# The Effect of Noise on Entropy

By Rosemarie Swanson and Stanley M. Swanson

*Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843-2128, USA*

(*Received* 11 *June* 1992; *accepted* 24 *September* 1992)

## Abstract

A complementary relationship between the entropy $(S)$ and the variance $(\sigma^2)$ of an electron-density map is derived by approximating the logarithmic term in the entropy expression by a series expansion around the average map density. The resulting expression is $S \simeq \ln N - \frac{1}{2}\sigma^2$, where $N$ is the number of grid points and $\sigma$ is the r.m.s. deviation from the mean in a map normalized to unit mean. The algebraic expression is of interest because it is consistent with and allows numerical evaluation of the surprising argument that noise *decreases* the entropy of a map. The argument is that a noise contribution by itself generates a certain variance that is independent of the atomic structure and that adds to the variance due to the structure. Increased variance corresponds to decreased entropy. This property of noise provides an intuitively reasonable justification for maximizing the entropy of an electron-density map in the quest for more readily interpretable maps of macromolecules. The entropy–variance relationship also extends the range of applicability of the entropy concept to maps with a limited amount of negative density. The approximation which leads to the entropy–variance relationship is most applicable where it is most likely to be useful – in experimental maps of relatively low structure definition.

## Introduction

The maximum-entropy approach to the phase problem relies on the notion that the optimum representation of the information in the data corresponds to a probability distribution with maximum entropy (Jaynes, 1978). In crystallography, this probability distribution is a normalized electron-density map which is consistent with a set of constraints (the experimental data) and does not contain unwarranted detail. Such a distribution is the flattest possible under the constraints (Collins & Prince, 1991). But why is the flattest map desirable, especially when one considers ideal atoms as points?

An initial answer is that in practical macromolecular problems atoms are far from being points, especially in the initial phasing stages. In fact, the main expression of phasing inadequacies in macromolecular maps is ambiguities in chain tracing – problems in working out the connectivity of atoms. At less than atomic resolution, spikes misrepresent the true density. These make map interpretation more, rather than less, difficult. A flat map is more helpful than a spiky map in establishing connectivity.

A second answer is the argument that, again at practical resolution, noisy maps are less flat than their noiseless counterparts. The qualitative and quantitative presentation of this second point and some numerical tests of its applicability at practical resolution are the focus of this paper. (The quantitative assessment of connectivity is a much more difficult problem that will be addressed separately.)

## Noise decreases entropy

Noise, it can be argued, increases the variance and decreases the entropy of an electron-density map, as compared to the same map without noise.

Consider a 'noiseless pseudo-density' map for a protein molecule. Such a map, on a grid at limited resolution, with no negative density, can be synthesized by summing Gaussian atomic density peaks with widths obtained from resolution, temperature factors and scattering factors. From this map, 'noiseless' structure factors may be obtained by an inverse Fourier transform.

Adding noise to the $F$'s will increase the variance of the resulting map. The difference between the noisy and noiseless maps is the same as a map calculated from the noise terms alone, because of the linearity of the Fourier transform. The noise-only map has a certain variance, as has the noiseless map. If the noise is assumed to be uncorrelated with the $F$'s, the variance of the combined map (the noisy $F$'s map) is the sum of the variances of the component maps and is greater than that of either map alone. The variance of the noiseless map results from the protein structure, and is the minimum possible variance, given that structural constraint.

The relation between the variance of a map and its entropy is a complementary one: a larger variance corresponds to a reduced entropy. In the limit of small variance, the quantitative relationship between entropy and variance is

$$S = \ln N - 0.5 \frac{\sum_{i=1}^{N} (d_i - \bar{d})^2}{N\bar{d}^2},$$

where $N$ is the number of grid points in the map, $d_i$ is the density value at a point and $\bar{d}$ is the average density value. This relationship is derived from a series expansion of the logarithm, but is numerically good to about 10%

