
The MaxEnt (Best Guess) Machine

Many useful thermodynamic results can be obtained from a knowledge of "number of accessible states" Ω as a function of work parameters like energy E , volume V , and number of atoms N . These include, for example, the law of equipartition for quadratic systems, the equation of state for an ideal gas, and the law of mass action for tracking chemical reaction equilibria. This note is designed to move beyond those results to cases (Gibbs called them ensembles) for which one has information about "expected averages". That information can be used to modify the assignment of equal *a priori* probabilities beyond simple maximization of $S = k\text{Log}[\Omega]$ in the microcanonical ensemble. Our notation here also treats energy as but one of many possible constraints in the statistical inference problem, so the results won't be restricted to their usual thermodynamic use in the study of heat.

One first writes entropy in terms of probabilities by defining for each probability a "surprisal" $s_i = k\text{Log}\left[\frac{1}{p_i}\right]$, in units of {bits, nats, hartleys, J/K} if k is respectively: $\left\{\frac{1}{\text{Log}2}, 1, \frac{1}{\text{Log}10}, 1.38 \times 10^{-23}\right\}$. Then entropy, i.e. the average value of this surprisal, becomes $S = k\text{Log}[\Omega]$ when the p_i are all equal. *Note:* Here $\text{Log}[\]$ refers to the natural and not the base-10 log. These relations translate seamlessly to continuous distributions for classical calculation (cf. Chapter 2 of Plischke and Bergersen) and density matrices for QM application (cf. Jaynes in *Phys Rev* **108** (1956) 171).

The problem

Our job is to maximize average surprisal S when all Ω accessible states are not equally probable...

$$\frac{S}{k} = \left\langle \frac{s}{k} \right\rangle = \sum_{i=1}^{\Omega} p_i \frac{s_i}{k} = \sum_{i=1}^{\Omega} p_i \text{Log} \left[\frac{1}{p_i} \right] \quad (1)$$

...subject to the usual normalization requirement that the probabilities add to 1, i.e. that...

$$\sum_{i=1}^{\Omega} p_i = 1, \quad (2)$$

along with the "expected average" constraints which for the r th of R constraints might take the form...

$$E_r = \langle e_r \rangle = \sum_{i=1}^{\Omega} p_i e_{ri}, \quad \forall r \in \{1, R\}. \quad (3)$$

The solution

The Lagrange method of undetermined multipliers tells us that the solution for the i th of Ω probabilities is simply...

$$p_i = \frac{1}{Z} e^{-\sum_{r=1}^R \lambda_r e_{ri}}, \quad \forall i \in \{1, \Omega\}, \quad (4)$$

where partition function Z (a kind of "constrained multiplicity") is defined to normalize probabilities as...

$$Z = \sum_{i=1}^{\Omega} e^{-\sum_{r=1}^R \lambda_r e_{ri}}. \quad (5)$$

Here λ_r is the Lagrange (or "heat") multiplier for the r th constraint, and e_{ri} is the value of the r th parameter when the system is in the i th accessible state. For example, when E_r is the energy E , λ_r is often written as $\frac{1}{kT}$. Values for these multipliers can be calculated by substituting the two equations above back into the constraint equations (2) and (3), or from the differential relations derived below.

A few "big picture" quantities

The resulting entropy (maximized with constraints) is...

$$\frac{S}{k} = \text{Log} [Z] + \sum_{r=1}^R \lambda_r E_r . \quad (6)$$

We can rearrange this expression (following Gibbs) by defining the *dimensionless availability* in natural units as...

$$A \equiv -\text{Log} [Z] = \sum_{r=1}^R \lambda_r E_r - \frac{S}{k} . \quad (7)$$

This quantity was minimized without constraint, and serves as common numerator for one dimensioned availability for each extensive variable E_r , namely $A_r \equiv A/\lambda_r$. For example, the "energy availability" is Helmholtz free energy $E-TS$, if $R=1$ and E_1 is a constraint on energy E . These dimensioned availabilities have also been minimized for a given value of their corresponding heat multiplier λ_r , assuming of course that the e_{ri} coefficients are held constant themselves (cf. Betts and Turner p.46).

Familiar ensemble contexts for this calculation include the Microcanonical Ensemble ($R=0$ so that $Z=S/k$), the Canonical Ensemble ($R=1$ and E_1 is energy E so that $\lambda_1 = 1/kT$ and $A_1=E-TS$), the Pressure Ensemble ($R=2$ with E_1 energy, $\lambda_1 = 1/kT$, E_2 volume, $\lambda_2=P/kT$, and $A_1=E+PV-TS$ =Gibbs Free Energy), and the Grand Canonical Ensemble (same as pressure except that $E_2=N$, $\lambda_2=-\mu/kT$, and $A_1 = E - \mu N - TS = \Omega_g =$ The Grand Potential).

