}[N] - \sum_{i=1}^{k} N\,p_i\,\mathrm{Log}[N\,p_i]$$

$$= N\,\mathrm{Log}[N] - N\,\mathrm{Log}[N]\sum_{i=1}^{k} p_i - N\sum_{i=1}^{k} p_i\,\mathrm{Log}[p_i] \tag{3}$$
$$\propto -\sum_{i=1}^{k} p_i\,\mathrm{Log}[p_i]$$

- This sum is therefore an approximation of the logarithm of the combinations (multiplicities) and is called *entropy*.
- Since the probabilities p_i sum to one and since the entropy is a concave function of the multiplicity, the entropy corresponding to the macrostate with the largest multiplicity – entropy at its maximum – occurs when all probabilities are equal, $p_i = k / N$, ∀ i.
- As we have learned from information theory, maximum entropy corresponds to maximum uncertainty. Recall the bit length of the uniform distribution was the longest average Shannon code and was equal to -Log[n] where n was the number of messages, which we can now equate with the number of bins in the coarse-grained state space.

Maximum Entropy Inference

- E. T. Jaynes (1957) recognized the generality of Shannon’s result in information theory and discovered this situation is not that different from that in statistical mechanics, where the physicist must assign probabilities to various energy states.
- He argued that the number of possibilities in either case was so great that the frequency interpretation of probability would clearly be absurd as a means for assigning probabilities.
- Instead, he showed that the equation for entropy is a measure of information that is to be understood as a measure of uncertainty or ignorance.
- Jaynes argued that imposing constraints on systems that change the maximum entropy distribution are just instances of using information for inference.
- Inferring the maximum entropy distribution subject to information about the system is what Jaynes referred to as the ***principle of maximum entropy inference*** (PME).
- The ensemble of “bins” and their relative probabilities describe a certain state of knowledge and for this reason Jaynes believed the connection between Shannon’s information theory and statistical mechanics was that the former justified the later.

Information as Constraints

- Maximum entropy distributions use *information* typically in the form of moment constraints. A constraint is anything that modifies a probability distribution.

- Intuitively, if you are completely certain about an event entropy (uncertainty) is at a minimum because its probability is 1 implying $H[p] = -p \operatorname{Log}[p] = 1 \operatorname{Log}[1] = 0$.
- On the other side of the spectrum, complete ignorance corresponds to maximum entropy (uncertainty) which we know corresponds to a uniform probability assignment $H[p] = -\operatorname{Log}[n]$
- In an economic context, examples of information we use to impose constraints (e.g. model closures) are a budget constraint, market clearing, the non-negativity of prices, stock-flow consistency, savings equal to investment, behavioral constraints such as utility maximization, conservation of value in exchange, full employment, functional distribution of income, equalization of the rate of profit, etc. So long as these constraints are binding we should expect to find persistent macroeconomic phenomena consistent with the PME.

Maximizing Entropy

- In the case of maximizing entropy in a bounded interval $[1, n]$ in the state space of x, the MAXENT distribution is the uniform distribution.
- Maximizing entropy subject only to a normalization constraint and bounded interval is expressed in the programming problem:

$$\begin{aligned} &\operatorname{Max} H[p] = -\sum_i p_i \operatorname{Log}[p_i] \qquad p_i \geq 0 \ \forall\, i, \quad i = 1, \cdots, n \\ &\qquad s.t. \sum_{i=1}^{n} p_i = 1 \end{aligned} \tag{4}$$

- The associated Lagrangian is:

$$\mathcal{L} = -\sum_{i=1}^{n} p_i \operatorname{Log}[p_i] - \lambda \sum_{i=1}^{n} p_i \tag{5}$$

- The first order conditions require:

$$\frac{\partial \mathcal{L}}{\partial p_i} = -\operatorname{Log}[p_i] - 1 - \lambda = 0 \tag{6}$$

- Solving for p_i,

$$p_i = e^{-1-\lambda} \tag{7}$$

- Plugging into the normalization constraint:

$$\sum_{i=1}^{n} e^{-1-\lambda} = 1 \tag{8}$$

- We get the uniform distribution

$$e^{-1-\lambda} = \frac{1}{n} \tag{9}$$

- The uniform distribution has an entropy

$$H[p_i] = -\sum_i p_i \operatorname{Log}[p_i] = -\sum_i \frac{1}{n} \operatorname{Log}\left[\frac{1}{n}\right] = -\operatorname{Log}[n] \tag{10}$$

- What would happen if we maximized entropy over the open interval $(-\infty, \infty)$?