Table 1. *Numerical comparison of entropy and variance*

| Map identity | Resolution | $\Delta S^+$ | $\frac{1}{2}\sigma^2$ | $S^+$ | $N$ | $\% \geq 0$ | % optiony | $d_{\max}/\bar{d}$ | $d_{\min}/\bar{d}$ |
|---|---|---|---|---|---|---|---|---|---|
| CPG MIR | 2.0 | 0.0924 | 0.0892 | 12.2051 | 220320 | 99.4 | 90.7 | 4.02 | −0.48 |
| No $F_{000}$ | | 0.3449 | 0.3606 | 11.2317 | | 48.4 | 34.3 | 3.02 | −1.48 |
| $(-d)$ | | n/a | 0.5933 | 11.7095 | | 99.4 | 55.3 | 4.02 | −0.48 |
| | | | | | | | | | |
| 3EST | 3.0 | 0.2217 | 0.2383 | 12.1666 | 277200 | 84.6 | 69.4 | 7.26 | −1.55 |
| No $F_{000}$ | | 0.4454 | 0.5125 | 11.3366 | | 47.2 | 30.2 | 6.26 | −2.55 |
| $(-d)$ | | n/a | 0.5000 | 12.0325 | | 84.6 | 60.7 | 7.26 | −1.55 |
| | | | | | | | | | |
| XSU | 3.0 | 0.4016 | 0.3336 | 11.0731 | 96255 | 100.0 | 66.9 | 3.07 | 0.00 |
| No $F_{000}$ | | 0.2214 | 0.2022 | 10.5244 | | 48.2 | 38.7 | 2.07 | −1.00 |
| | | | | | | | | | |
| 1BP2 $F_c$ | 1.7 | 0.7395 | 1.0273 | 11.2004 | 230400 | 66.4 | 31.8 | 28.3 | −3.96 |
| med | | 0.3748 | 2.6995 | 11.9270 | | 100.0 | 65.7 | 312.6 | 0.27 |
| $F_o$ med | | 0.4590 | 0.8788 | 11.5287 | | 69.6 | 44.1 | 143.2 | −12.23 |
| sms1 | | 0.0752 | 0.1275 | 12.7219 | | 100.0 | 92.7 | 28.3 | 0.51 |
| $F_o$ sms1 | | 0.6704 | 1.0031 | 11.2705 | | 66.5 | 34.1 | 59.2 | −5.24 |
| sms2 | | 0.3470 | 0.7697 | 11.9998 | | 100.0 | 70.4 | 28.3 | 0.23 |
| $F_o$ sms2 | | 0.6769 | 1.0183 | 11.2662 | | 66.7 | 33.9 | 38.5 | −4.53 |
| | | | | | | | | | |
| Bi-level | n/a | 0.6931 | 0.5000 | n/a | n/a | 100.0 | 50.0 | 2.00 | 0.00 |
| | | 0.6617 | 0.4900 | | | | 51.6 | 1.99 | 0.01 |
| | | 1.3863 | 1.5000 | | | | 25.0 | 4.00 | 0.00 |
| | | 1.3339 | 1.4700 | | | | 26.3 | 3.97 | 0.01 |
| | | 0.5548 | 0.6734 | | | | 57.4 | 3.01 | 0.33 |
| | | 2.0794 | 3.5000 | | | | 12.5 | 8.00 | 0.00 |
| | | 0.9165 | 1.5712 | | | | 40.0 | 5.67 | 0.33 |

Notes: $\Delta S^+ = \ln(N^+) - S^+$ is the logarithm of the number of points $\geq 0$ minus the entropy over the same set of points. $N^+$ is the number of points with density $\geq 0$. $\bar{d}$ is the average density when $F_{000}$ is included. 'n/a' stands for 'not applicable'.

for experimental electron-density maps. Taking this as an alternative approximate entropy has an advantage over the logarithmic definition because it can be applied to maps with a limited amount of negative density (although the mean $\bar{d}$ cannot approach zero). A key point in the derivation is to expand the logarithm around the average density (see *Appendix*).

In the region where our approximation is good, the sum of the entropy and one-half of the always-positive variance of the map is $\ln N$, the maximum possible entropy for $N$ points. Any increase of the variance due to noise must necessarily decrease the entropy, and this is a justification for using entropy as a figure of merit for choosing among phase sets. The approximation seems good enough for 'real' maps to serve as a heuristic characterization of what the entropy measure is telling us. The comparison of entropies in this approximation reduces to comparing suitably normalized variances.