There is much more to go, as the process of transcribing and synthesizing disparate notes continues. Along the way, I suspect that a better understanding of the validity and limits of "altered looks" at heat (and other) capacities, as bits of information lost per two-fold increase in the corresponding extensive variable or it's Lagrange multiplier, will emerge (*Amer J Phys* **71**:1142-1151).

Internal Work, Heat, and Irreversibility

Values of the e_{ri} , which represent the values of parameter r associated with the i th alternative state, are often not themselves constant but instead depend on the value of a set of "work parameters" X_m , for values of m between 1 and M . For example in the Canonical Ensemble case, the energies of the various allowed states may depend on volume V or particle number N . In that case we can define work-types for each constraint r in terms of the rate at which E_r changes with X_m as follows...

$$-\delta W_r = \sum_{m=1}^M \delta X_m \left(-\frac{\partial E_r}{\partial X_m} \right)_{X_{s \neq m}, \lambda_r} = -\sum_{m=1}^M \delta X_m \sum_{i=1}^{\Omega} \left(e_{ri} \frac{\partial p_i}{\partial X_m} + p_i \frac{\partial e_{ri}}{\partial X_m} \right), \quad \forall r \in \{1, R\} \quad (8)$$

Note that the various "work increments" δW_r have the same units as the constrained parameters E_r to which they correspond. Also, we have left open the possibility that changes in X_m may alter probabilities directly, e.g. by making new volume available for free expansion, rather than simply via their effect on the state parameters e_{ri} . As we will show, this allows us to mathematically incorporate "irreversible" changes in entropy by averaging this term over all work parameters and all constraints...

$$\frac{\delta S_{irr}}{k} = \sum_{r=1}^R \lambda_r \sum_{m=1}^M \delta X_m \sum_{i=1}^{\Omega} e_{ri} \left(\frac{dp_i}{dX_m} \right)_{X_{s \neq m}, \lambda_r} . \quad (9)$$

If we further define "heat increments" δQ_r , of the r th type as...

$$\delta Q_r = \sum_{i=1}^{\Omega} e_{ri} \sum_{u=1}^R \delta \lambda_r \left(\frac{dp_i}{d\lambda_u} \right)_{\lambda_{s \neq u}, X_m}, \quad \forall r \in \{1, R\}, \quad (10)$$

we can then obtain from the definitions above a couple of familiar differential relationships...

$$\delta Q_r - \delta W_r = \sum_{i=1}^{\Omega} (e_{ri} \delta p_i + p_i \delta e_{ri}) = \delta E_r, \quad \forall r \in \{1, R\}, \quad (11)$$

and

$$\sum_{r=1}^R \lambda_r \delta Q_r + \frac{\delta S_{irr}}{k} = \sum_{i=1}^{\Omega} \delta p_i \sum_{r=1}^R \lambda_r e_{ri} = \frac{\delta S}{k} \quad (12)$$

Although these relationships are familiar as forms of the first and second laws of thermodynamics, respectively, for the case when $R=1$ and E_1 is energy, note that they don't yet contain any physics. The familiar physics applies only if we postulate that E_r represents a conserved quantity in transfers between systems, and that δS_{irr} can only increase in time.

Partials, and symmetry between ensembles

Since different thermodynamic "ensembles" often switch the status of a given extensive variable from constraint E_r to work parameter X_m , the symmetry of the equations with respect to these quantities can be better seen if we define M "work multipliers" J_m , analogous to the R "heat multipliers" λ_r , as averages over all constraints E_r of the rate at which the e_{ri} depend on the work parameters to which they correspond, i.e.

$$J_m = \sum_{r=1}^R \lambda_r \sum_{i=1}^{\Omega} p_i \left(- \frac{de_{ri}}{dX_m} \right)_{X_{s \neq m}, \lambda_r}, \quad \forall m \in \{1, M\} \quad (13)$$

Then we can also write...

$$\sum_{m=1}^M J_m \delta X_m = \frac{\delta S}{k} - \sum_{r=1}^R \lambda_r \delta E_r = \frac{\delta S_{irr}}{k} + \sum_{r=1}^R \lambda_r \delta W_r = \delta \text{Log} [Z] + \sum_{r=1}^R E_r \delta \lambda_r. \quad (14)$$

Equivalently, therefore, we gain the following increment (hence partial derivative) relationships for entropy and availability...