Statistical Equilibrium

- The extraordinary discovery of statistical mechanics consisted in the realization that while the microscopic state of the system was highly disordered, its statistics expressed in terms of energy level histograms tend to “relax” very rapidly to an *equilibrium distribution*.
- It is then possible to identify the parameters of the equilibrium distribution with phenomenal observables like pressure and temperature.
- Such a ***statistical equilibrium distribution*** is defined as the macrostate that corresponds to the largest number of microstates, or equivalently, as the maximum entropy distribution consistent with any constraints imposed on the system.

Maximum Entropy in Thermodynamics

- The momentum of each particle determines its *kinetic energy*. If a system is isolated, the kinetic energy will determine the total energy which will not change over time (first law of thermodynamics).
- Boltzmann (1871) used ensemble reasoning to find the most probable macrostate (distribution of energy) of a gas constrained, for example, only by its energy. If the energy states (bins) are $\{x_1, \ldots, x_k\}$ and the average energy per particle is $\langle x \rangle$, the maximum entropy program is:

$$\begin{aligned} &\underset{\{p_i \geq 0\, \forall i\}}{\text{Max}}\; H[p] = -\sum\nolimits_i p_i \operatorname{Log}[p_i] \\ &\text{subject to} \sum\nolimits_i p_i = 1 \\ &\qquad\qquad\;\; \sum\nolimits_i p_i\, x_i = \langle x \rangle \end{aligned} \tag{11}$$

- This is a mathematical programming problem quite familiar to economists. The Lagrangian associated with this constrained-maximization problem is

$$\mathcal{L}[p;\, \lambda,\, \mu] = -\sum\nolimits_i p_i \operatorname{Log}[p_i] - \mu\left(\sum\nolimits_i p_i - 1\right) - \lambda\left(\sum\nolimits_i p_i\, x_i - \langle x \rangle\right) \tag{12}$$

- Entropy is a strictly concave function of the probability distributions describing the system. Since the constraints are typically linear or convex functions, the first-order conditions are *necessary* and *sufficient* for a maximum.

$$\begin{aligned} &\frac{\partial \mathcal{L}}{\partial p_i} = -\operatorname{Log}[p_i] - 1 - \mu - \lambda\, x_i = 0 \\ &p_i = e^{-(1+\mu)}\, e^{-\lambda\, x_i} \end{aligned} \tag{13}$$

- Plugging this solution into the normalization constraint:

$$\begin{aligned} &\sum\nolimits_i e^{-(1+\mu)}\, e^{-\lambda\, x_i} = 1 \\ &e^{-(1+\mu)} \sum\nolimits_i e^{-\lambda\, x_i} = 1 \\ &e^{-(1+\mu)} = \frac{1}{\sum_i e^{-\lambda\, x_i}} \\ &p_i = \frac{e^{-\lambda\, x_i}}{\sum_i e^{-\lambda\, x_i}} \propto e^{-\lambda\, x_i} \end{aligned} \tag{14}$$

- The family of distributions of this form are known as the exponential family and Eq. 7 is called the Maxwell-Boltzmann-Gibbs distribution. The function $Z[\lambda] = \sum_i e^{-\lambda\, x_i}$ is called the *partition function.*
- The partition function allows us to calculate all the thermodynamic properties of the system.

$$-\frac{\partial Z[\lambda]}{\partial \lambda} = \langle x \rangle \text{ Thermodynamic energy}$$
$$-\frac{\partial^2 Z[\lambda]}{\partial \lambda^2} = \left\langle (\Delta x)^2 \right\rangle \text{ Energy variance (fluctuations)} \tag{15}$$