## Numerical examples

Numerical tests of the entropy–variance relationship were applied to maps with different degrees of refinement, with different resolutions and with different amounts of negative density. The relationship holds best for 'real' electron-density maps (low to moderate atom definition) and is worst for maps which have been made exponentially positive. The relationship has been examined quantitatively in these cases:

(1) CPG – a 2.0 Å smoothed multiple isomorphous replacement (MIR) map of carboxypeptidase Aγ (R. Swanson, unpublished);

(2) 3EST – a 3.0 Å $2F_o - F_c$ map from refined native elastase (Meyer, Cole, Radhakrishnan & Epp, 1988);

(3) XSU – a corresponding 3.0 Å 'noiseless pseudo-density' map also based on elastase coordinates;

(4) 1BP2 – a 1.7 Å $F_c$ map of phospholipase A2 (Dijkstra, Kalk, Hol & Drenth, 1981), and various 'positivized' maps created from it;

(5) Bi-level – a set of simplified maps (only two density levels) for illustration and comparison of the effects of certain density changes.

The results are shown in Table 1 in the form of the difference between the maximum entropy and the observed entropy (both based on positive-density points – column $\Delta S^+$), compared with an estimate of the same quantity based on variance (column $\frac{1}{2}\sigma^2$). The observed entropy (column $S^+$) and the total number of points in each map (column $N$) are given for reference. The percentage of non-negative density points (column $\% \geq 0$) and the percentage of points in a uniform density that would yield an entropy equal to the observed entropy (column % optiony) are shown to facilitate comparison between the examples. The quantity 'optiony', on which '% optiony' is based, is the exponential of an entropy value (Swanson, 1990). Percent optiony corresponds to the non-zero volume of the cell in a 1-0 bi-level density with the same entropy as the actual map. The last two columns $(d_{\max}/\bar{d}, d_{\min}/\bar{d})$ in the table show the maximum and minimum densities (scaled by the average density $\bar{d}$) as indicators of the non-uniformity of the density.

Although the inclusion of $F_{000}$ is important to entropy calculations, $F_{000}$ is often not readily available in the earlier stages of protein map calculations. To explore the

consequences of omitting the $F_{000}$ term, the table shows values for CPG, 3EST and XSU maps with and without $F_{000}$ included. The entropy values and variances for the first two lines for each of the maps mentioned are based only on points with density greater than zero. The third line for each $(-d)$ is based on all points, including negative densities. The entropy quoted on the third line of each set is estimated from the variance only since entropy cannot be calculated directly when negative numbers are included. The variance approximation to the entropy is little affected by the omission of $F_{000}$, even though the overall entropy, as calculated by omitting negative density points, is significantly (and unrealistically) reduced.

A rule of thumb for the contribution of $F_{000}$ to the average value of a map can be proposed from experience with MIR maps and a brief theoretical argument: the average value of a map with $F_{000}$ included is probably in the range $0.7$–$1.0\sigma$, where $\sigma$ is calculated using all the points in the map prepared without $F_{000}$. The argument is: (1) our experience shows that roughly 25% of points are negative in an MIR map with the correct value of $F_{000}$ included, (2) the density distribution in an initial map is fairly close to Gaussian, and (3) for a Gaussian distribution of density values, a value of the map mean of $0.7\sigma$ would make density values less than zero account for roughly 25% of the points; a mean equal to $\sigma$ will give approximately 15% negative density points.

The first purpose of the examples is to evaluate the agreement between the straight entropy calculation and its estimate from the map variance ($\Delta S^+$, $\frac{1}{2}\sigma^2$). The values agree to 10% or better for the two experimental maps, CPG and 3EST, and are not too much worse for the 'pseudo-density' XSU map. The agreement is substantially worse for the 1BP2 examples. The maps for which the variance and entropy agreement is poor have an artificial sharpness and lack a bulk solvent model; they fall outside the range where the variance approximation is good. Thus the approximation is best where it is most likely to be useful, namely in real maps, rather than in model maps.

The second purpose of the examples is to examine the effectiveness of the variance approximation to the entropy when negative density is included. On the third lines for CPG and 3EST, this question has been addressed (maps include $F_{000}$ and negative density). Negative density has been included in the variance calculation, and entropy has been estimated on this basis. This estimation yields a percent optiony value that lies between the percent optiony values based on only the positive density for the two cases, $F_{000}$ included or not included ('no $F_{000}$'). This extension of the entropy concept to include negative densities thus seems numerically plausible when compared against entropy estimates based on truncation of negative density.