$$\begin{aligned} \frac{\delta S}{k} &= \sum_{m=1}^M J_m \delta X_m + \sum_{r=1}^R \lambda_r \delta E_r \\ \implies \left\{ \begin{aligned} J_m &= \left(\frac{\partial S/k}{\partial X_m} \right)_{X_{s \neq m}, E_r}, \\ \lambda_r &= \left(\frac{\partial S/k}{\partial E_r} \right)_{E_{s \neq r}, X_m} \end{aligned} \right\}. \end{aligned} \quad (15)$$

$$\begin{aligned} \delta \text{Log} [Z] &= \sum_{m=1}^M J_m \delta X_m - \sum_{r=1}^R E_r \delta \lambda_r \\ \implies \left\{ \begin{aligned} J_m &= \left(\frac{\partial \text{Log} [Z]}{\partial X_m} \right)_{X_{s \neq m}, \lambda_r} = - \left(\frac{\partial A}{\partial X_m} \right)_{X_{s \neq m}, \lambda_r}, \\ E_r &= - \left(\frac{\partial \text{Log} [Z]}{\partial \lambda_r} \right)_{\lambda_{s \neq r}, X_m} = \left(\frac{\partial A}{\partial \lambda_r} \right)_{\lambda_{s \neq r}, X_m} \end{aligned} \right\} \end{aligned} \quad (16)$$

These partial derivative relationships can be quite useful. For example, if E_1 is energy E then uncertainty slope $\delta S/\delta E$ at constant V and N is $\lambda_1=1/kT$. Also note that for the special case when S , E_r and X_m are extensive variables (i.e. size proportional) that the first of these relationships (cf. page 46 of Betts and Turner) yields the Gibbs-Duhem relation:

$$\frac{S}{k} = \sum_{m=1}^M J_m X_m + \sum_{r=1}^R \lambda_r E_r. \quad (17)$$

In this case, given S as a function of $M+R$ variables, one might then conjecture this larger collection of partial derivative relationships...

$$E_r = \left(\frac{\partial S/k}{\partial \lambda_r} \right)_c = \left(\frac{\partial S/k}{\partial \lambda_r} \right)_Y = \left(\frac{\partial A}{\partial \lambda_r} \right)_c = \left(\frac{\partial A}{\partial \lambda_r} \right)_Y \quad (18)$$

$$\lambda_r = \left(\frac{\partial S/k}{\partial E_r} \right)_\infty = \left(\frac{\partial S/k}{\partial E_r} \right)_\supset = - \left(\frac{\partial A}{\partial E_r} \right)_\infty = - \left(\frac{\partial A}{\partial E_r} \right)_\supset \quad (19)$$

$$X_m = \left(\frac{\partial S/k}{\partial J_m} \right)_Y = \left(\frac{\partial S/k}{\partial J_m} \right)_\supset = \left(\frac{\partial A}{\partial J_m} \right)_Y = \left(\frac{\partial A}{\partial J_m} \right)_\supset \quad (20)$$

$$J_m = \left(\frac{\partial S/k}{\partial X_m} \right)_\infty = \left(\frac{\partial S/k}{\partial X_m} \right)_c = - \left(\frac{\partial A}{\partial X_m} \right)_\infty = - \left(\frac{\partial A}{\partial X_m} \right)_c \quad (21)$$

Here we've adopted a short hand notation for constraints, where $c = \lambda X$ (the default) refers to "ensemble constraints" (control parameters held constant), $\infty = EX$ refers to "no-work constraints" (extensive variables held constant), $\supset = EJ$ means that dependent variables only are constant, and $Y = \lambda J$ refers to "multiplier constraints" (intensive variables held constant), where here letter pairs denote which variable families from the set $\{E_r, \lambda_r, X_m, J_m\}$ are involved.

For irreversible changes, we can also write...

$$\frac{\delta S_{irr}}{k} = \sum_{m=1}^M J_m \delta X_m - \sum_{r=1}^R \lambda_r \delta W_r \quad (22)$$

Fluctuation Constraints

The solution above also gives rise to some very powerful and general relationships between the size of fluctuations in constrained quantities (like energy E) and their rate of change with respect to the Lagrange multipliers (e.g. the specific heat, which of course is simply related to the rate at which uncertainty slope $1/kT$ changes with E).

First, note that $\frac{dp_i}{d\lambda_r} = p_i(E_r - e_{ri}) \forall i \in \{1, \Omega\}$. Using this, one can show that...

$$\begin{aligned} \delta p_i &= \sum_{r=1}^R \delta \lambda_r \frac{dp_i}{d\lambda_r} + \sum_{m=1}^M \delta X_m \frac{dp_i}{dX_m} \\ &= p_i \sum_{r=1}^R \delta \lambda_r (E_r - e_{ri}) + \sum_{m=1}^M \delta X_m \frac{dp_i}{dX_m}. \end{aligned} \quad (23)$$

From this it follows quite generally that the cross variance between parameters r and s is minus the partial derivative of E_r with respect to λ_s , where the partial is taken for all $\lambda_{u \neq s}$ constant, i.e.

$$\begin{aligned} \sigma^2_{E_r E_s} &= \langle e_r e_s \rangle - \langle e_r \rangle \langle e_s \rangle \\ &= - \left(\frac{\partial E_r}{\partial \lambda_s} \right)_{\lambda_{u \neq s}, X_m} = - \left(\frac{\partial E_s}{\partial \lambda_r} \right)_{\lambda_{u \neq r}, X_m} \\ &= - \left(\frac{\partial^2 A}{\partial \lambda_r \partial \lambda_s} \right)_{\lambda_{u \neq r, s}, X_m}, \quad \forall r, s \in \{1, R\}. \end{aligned} \quad (24)$$