- The constrained maximization problem implicitly calculates shadow prices for each of the constraints, which are uniform over the subsystems and characterize its properties in equilibrium.
- The Lagrange multiplier or shadow price λ has the dimensions of entropy per unit of energy; its inverse is *temperature*, which can be measured in Kelvins (from absolute zero) as $k\,T = \frac{1}{\lambda}$ where k is Boltzmann's constant.
- We can see that any number of macrostates for which the constraints agree with the information represents a possible distribution, compatible will all this is specified.
- Out of the millions of such possible distributions, the one that is most likely to be realized is that which is most probable, i.e. the one that can be realized in the greatest number of ways, or equivalently, the macrostate with the greatest multiplicity.

The Pareto Distribution

- If we maximize entropy of a system defined over the support $[x_{\min}, \infty)$ and constrained to have a given geometric mean the MAXENT program is:

$$\begin{aligned} \text{Max } H[p] \; &= -\int_{x_{\min}}^{\infty} p[x] \,\text{Log}[p[x]] \,dx \\ s.t \; &\int_{x_{\min}}^{\infty} p[x] \,dx = 1 \text{ and} \\ &\int_{x_{\min}}^{\infty} p[x] \,\text{Log}[x] \,dx = \hat{x} \end{aligned} \tag{16}$$

- The associated Lagrangian is:

$$\mathcal{L} = -\int_{x_{\min}}^{\infty} p[x] \,\text{Log}[p[x]] \,dx - \mu\left(\int_{x_{\min}}^{\infty} p[x] \,dx - 1\right) - \lambda\left(\int_{x_{\min}}^{\infty} p[x] \,\text{Log}[x] \,dx - \hat{x}\right) \tag{17}$$

- The first order conditions require:

$$\frac{\partial \mathcal{L}}{\partial p[x]} = -\text{Log}[p[x]] - 1 - \mu - \lambda \text{Log}[x] = 0 \tag{18}$$

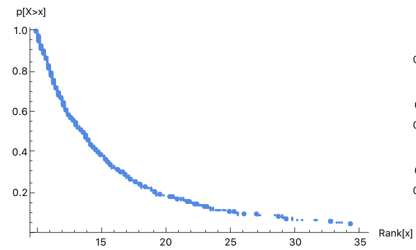
- Solving for $p[x]$,

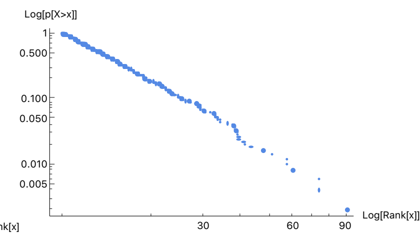
$$p[x] = e^{-(1-\mu)}\, e^{-\lambda \,\text{Log}[x]} \tag{19}$$

- Plugging this solution into the normalization constraint:

$$\begin{aligned} &\int_{x_{\min}}^{\infty} e^{-(1+\mu)}\, e^{-\lambda \,\text{Log}[x]} \,dx = 1 \\ &e^{-(1+\mu)} \int_{x_{\min}}^{\infty} e^{-\lambda \,\text{Log}[x]} \,dx = 1 \\ &e^{-(1+\mu)} = 1 \Big/ \left(\int_{x_{\min}}^{\infty} e^{-\lambda \,\text{Log}[x]} \,dx\right) \\ &p[x] = e^{-\lambda \,\text{Log}[x]} \Big/ \left(\int_{x_{\min}}^{\infty} e^{-\lambda \,\text{Log}[x]} \,dx\right) = c\, e^{-\lambda \,\text{Log}[x]} = c\, x^{-\lambda} \quad \text{for } \; x \geq x_{\min} \end{aligned} \tag{20}$$

- The statistical equilibrium distribution for this constraint is the Pareto distribution or power law. A power law is a scale invariant distribution.
- Notice Log[p[x]] = Log[c] – λ Log[x] which is linear on log scale.
- A common way of identifying power laws in empirical data is plot the *complementary cumulative distribution function (CCDF).* A CCDF curve shows e.g. how many people are at or above a given level of income.