The third purpose of the examples is to examine numerically the entropy decrease with the addition of noise. This question is addressed in the 1BP2 examples, where a map based on $F_c$'s (from protein coordinates, no bulk solvent

model) is processed in three different ways to make it positive definite. Each of these is taken as a model of a 'perfect' map. Noise is added by recalculating the map with the 'perfect' phases, but using artificially generated $F_o$'s [the $F_o$'s are the $F_c$'s with noise ($|\Delta F|/F \simeq 0.08$) added by the *PHONYD* program in the *XTAL3.0* system (Stewart, 1990)]. In each case (compare the lines 'xxx' with the immediately following lines '$F_o$xxx'), the difference between the map entropy and the maximum entropy is greater for the 'noisy' maps, as shown in the $\Delta S^+$ column of Table 1, and as shown by the decreases in the '% optiony' column for the noisy map as compared with the 'perfect' map. This decrease in entropy is large, although not as large as the accompanying change in variance (which falls well outside the range where our variance approximation to the entropy is valid). The positivizations 'sms1' and 'sms2' preserve the original map maximum rather than sharpening the map as 'med' does; all positivizations preserve the mean of the map. [The 'med' map is transformed by *MEDENS* in *XTAL3.0* (Collins & Stewart, 1990), with the map sharpening parameter, *sh* at its default value of 1, which corresponds to the 'natural' value in the derivation by Collins & Prince (1991).] In all cases, as expected, the addition of noise decreases entropy.

The final entries in the table ('Bi-level') are simplified two-level maps with the correct proportion of high and low densities to have $\bar{d} = 1$. They provide some insight into the range of validity of the approximation. For example, the first line in this group gives values for a map which consists, half and half, of density points with values 2 and 0. The following line shows how these values change for a map that has been slightly 'blurred' with density values of 1.99 and 0.01. The entropy increases (recall that $\Delta S^+$ is the difference between maximum possible entropy and observed entropy) and the variance decreases. The volume of the cell effectively occupied by density slightly increases. The variance approximation is slightly better for the slightly higher entropy flatter map. The next three lines present values for maps with one-quarter high values (4, 3.97, 3.01) and three-quarters low values (0.0, 0.01, 0.33, respectively). Again, the variance approximation is best for the lowest variance case. Note the substantial influence on total entropy and effective volume occupied (% optiony) of spreading low levels of density over the 'solvent' region ($d_{\min} = d_{\max}/9$ for line 5 of 'Bi-level'). Similar behavior is shown for the case of one-eighth high values (8.0, 5.67) and seven-eighths low values (0.0, 0.33 respectively). On the basis of these examples, it appears that the variance approximation to the entropy is reasonably good for variances up to 0.5, relative to a map normalized to $\bar{d} = 1$ (*i.e.* $\sigma < 0.7\bar{d}$).

## Concluding remarks

First, the effect of noise on entropy illustrated here justifies the use of entropy maximization in an intuitively plausible way. Second, the relation between entropy and variance

appears to be a valid approximation in the region where it will be most useful, namely, in maps where atom definition is low. Third, the variance approximation extends the entropy concept to densities with negative values.

## APPENDIX

Let a non-negative density be sampled in $N$ places on a regular grid, with values $d_i$. We divide by $D = \sum_i d_i$ to form the reduced density $r_i = d_i/D$. The reduced density sums to 1, and can be interpreted as a probability in the sense used for the information entropy

$$S = -\sum_{i=1}^{N} r_i \ln r_i.$$

Further, rewrite $r_i$ in terms of a fractional deviation from the mean

$$r_i = \bar{r}(1 + x_i),$$

where $\bar{r} = 1/N$ is the mean reduced density. Thus, by definition, $-1 \leq x_i \leq N - 1$ and $\sum_i x_i = 0$. The entropy is then

$$
\begin{aligned}
S &= -\sum_{i=1}^{N} \bar{r}(1 + x_i) \ln[\bar{r}(1 + x_i)] \\
&= -\bar{r} \sum \ln \bar{r} - \bar{r} \ln \bar{r} \sum x_i \\
&\quad - \bar{r} \sum (1 + x_i) \ln(1 + x_i) \\
&= \ln N - (1/N) \sum_{i=1}^{N} (1 + x_i) \ln(1 + x_i).
\end{aligned}
$$

With the assumption $|x_i| < 1$, we may expand the logarithm, and drop the first-order $\sum x_i$ which is zero, to get

$$\hat{S} = \ln N - (1/N) \sum_{i=1}^{N} (x_i^2/2 - x_i^3/6 + x_i^4/12 + \ldots).$$