The latter equality gives rise to the Onsager reciprocity relations of non-equilibrium thermodynamics. For the special case when

$r=s$, the above expression also fixes the variance (standard deviation squared) of r as

$$\sigma_{E_r}^2 = - \left(\frac{\partial E_r}{\partial \lambda_r} \right)_{\lambda_{s \neq r}, X_m} = \frac{C_{E_r} / k}{\lambda_r^2}, \quad \forall r \in \{1, R\}. \quad (25)$$

The C_{E_r} on the right hand side of this equation is the "differential E_r -capacity" (e.g. the differential heat capacity when E_r is energy) under ensemble constraints (i.e. for constant $\lambda_{s \neq r}$ and X_m). Since the left hand side of this equation seems likely to be positive, the equation says that, for example, temperature is likely to increase with increasing energy. This turns out to be true even for systems like spin systems which exhibit negative absolute temperatures, provided we recognize that negative absolute temperatures are in fact higher than positive absolute temperatures i.e. that the relative size of temperatures must be determined from their reciprocal ($1/kT$) ordering. Conversely, it says that when heat capacity is singular (e.g. during a first order phase change), the fluctuation spectrum will experience a spike as well.

The above relation also prompts a closer look at the "ensemble multiplicity exponent for E_r "...

$$E_r \left(\frac{\partial S/k}{\partial E_r} \right)_{\lambda_{s \neq r}, X_m} \quad (26)$$

...and the "ensemble multiplicity exponent for λ_r "...

$$\begin{aligned} -\lambda_r \left(\frac{\partial S/k}{\partial \lambda_r} \right)_{\lambda_{s \neq r}, X_m} \\ &= -\lambda_r \left(\frac{\partial S/k}{\partial E_r} \right)_{\lambda_{s \neq r}, X_m} \left(\frac{\partial E_r}{\partial \lambda_r} \right)_{\lambda_{s \neq r}, X_m} \\ &= \frac{\xi_{E_r}}{\lambda_r} \frac{C_{E_r}}{E_r} \frac{C_{E_r}}{k} = \frac{\left(\frac{\partial S/k}{\partial E_r} \right)_{\lambda_{s \neq r}, X_m} C_{E_r}}{\left(\frac{\partial S/k}{\partial E_r} \right)_{E_{s \neq r}, X_m} k}. \end{aligned} \quad (27)$$

One can see here that mixed constraints are involved, making the relationships messy at best. The solution, discussed below, involves defining "no-work multiplicity exponents" instead.

Perhaps a similar relationship for fluctuations (not proven here) also exists for the work multipliers J_m , i.e. ...

$$\begin{aligned} \sigma_{J_m J_n}^2 &= - \left(\frac{\partial J_n}{\partial X_m} \right)_{X_{s \neq m}, \lambda_r} = - \left(\frac{\partial J_m}{\partial X_n} \right)_{X_{s \neq n}, \lambda_r} \\ &= \left(\frac{\partial^2 A}{\partial X_m \partial X_n} \right)_{X_{s \neq m, n}, \lambda_r}, \quad \forall m, n \in \{1, M\}. \end{aligned} \quad (28)$$

... and for other control variable ($J_m \lambda_r$) combinations...

$$\begin{aligned} \sigma_{J_m \lambda_r}^2 &= - \left(\frac{\partial J_m}{\partial \lambda_r} \right)_{X_m, \lambda_{s \neq r}} = - \left(\frac{\partial E_r}{\partial X_m} \right)_{X_{s \neq m}, \lambda_r} \\ &= \left(\frac{\partial^2 A}{\partial X_m \partial \lambda_r} \right)_{X_{s \neq m}, \lambda_{s \neq r}}, \quad \forall m \in \{1, M\}, \quad r \in \{1, R\}. \end{aligned} \quad (29)$$

Partition Function, Availability Slope, and Multiplicity Exponent Insights

The utility of knowing the partition function Z , or equivalently the dimensionless availability $A = -\text{Log}[Z]$, emerges from these equations as well. This quantity has long been discussed in the literature as a generalization of free energy. I'm guessing that we'll be able to show that it determines the direction that conserved extensive quantities are likely to flow when random exchange is allowed under ensemble constraints, although that remains to be done. It is easy to see that it reduces to $\frac{S}{k}$ in the microcanonical case, when all control variables are extensive quantities. Although A is here cast in natural units, one can easily define $A = A'/k$, the latter of which may take any units that entropy takes. Many macroscopic quantities can easily be calculated in terms of it. For example...