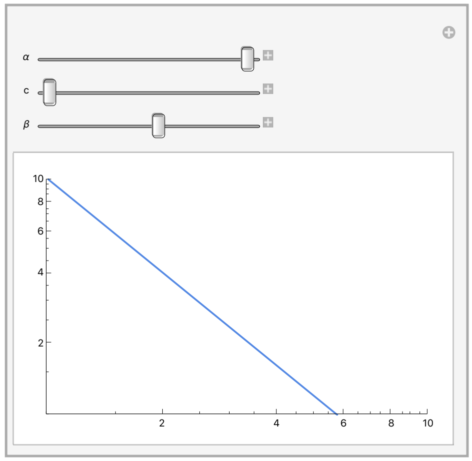
- When $\lambda \approx 1$ a particular case of the power law emerges which is called Zipf's law. Power laws have been observed in many empirical size distributions in economics such as city sizes, word frequencies, proportional random growth, income, and wealth.
- As Schumpeter remarked in 1949 while discussing the power law, "Few if any economists seem to have realized the possibilities that such invariants hold for the future of our science. In particular, nobody seems to have realized that the hunt for, and the interpretation of, invariants of this type might lay the foundations for an entirely novel type of theory."

Two-Class Income Distribution

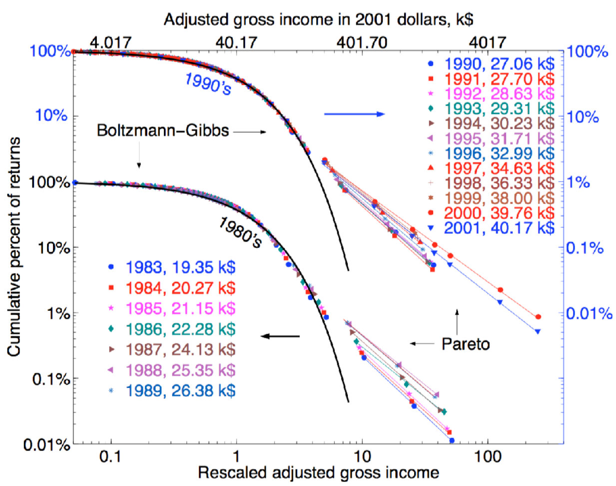
- The modeling of economic size distributions originated around 1897 with the work of Vilfredo Pareto on the distribution of income.
- Pareto noticed that for many countries a logarithmic plot of the number of incomes N above $x_{\min}$ against the logarithm of x yields points on a straight line with slope $-\lambda$. A lower λ meant a higher degree of inequality in the distribution.
- Power laws are an example of a **scaling law** described by relation of the type $y = \alpha\, x^{\beta}$. The same functionality exists across the spectrum for a scaling law. If you scale x by some constant, say 10, then we see that

$$y = \alpha(10\, x)^{\beta} = \alpha\, 10^{\beta}\, x^{\beta} \tag{21}$$

- That is, y scales by 10^{β}.



- The physicist Victor Yakovenko examined IRS data and found that income actually follows a two-class distribution.



Other Well Known MAXENT Distributions

- The principle of maximum entropy can be used to derive most well known distributions given appropriate constraints. Some examples are:

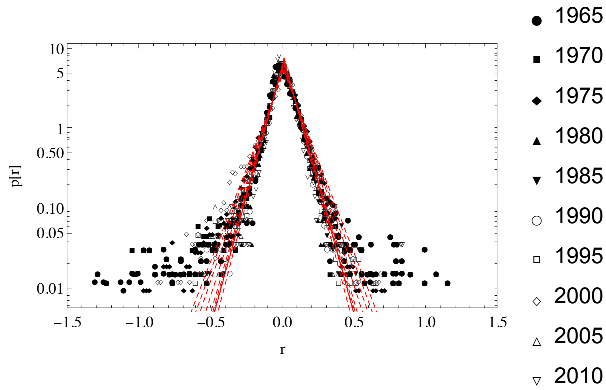
Constraint	Support	Distribution
E[x]	[0, ∞)	Exponential
E[Log[x]]	[k, ∞)	Pareto
E[|x|]	(−∞, ∞)	Laplace
E[x], E[Log[x]]	[0, ∞)	Gamma
E[x], E[x^2]	(−∞, ∞)	Normal
E[Log[(1 + x^2)]]	(−∞, ∞)	Cauchy
E[Log[x]], E[Log[x]^2]	[0, ∞)	Log Normal
E[x], E[|x|]	(−∞, ∞)	Asymetric Laplace
E[x]	[0, ∞), x ∈ ℕ	Geometric

Incomplete Information

- The essence of Jaynes' argument is that when we make inferences based on incomplete information, we should draw them from that probability distribution that has the maximum entropy permitted by the information we do have.
- The maximum entropy method is useful when we do not have complete information on the state of a system.
- Maximizing entropy subject to constraints that express what information we do have results in a probability distribution over states of the system that best expresses our state of knowledge of the system.
- The principle of maximum entropy is so powerful because it is based on a combinatorial theorem that says to choose any other distribution would amount to ignoring the vast majority of all possibilities allowed by the data.
- Intuitively, distributions with higher entropy (multiplicity) can be realized in a greater number of ways. Because the maximum entropy distribution can be realized in so many ways, as Max Plank put it, Nature will appear to have a "strong preference for situations of higher entropy".
- An important consequence of the MAXENT distribution is that "if the predictions prove to be wrong, then induction has served its real purpose; we have learned that our hypotheses are wrong or incomplete, and from the nature of the error we have a clue as to how they might be improved... As Harold Jeffreys explained long ago, induction is most valuable to a scientist just when it turns out to be wrong; only then do we get new fundamental knowledge." E. T. Jaynes (2003, p. 311)

Statistical Equilibrium Economics: Profit Rates

- Adam Smith argued that it was the competitive disposition of capital to persistently seek higher rates of profit and that through competition a tendency of the equalization of profit rates across all competitive industries would emerge.
- There are two parts to Smith's theory:
  - Individual capitalists, who take profit rates in particular sectors of production as determined by forces beyond their control, move capital from low profit rate sectors to high profit rate sectors.
  - The movement of capital into (out of) a sector tends to lower (raise) the profit rate in that sector
- The important point is that the tendential gravitation of each industry's rate of profit is an unintended consequence of the entry/exit decisions of many individual firms.
- Resolution of the process of competition emerges at the macroscopic scale as captured by the distribution of the prime mover of capital, the rate of profit.



- Both of these factors are essential to the process of convergence of profit rates.
  - If individual capitalists don't pay any attention to profit rates in deciding where to invest, their investments would be random and there would be no tendency toward profit rate convergence.
  - If the movement of capital into or out of a sector had no impact on the profit rate in the sector through competition, even if capitalists do seek higher profit rates, their actions would not lead to a convergence of profit rates.
- These considerations indicate that we need to think of the statistical equilibrium model of the profit rate in terms of a joint frequency distribution over the actions of enter and exit, and the distribution of profit rates among sectors of production, $f[\{\text{entry}, \text{exit}\}, r]$.