The leading term is exactly half the variance about the mean $\bar{r}$, when expressed in the dimensionless variable $x_i$. [Note that $x_i = (d_i - \bar{d})/\bar{d}$; a density normalized in this way would sum to $N$.] Hence to lowest order,

$$\hat{S} = \ln N - \frac{1}{2}\sigma^2 = \ln N - \frac{1}{2N} \text{variation},$$

where the variance or variation would be computed with the density normalized to $\bar{d} = 1$. [Variation is used in the sense of Fisher (1960): the sum over all points of the squared difference from the mean, and is equal to $N$ times the square of the standard deviation for an $N$ point map.] For a symmetric distribution, the cubic term would also disappear, and although real density distributions are not symmetric, we do not expect significant contributions in the region where our approximation is good.

Our result can be seen formally as a special case of a more general result of Jaynes (1979, equation $A4$) relating the decrease of entropy near a constrained maximum to the $\chi^2$ measure of the difference between that maximum point and a nearby point.

In this approximation, the comparison of two entropies with the same mean and the same $N$ is the difference between two quadratic sums

$$\hat{S}_x - \hat{S}_z = \frac{1}{2N} \sum (z_i^2 - x_i^2) = \frac{1}{2N} \sum \varepsilon_i(z_i + x_i),$$

where $\varepsilon_i = z_i - x_i$. By the mean value theorem, the difference between the original entropies is

$$
\begin{aligned}
S_x - S_z &= \frac{1}{N} \sum (z_i - x_i)[\ln(1 + y_i) + 1] \\
&= \frac{1}{N} \sum \varepsilon_i \ln(1 + y_i),
\end{aligned}
$$

where $y_i$ lies strictly between $x_i$ and $z_i$ and the term $\sum(z_i - x_i)$ disappears in the sum. If $|\varepsilon_i| \ll (z_i + x_i)/2$, then $y_i \simeq (z_i + x_i)/2 = \mu_i$ and an expansion of the logarithm gives $\hat{S}_x - \hat{S}_z$ as its leading term when $|y_i| \ll 1$.

Both of these expressions show clearly how $\varepsilon_i$ must be correlated with $\mu_i$ in order to make $S_x > S_z$: to get positive terms in the sum, when the mean (or $y_i$) is $< 0$, $x_i$ must on the average be $> z_i$ to make $\varepsilon_i < 0$, or when the mean is $> 0$, $x_i$ must on the average be $< z_i$. In both cases, $x_i$ is closer to zero than $z_i$ (the corresponding $d_i$ is closer to $\bar{d}$) and the density is flatter, on the average, point by point. See Collins & Prince (1991) for another derivation.

## References

COLLINS, D. M. & PRINCE, E. (1991). *Exponential Density: Exact Fitting of Structure Moduli by Entropy Maximization.* In *Crystallographic Computing 5: From Chemistry to Biology*, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 308–316. Oxford Univ. Press.

COLLINS, D. M. & STEWART, J. M. (1990). *MEDENS.* In *XTAL3.0 Reference Manual*, edited by S. R. HALL & J. M. STEWART, pp. 149–151. Univs. of Western Australia, Australia, and Maryland, USA.

DIJKSTRA, B. W., KALK, K. H., HOL, W. G. J. & DRENTH, J. (1981). *J. Mol. Biol.* **147**, 97–123.

FISHER, R. A. (1960). *The Design of Experiments*, 8th ed., pp. 52–58. Edinburgh: Oliver and Boyd.

JAYNES, E. T. (1978). *Where Do We Stand on Maximum Entropy?* In *Papers on Probability, Statistics, and Statistical Physics*, edited by R. D. ROSENKRANTZ, pp. 211–314. Dordrecht: Reidel.

JAYNES, E. T. (1979). *Concentration of Distributions at Entropy Maxima.* In *Papers on Probability, Statistics, and Statistical Physics*, edited by R. D. ROSENKRANTZ, pp. 317–336. Dordrecht: Reidel.

MEYER, E., COLE, G., RADHAKRISHNAN, R. & EPP, O. (1988). *Acta Cryst.* **B44**, 26–38.

STEWART, J. M. (1990). *PHONYD.* In *XTAL3.0 Reference Manual*, edited by S. R. HALL & J. M. STEWART, pp. 209–210. Univs. of Western Australia, Australia, and Maryland, USA.

SWANSON, R. M. (1990). *J. Chem. Educ.* **67**, 206–208.