$$\frac{S}{k} = -A + \sum_{r=1}^R \lambda_r \left(\frac{\partial A}{\partial \lambda_r} \right)_{\lambda_{s \neq r}, X_m}$$

$$E_r = \left(\frac{\partial A}{\partial \lambda_r} \right)_{\lambda_{s \neq r}, X_m}, \quad \forall r \in \{1, R\} \quad (31)$$

$$J_m = - \left(\frac{\partial A}{\partial X_m} \right)_{X_{s \neq m}, \lambda_r}, \quad \forall m \in \{1, M\} \quad (32)$$

$$\sigma_{E_r}^2 = \sqrt{- \left(\frac{\partial^2 A}{\partial \lambda_r^2} \right)_{\lambda_{s \neq r}, X_m}}, \quad \forall r \in \{1, R\} \quad (33)$$

$$\sigma_{E_r E_s}^2 = - \left(\frac{\partial^2 A}{\partial \lambda_r \partial \lambda_s} \right)_{\lambda_{u \neq r, s}, X_m}, \quad \forall r, s \in \{1, R\}. \quad (34)$$

Should the additional fluctuation relations above apply, we would also have (for example)...

$$\sigma_{J_m}^2 = \sqrt{- \left(\frac{\partial^2 A}{\partial X_m^2} \right)_{X_{s \neq m}, \lambda_r}}, \quad \forall m \in \{1, M\} \quad (35)$$

Aprupo our work on natural units for heat capacities, note also quite generally that...

$$\lambda_r E_r = \lambda_r \left(\frac{\partial A}{\partial \lambda_r} \right)_{\lambda_{s \neq r}, X_m} = E_r \left(\frac{\partial S/k}{\partial E_r} \right)_{E_{s \neq r}, X_m} \equiv \xi_{E_r}, \quad \forall r \in \{1, R\} \quad (36)$$

$$J_m X_m = -X_m \left(\frac{\partial A}{\partial X_m} \right)_{X_{s \neq m}, \lambda_r} = X_m \left(\frac{\partial S/k}{\partial X_m} \right)_{X_{s \neq m}, E_r} \equiv \xi_{X_m}, \quad \forall m \in \{1, M\} \quad (37)$$

$$\frac{C_{E_r}^{\otimes}}{k} \equiv -\lambda_r^2 \left(\frac{\partial E}{\partial \lambda_r} \right)_{E_{s \neq r}, X_m} = -\lambda_r \left(\frac{\partial S/k}{\partial \lambda_r} \right)_{E_{s \neq r}, X_m} \equiv \xi_{\lambda_r}, \quad \forall r \in \{1, R\} \quad (38)$$

$$\frac{C_{E_r}}{k} \equiv -\lambda_r^2 \left(\frac{\partial E}{\partial \lambda_r} \right)_{\lambda_{s \neq r}, X_m} = -\lambda_r^2 \left(\frac{\partial^2 A}{\partial \lambda_r^2} \right)_{\lambda_{s \neq r}, X_m} = \lambda_r^2 \sigma_{E_r}^2, \quad \forall r \in \{1, R\} \quad (39)$$

and perhaps the conjectured relation below, will follow as well...

$$\frac{C_{X_m}^{\otimes}}{k} \equiv -J_m^2 \left(\frac{\partial E}{\partial J_m} \right)_{E_r, X_{s \neq m}} = -J_m \left(\frac{\partial S/k}{\partial J_m} \right)_{E_r, X_{s \neq m}} \equiv \xi_{J_m}, \quad \forall m \in \{1, M\} \quad (40)$$

All of the quantities which define a ξ ("multiplicity exponent") parameter represent the number of base-b units of entropy increase per b-fold increase in the constrained quantity E_r or it's Lagrange multiplier λ_r , under no-work conditions. These quantities can also be thought of as "integral E_r -capacities" $\lambda_r E_r$ and "differential E_r capacities" $-\lambda_r^2 \left(\frac{\partial E}{\partial \lambda_r} \right)_{E_{s \neq r}, X_m}$, respectively. Note that the former also provide information on base-b availability changes under ensemble constraints, per b-fold change in a given multiplier. We argue elsewhere that these relationships provide insight into both natural units for heat capacity, and its utility in non-quadratic systems (e.g. those for which $1/kT$ makes more sense than kT).

Except for the entropy partials above which operate under no-work conditions (sometimes noted with \otimes), most if not all of the above partials operate under "ensemble conditions", i.e. with respect to X_m hold constant X_s for $s \neq m$, and λ_r for all r , while those with respect to λ_r hold constant λ_s for $s \neq r$, and X_m for all m . Thus in availability relations as in the chosen ensemble thought experiment itself, the λ_r and X_m serve as the "control variables". For example, microcanonical ensembles and availability relations operate with external control of all work parameters. Canonical ensemble systems and availability relations operate in a temperature-controlled heat bath but with other work parameters fixed. Pressure ensemble systems (and availability relations) also have their pressure rather than volume held constant, etc.