Statistical Equilibrium Economics: Profit Rates

- We have a system in which an outcome, r, is brought into statistical equilibrium by actions, $a = \{\text{enter}, \text{exit}\}$, of participants in an institutional structure of competition.
- An equilibrium of the system is represented by a joint frequency, $f[a, r]$, which determines marginal frequencies, $f[a] = \int f[a, r]\, dr$, $f[r] = \sum_a f[a, r]$, and conditional frequencies, $f[a \mid r] = \frac{f[a,r]}{f[r]}$, $f[r \mid a] = \frac{f[a,r]}{f[a]}$, interpretable as the causal forces constituting the statistical equilibrium.
- Data often provides only limited information about the statistical equilibrium, for example, observations on the marginal frequency of outcomes, $f[r]$, which must be supplemented by theoretical prior constraints in order to make inferences about the joint frequency $f[a, r]$.
- In this case, the data are in the macroscopic *effects* domain, but our interest also lies in the microscopic *causal* domain.
- This ill-posed underdetermined problem of inference is an ideal situation for employing the PME to infer the missing information.
- The statistical equilibrium of this type of system represents the interaction of two real forces. First, the capitalist firms moving capital into and out of particular sectors of production have to respond to outcome signals (profit rate differentials) by actually acting (entering or exiting subsectors of production).
- This means that the conditional action frequency, $f[a \mid r]$, must respond to the level of the outcome variable, r.

- Second, there has to be some impact of actions (entry and exit decisions) on the outcome variable (entry reducing profit rates and exit increasing them), so that the action behavior effectively feeds back on the outcome distribution.
- This means that the conditional outcome frequencies, $f[r \mid a]$, must vary with the action.

Firm Behavioral Constraint

- The condition that $f[r \mid a] \neq f[r]$ and $f[a \mid r] \neq f[a]$ is a just an example of the information we supply as constraints. In this case, we are implying a mutual conditional dependence of entry/exit decisions and the rate of profit.
- First consider a constraint on $f[a \mid r]$. Theory suggests individual capitalists, who take profit rates in particular sectors of production as determined by forces beyond their control, move capital from low profit rate sectors to high profit rate sectors.
- We can consider $f[a \mid r] : a \times r \to (0, 1)$ as the firms' mixed strategy over actions $a = \{\text{enter}, \text{exit}\}$. If firms choose the action that maximizes their expected payoff $\sum_a f[a \mid r]\, r$ the distribution over actions is the solution to:

$$\begin{aligned} &\underset{\{f[a \mid r] \geq 0\}}{\text{Max}} \sum\nolimits_{\{\text{enter,exit}\}} f[a \mid r]\,(r - \mu) \\ &\text{subject to } \sum\nolimits_{\{\text{enter,exit}\}} f[a \mid r] = 1 \end{aligned} \tag{22}$$

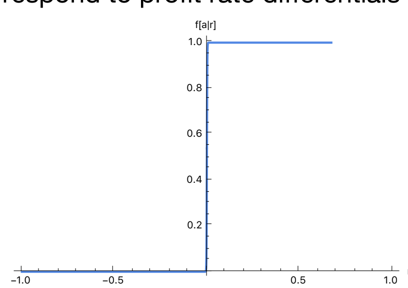
- The solution to this programming problem is the Dirac delta distribution:

$$\hat{f}[a \mid r] = \delta_x\left[a - \hat{a} \mid (r - \mu)\right] \tag{23}$$

- In this case the firms will always put all weight on the payoff-maximizing action and zero weight on any other action.
- As a probability measure on $\mathbb{R}$, this is characterized by the Heaviside unit step function:

$$\theta[a \mid r] = \begin{cases} 0 & \text{if } r \geq \mu \\ 1 & \text{if } r < \mu \end{cases} \tag{24}$$

- This implies that firms are able to respond to profit rate differentials instantaneously.



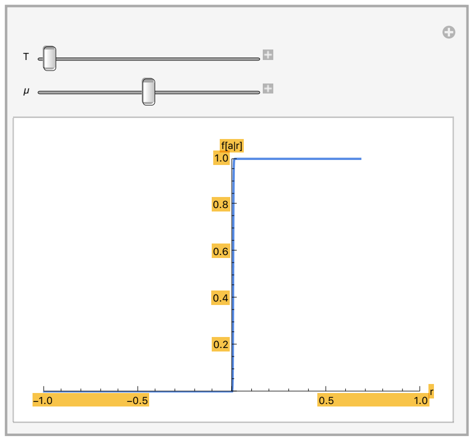
- From thermodynamic perspective, firms are operating at a zero behavioral temperature. From an information theoretic perspective, this implies firms are processing market information with infinite bandwidth and zero uncertainty.
- To introduce variability in choice we can make the no so unrealistic assumption that firms face a minimum entropy constraint.