Thermodynamic examples for the monatomic ideal gas

Ensemble	General	MicroCanonical	Canonical	Pressure	Grand
System Model	E_r sharing	isolated	U sharing	U, V sharing	U, N sharing
Control Variables	$\lambda_r X_m$	UVN	βVN	$\beta \gamma N$	$\beta \alpha V$
Dependent Variables	$E_r J_m$	$\beta \gamma \alpha$	U $\gamma \alpha$	UV α	UN γ
Partition Fn Z	$e^{-\lambda} = \sum_{i=1}^{\infty} e^{-\lambda_r E_r}$	$(\frac{V e}{N})^N (\frac{4 \pi e m U}{3 N h^2})^{\frac{3N}{2}}$	$(\frac{V e}{N})^N (\frac{2 \pi m}{\beta h^2})^{\frac{3N}{2}}$	$(\frac{1}{\gamma})^N (\frac{2 \pi m}{\beta h^2})^{\frac{3N}{2}}$	$e^{-\alpha} (\frac{2 \pi m}{h^2 \beta})^{\frac{3N}{2}}$
Availability A [nats]	$\sum_{r=1}^R \lambda_r E_r - \frac{S}{k} = -\text{Log}[Z]$	$-\frac{S}{k} = -N(\langle \alpha \rangle + \frac{5}{2})$	$\beta U - \frac{S}{k} = -N(\langle \alpha \rangle + 1)$	$\beta U + \gamma V - \frac{S}{k} = -N\langle \alpha \rangle$	$\beta U + \alpha N - \frac{S}{k} = -\langle N \rangle$
Energy $\langle U \rangle$	$\frac{\partial A}{\partial \beta}$	Fixed	$\frac{3N}{2\beta}$	$\frac{3N}{2\beta}$	$\frac{3\langle N \rangle}{2\beta}$
Coldness $\langle \beta \equiv \frac{1}{kT} \rangle$	$-\frac{\partial A}{\partial U}$	$\frac{3N}{2U}$	Fixed	Fixed	Fixed
Volume $\langle V \rangle$	$\frac{\partial A}{\partial \gamma}$	Fixed	Fixed	$\frac{N}{\gamma}$	Fixed
FreeExpCoeff $\langle \gamma \equiv \frac{p}{kT} \rangle$	$-\frac{\partial A}{\partial V}$	$\frac{N}{V}$	$\frac{N}{V}$	Fixed	$\frac{\langle N \rangle}{V}$
# Particles $\langle N \rangle$	$\frac{\partial A}{\partial \alpha}$	Fixed	Fixed	Fixed	$V e^{-\alpha} (\frac{2 \pi m}{\beta h^2})^{\frac{3N}{2}}$
ChemAffinity $\langle \alpha \equiv \frac{-\mu}{kT} \rangle$	$-\frac{\partial A}{\partial N}$	$\text{Log}[\frac{V}{N} (\frac{4 \pi m U}{3 h^2 N})^{\frac{3}{2}}]$	$\text{Log}[\frac{V}{N} (\frac{2 \pi m}{\beta h^2})^{\frac{3}{2}}]$	$\text{Log}[\frac{1}{\gamma} (\frac{2 \pi m}{\beta h^2})^{\frac{3}{2}}]$	Fixed
Entropy S/k [nats]	$\text{Log}[Z] + \sum_{r=1}^R \lambda_r E_r$	$N(\langle \alpha \rangle + \frac{5}{2})$	$N(\langle \alpha \rangle + \frac{5}{2})$	$N(\langle \alpha \rangle + \frac{5}{2})$	$\langle N \rangle (\alpha + \frac{5}{2})$
$\xi_U \equiv U (\frac{\partial S}{\partial U})_{\otimes} = \frac{U}{kT}$	$\beta U = \beta (\frac{\partial A}{\partial \beta})$	$\frac{3N}{2}$	$\frac{3N}{2}$	$\frac{3N}{2}$	$\frac{3\langle N \rangle}{2}$
$\xi_{\beta} \equiv -\beta (\frac{\partial S}{\partial \beta})_{\otimes} = \frac{C_{Vib}}{k}$	$-\beta^2 (\frac{\partial U}{\partial \beta})_{\otimes} = (\frac{\partial U}{\partial (kT)})_{\otimes}$	$\frac{3N}{2}$	$\frac{3N}{2}$	$\frac{3N}{2}$	$\frac{3\langle N \rangle}{2}$
$\xi_V \equiv V (\frac{\partial S}{\partial V})_{\otimes} = \frac{PV}{kT}$	$\gamma V = \gamma (\frac{\partial A}{\partial \gamma})$	N	N	N	$\langle N \rangle$
$\xi_{\gamma} \equiv -\gamma (\frac{\partial S}{\partial \gamma})_{\otimes} = \frac{C_{Vib}}{k}$	$-\gamma^2 (\frac{\partial V}{\partial \gamma})_{\otimes} = (\frac{\partial V}{\partial (kT/P)})_{\otimes}$	N	N	N	$\langle N \rangle$
$\xi_N \equiv N (\frac{\partial S}{\partial N})_{\otimes} = \frac{-\mu N}{kT}$	$\alpha N = \alpha (\frac{\partial A}{\partial \alpha})$	$N\langle \alpha \rangle$	$N\langle \alpha \rangle$	$N\langle \alpha \rangle$	$\langle N \rangle \alpha$
$\xi_{\alpha} \equiv -\alpha (\frac{\partial S}{\partial \alpha})_{\otimes} = \frac{C_{Vib}}{k}$	$-\alpha^2 (\frac{\partial N}{\partial \alpha})_{\otimes} = (\frac{-\partial N}{\partial (kT/\mu)})_{\otimes}$	$-N\langle \alpha \rangle$	$-N\langle \alpha \rangle$	$-N\langle \alpha \rangle$	$-\langle N \rangle \alpha$
Variance in U	$\sigma_U^2 = -\frac{\partial U}{\partial \beta} = -\frac{\partial^2 A}{\partial \beta^2}$	-	$\frac{3N}{2\beta^2}$	$\frac{3N}{2\beta^2}$	$\frac{3\langle N \rangle}{2\beta^2} (\alpha + \frac{5}{2})$
Variance in β	$\sigma_{\beta}^2 = -\frac{\partial \beta}{\partial U} = \frac{\partial^2 A}{\partial U^2}$	-	-	-	-
Variance in V	$\sigma_V^2 = -\frac{\partial V}{\partial \gamma} = -\frac{\partial^2 A}{\partial \gamma^2}$	-	-	$\frac{N}{\gamma^2}$	-
Variance in γ	$\sigma_{\gamma}^2 = -\frac{\partial \gamma}{\partial V} = \frac{\partial^2 A}{\partial V^2}$	$\frac{N}{V^2}$	$\frac{N}{V^2}$	-	0
Variance in N	$\sigma_N^2 = -\frac{\partial N}{\partial \alpha} = -\frac{\partial^2 A}{\partial \alpha^2}$	-	-	-	$\langle N \rangle$
Variance in α	$\sigma_{\alpha}^2 = -\frac{\partial \alpha}{\partial N} = \frac{\partial^2 A}{\partial N^2}$	$\frac{5N}{2N^2}$	$\frac{N}{N^2}$	0	-
CoVariance: V, N	$\sigma_{VN}^2 = -\frac{\partial V}{\partial \alpha} = -\frac{\partial N}{\partial \gamma} = -\frac{\partial^2 A}{\partial \alpha \partial \gamma}$	-	-	-	-
CoVariance: γ, α	$\sigma_{\gamma\alpha}^2 = -\frac{\partial \gamma}{\partial \alpha} = -\frac{\partial \alpha}{\partial \gamma} = \frac{\partial^2 A}{\partial N \partial V}$	$-\frac{1}{V}$	$-\frac{1}{V}$	-	-
CoVariance: N, U	$\sigma_{NU}^2 = -\frac{\partial U}{\partial \alpha} = -\frac{\partial V}{\partial \beta} = -\frac{\partial^2 A}{\partial \alpha \partial \beta}$	-	-	-	$-\frac{3\langle N \rangle}{2}$
CoVariance: α, β	$\sigma_{\alpha\beta}^2 = -\frac{\partial \alpha}{\partial \beta} = -\frac{\partial \beta}{\partial \alpha} = \frac{\partial^2 A}{\partial U \partial N}$	$-\frac{3}{2U}$	-	-	-
CoVariance: U, V	$\sigma_{UV}^2 = -\frac{\partial U}{\partial \gamma} = -\frac{\partial V}{\partial \beta} = -\frac{\partial^2 A}{\partial \gamma \partial \beta}$	-	-	0	-
CoVariance: β, γ	$\sigma_{\beta\gamma}^2 = -\frac{\partial \beta}{\partial \gamma} = -\frac{\partial \gamma}{\partial \beta} = \frac{\partial^2 A}{\partial V \partial U}$	0	-	-	-
CoVariance: α, U	$\sigma_{N\beta}^2 = \frac{\partial \alpha}{\partial \beta} = \frac{\partial U}{\partial N} = \frac{\partial^2 A}{\partial N \partial \beta}$	-	$\frac{3}{2\beta}$	$\frac{3}{2\beta}$	-
CoVariance: α, V	$\sigma_{N\gamma}^2 = \frac{\partial \alpha}{\partial \gamma} = \frac{\partial V}{\partial N} = \frac{\partial^2 A}{\partial N \partial \gamma}$	-	-	$\frac{1}{\gamma}$	-
CoVariance: γ, U	$\sigma_{V\beta}^2 = \frac{\partial \gamma}{\partial \beta} = \frac{\partial U}{\partial V} = \frac{\partial^2 A}{\partial V \partial \beta}$	-	0	-	$\frac{3\langle \gamma \rangle}{2\beta}$
CoVariance: γ, N	$\sigma_{V\alpha}^2 = \frac{\partial \gamma}{\partial \alpha} = \frac{\partial N}{\partial V} = \frac{\partial^2 A}{\partial V \partial \alpha}$	-	-	-	$\langle \gamma \rangle$