$$\begin{aligned}&\underset{\{f[a\mid r]\geq 0\}}{\text{Max}} \sum\nolimits_{\{\text{enter},\text{exit}\}} f[a\mid r]\,(r-\mu)\\ &\text{subject to } \sum\nolimits_{\{\text{enter},\text{exit}\}} f[a\mid r]=1\\ &\text{and} \quad -\sum\nolimits_{\{\text{enter},\text{exit}\}} f[a\mid r]\,\text{Log}[f[a\mid r]]\geq H_{\min}\end{aligned} \tag{25}$$

- The solution to this programming problem is called the "logit quantal response" distribution:

$$f[a\mid r]=\frac{e^{-\frac{r-\mu}{T}}}{1+e^{-\frac{r-\mu}{T}}} \tag{26}$$

- Where T is the shadow price of information or "behavioral temperature."



Profit Rate Impact Constraint

- Now consider the constraint on the conditional distribution $f[r \mid a]$. Here theory suggest the movement of capital into (out of) a sector tends to lower (raise) the profit rate in that sector.
- The impact of this negative feedback can be expressed by constraining the difference between the part of the mean outcome due to agents who take the action, $E[r \mid \text{enter}]\, f[\text{enter}]$, and the part due to agents who do not take the action, $E[r \mid \text{exit}]\, f[\text{exit}]$:

$$\begin{aligned}&E[r\mid \text{enter}]\,f[\text{enter}]-E[r\mid \text{exit}]\,(1-f[\text{exit}])\\ &=\int_r f[\text{enter},r]\,r\,dr-\int_r f[\text{exit},r]\,r\,dr\\ &=\int_r \frac{1}{1+e^{-\frac{r-\mu}{T}}}f[r]\,r\,dr-\int_r \frac{e^{-\frac{r-\mu}{T}}}{1+e^{-\frac{r-\mu}{T}}}f[r]\,r\,dr\\ &=\int_r \text{Tanh}\left[\frac{r-\mu}{2T}\right]f[r]\,r\,dr=\delta\end{aligned} \tag{27}$$

- The smaller is δ, the more effective are entry and exit in changing profit rates, so δ is an indirect measure of the effectiveness of competition.

MAXENT Distribution

- Since we are interested in a statistical model of $f[a, r]$ we can maximize the joint entropy $-\sum_a \int_r f[a, r]\, \text{Log}[a, r]\, dr$ subject to the two constraints on the conditional distributions.

- It’s well known that the joint entropy can be written as the entropy of the marginal distribution plus the average entropy of the conditional distribution (just use the product rule):

$$-\sum_a \int_r f[a, r] \, \mathrm{Log}[a, r] \, dr = -\int_r f[r] \, \mathrm{Log}[f[r]] \, dr - \int_r f[r] \sum_a f[a \mid r] \, \mathrm{Log}[f[a \mid r]] \, dr \tag{28}$$

- Since we know $f[a \mid r] = \frac{e^{-\frac{r-\mu}{T}}}{1+e^{-\frac{r-\mu}{T}}}$ let $H_{T,\mu}[r] = -\left(\frac{1}{1+e^{-\frac{r-\mu}{T}}} \mathrm{Log}\left[\frac{1}{1+e^{-\frac{r-\mu}{T}}}\right] + \frac{e^{-\frac{r-\mu}{T}}}{1+e^{-\frac{r-\mu}{T}}} \mathrm{Log}\left[\frac{e^{-\frac{r-\mu}{T}}}{1+e^{-\frac{r-\mu}{T}}}\right]\right)$. The MAXENT problem becomes:

$$\begin{aligned} &\underset{\{f[r]\}}{\mathrm{Max}} \; -\int_r f[r] \, \mathrm{Log}[f[r]] \, dr + \int_r f[r] \, H_{T,\mu}[r] \, dr \\ &\text{subject to} \; \int_r f[r] \, dr = 1 \\ &\qquad \int_r f[r] \, r \, dr = \xi \\ &\qquad \int_r \mathrm{Tanh}\left[\frac{r-\mu}{2T}\right] f[r] \, r \, dr \leq \delta \end{aligned} \tag{29}$$

- The Lagrangian associated with this programming problem is:

$$\begin{aligned} \mathcal{L}[f[r], \lambda, \gamma, \beta] = &-\int_r f[r] \, \mathrm{Log}[f[r]] \, dr + \int_r f[r] \, H_{T,\mu}[r] \, dr - \\ &\lambda\left(\int_r f[r] \, dr - 1\right) - \gamma\left(\int_r f[r] \, r \, dr - \xi\right) - \beta\left(\int_r \mathrm{Tanh}\left[\frac{r-\mu}{2T}\right] f[r] \, r \, dr - \delta\right) \end{aligned} \tag{30}$$

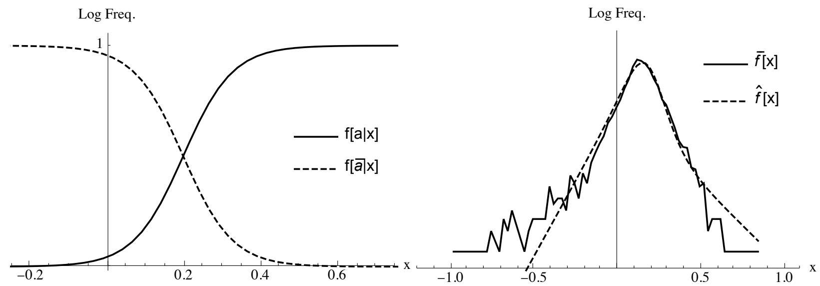
- The first-order conditions for maximizing entropy of the joint frequencies and conditional frequency require

$$f[r] = \left(e^{H_{T,\mu}[r]} \, e^{-\gamma r} \, e^{-\beta \, \mathrm{Tanh}\left[\frac{r-\mu}{2T}\right] f[r] \, r}\right) \Big/ \left(\int e^{H_{T,\mu}[r]} \, e^{-\gamma r} \, e^{-\beta \, \mathrm{Tanh}\left[\frac{r-\mu}{2T}\right] f[r] \, r} \, dr\right) \tag{31}$$

- This the maximum entropy (statistical equilibrium) distribution of the rate of profit.

MAXENT Distribution

- The MAXENT program would like to make the marginal profit rate frequency uniform, in order to maximize entropy, but this does not fit the tent-shaped data distribution.
- Once the frequencies of entry and exit conditional on the profit rate are constrained to be logit quantal responses, a uniform marginal profit rate distribution will imply the largest possible difference between expected profit rates conditional on entry and exit.
- But we also constrain the difference between the expected profit rate conditional on entry and the average profit rate (which implicitly also determines the expected profit rate conditional on exit).
- The MAXENT program can meet this constraint only by “bending” the marginal profit rate distribution around the average profit rate in order to reduce the difference between the expected rates conditional on entry and exit.
- This effect represents the impact of competition in reducing profit rates as the result of entry and increasing them as the result of exit.
- The predictive relevance of maximum entropy inference is conditional on the ability of the statistical model to produce observable regularities in the system under analysis.
- In the profit rate case, the model predicts the distribution of profit rates extremely well in addition to predicting the unobserved quantal actions of firms.



- This is just the probability distribution that achieves maximum entropy over the relevant domain and is a logically equivalent representation of our state of knowledge. That is, the maximum entropy distribution expresses what we know while being maximally non-committal toward what we do not.

MAXENT Dynamics

- A statistical equilibrium model does not determine any unique individual dynamics, but they do ensure that the estimated statistical equilibrium will be in the high probability manifold of the state space.
- That is, the model provides a complete dynamic description of all individual behavior through the ensemble averages. These tell us which dynamics are most likely given the configuration of the state space, not the deterministic time path of a representative individual that can be aggregated over to provide macroscopic properties of the system.