Kullback-Leibler divergence or "net-surprisal"

KL divergence of reference probability set $\{p_o\}$ from a system state $\{p\}$ is defined as the "net-surprisal"...

$$I_{\text{net}} \equiv k \sum_{i=1}^{\Omega} p_i \text{Log} \left[\frac{p_i}{p_{oi}} \right] \geq 0 \text{ by Gibbs Inequality.} \quad (41)$$

From equations (4), (5) and (6) above, one can rewrite this (arXiv:physics/9611022) as ...

$$\frac{I_{\text{net}}}{k} = \sum_{r=1}^R \lambda_{ro} (E_r - E_{ro}) - \left(\frac{S}{k} - \frac{S_o}{k} \right) \geq 0. \quad (42)$$

Available Work: KL divergence of "ambient from actual"

Let's treat the reference state as that of an ambient reservoir, the system state as that of an unequilibrated subsystem with access to the reservoir, and the observables E_r as quantities that are conserved on transfer between subsystem and ambient. The subsystem's positive KL divergence or net-surprisal then becomes the entropy gained on transfer of excess subsystem observables E_r to ambient, minus the entropy lost by the subsystem in the process (which is generally less if the conditions encourage spontaneous flow). Done reversibly, this entropy increase could make possible an entropy decrease somewhere else. Hence this measure of *departure from equilibrium* is the subsystem's "availability in entropy units", as mentioned by J. W. Gibbs. Multiplied by ambient temperature (i.e. the reciprocal of energy's Lagrange multiplier), it yields the subsystem's *available work* or (in engineering terms) the *exergy difference* between subsystem and ambient.

Mutual Information: KL divergence of "uncorrelated from correlated"

Now we turn to the question of correlated systems, i.e. those for which subsystem correlations makes entropy a non-local and non-extensive quantity. This is easy to see since given two subsystems with state indices i and j , one writes the mutual information from equation (41) as...

$$M_{12} \equiv k \sum_{i=1}^{\Omega_1} \sum_{j=1}^{\Omega_2} p_{ij} \text{Log} \left[\frac{p_{ij}}{p_i p_j} \right] = S_1 + S_2 - S_{12} \geq 0, \quad (43)$$

where the p_i and p_j are marginal probabilities defined e.g. by $p_i \equiv \sum_{j=1}^{\Omega_2} p_{ij}$. From this expression it's easy to show that the entropy of the whole system $S_{12} = S_1 + S_2 - M_{12}$, where M_{12} resides in neither system but instead in the correlation between the two. In *communications theory*, *clade analysis*, and *quantum computing* the KL divergence of uncorrelated from correlated, in this sense, can be used to measure the mutual information associated with *fidelity*, *inheritance*, and *entanglement* respectively. In studies of evolving correlation-based complexity (a kind of natural history of invention), the mutual information of subsystem correlations is part of a larger story as well.

Akaike Information Criterion: KL divergence of "model from reality"

In ecology and related fields, KL-divergence of "model from reality" is useful even if the only clues we have about reality are some experimental measurements, since it tells how to rank models against experimental data according to the residuals that they *don't* account for. Specifically Akaike's criterion allows one to estimate the KL-divergence of a model from reality, to within a constant additive term, by a function (like the squares summed) of the deviations observed between data and the model's predictions. Estimates of such divergence, for models that share the same additive term, can in turn be used to choose between models.

Layered mutuality: KL divergences in-out wrt/boundaries in a multiscale network

The foregoing are applications which by and large work on one level of organization at a time. Chaisson's cosmic evolution, on the other hand, can be seen as developments *after neutral atom formation* of sub-system correlations (i.e. mutual information) looking inward and outward with respect to a layered series of physical boundaries. The formation of these subsystem correlations for the most part has been powered by the reversible thermalization of energy in the form of available work (cf. *Complexity* 13:18-26) . This yields an integrative (cross-disciplinary) view of these correlations, a fringe benefit of which might be an incentive for communications informed simultaneously to correlations on more than one level (cf. arXiv:physics/0603068).

Contact and Copyright Information

This note represents insights provided by numerous colleagues, and has benefited in particular from discussions with, and notes provided by, the late E. T. Jaynes. The person responsible for mistakes (and to whom you can forward suggestions) is P. Fraundorf in Physics and Astronomy at UM-St. Louis (pfraundorf AT umsl.edu).

Notes