Statistical estimation of the number of minima in a function with a finite number of variables

T. A. Kaplan
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824
(Received 8 August 1985)

The statistical estimate of the number of local minima in an energy function obtained by a finite random sampling, due to Walker and Walstedt (WW), is clarified. In particular, it is found that an additional assumption, of the Bayesian type, seems to be needed, and the consequences are discussed in some detail. The relaxation of the explicit assumption of WW, that each minimum has equal *a priori* probability of being sampled, is discussed briefly.

The study of local minima in classical energy functions is important in a variety of fields, e.g., structure and thermodynamic behavior of solids, including glasses and spin glasses. Most of the results in this field come from computer studies of systems of finite numbers N of particles. For example, recent work has studied local minima for nonmagnetic particles interacting via Lennard-Jones and similar potentials,[1,2] and for magnetic particles whose positions are fixed, their magnetic moments (i.e., spins) interacting according to a Heisenberg or similar model.[3]

The problem of determining the ground states (i.e., the states with the absolute—or global— minimum energy) is enormously difficult[4,5] for large N, as is the problem of determining all the metastable states (local minima with energy above the ground-state energy), or all those lying close to the ground state, even for moderate N.[6] The determination of the number of local minima (LM) for small N is felt[1-3] to have been accomplished both in the nonmagnetic and the magnetic models studied. The discussion supporting this in the nonmagnetic cases[1,2] seems to be purely intuitive, while an explicit probabilistic or statistical argument was given for the magnetic case.[3]

My concern here is with the latter probabilistic argument (an argument which is not at all limited by the type of model used for the energy). My purpose is to clarify the assumptions needed for the argument, and to indicate possible generalizations.

The argument[3] is as follows. A set of M "quenches" is made, each quench starting from a randomly chosen state (or point in the configuration space), each quench leading to a definite local minimum. These quenches result in $K(\leq M)$ independent (i.e., distinct[7]) local minima. According to Walker and Walstedt,[3(a)] if we assume that all local minima "have equal likelihood of being detected," the probability that there are Z independent local minima and only $Z-1$ occurred in a random sample of M starting configurations is

$$P_1(M)=Z\left(1-\frac{1}{Z}\right)^M . \tag{1}$$

The similar case for the occurrence of $(Z-2)$ independent local minima is[3]

$$P_2(M)=\tfrac{1}{2}Z(Z-1)\left(1-\frac{2}{Z}\right)^M . \tag{2}$$

These were applied to the case $M=70$, $K=7$: $Z=8$ in (1) and $Z=9$ in (2) gave very small probabilities $P_1(70)$ and $P_2(70)$, leading to the conclusion $Z=7$ is highly likely. Also, another case $M=50$, $K=41$ was considered, for which it was concluded without detailed explanation that $Z\approx 200$ [see Ref. 3(b)].

I argue here that it is not possible on the basis of the stated assumptions to determine the (joint) probability that there are Z local minima and $Z-s$ occur on M trials. I also will derive the right-hand sides of (1) and (2) (and the generalization to $Z-s$ local minima) as results of different probabilistic statements, and, finally, come up with essentially the same conclusions as Walker and Walstedt.[3(a),3(b)]

What *can* be calculated is the conditional probability $P(M,Z-s\,|\,Z)$ that on M trials $Z-s$ different local minima will occur, *if* there are Z local minima altogether. [This is not the right-hand side of (1) for $s=1$ or of (2), for $s=2$, as will be seen.] We are then interested in the "inverse" question: What is the conditional probability $P(Z\,|\,M,K)$ that there are Z local minima if on M trials K local minima occur? This is written down according to the Bayesian approach[8] as follows. A general relation of probability theory is

$$P(Z\,|\,M,K)=P(M,K\,|\,Z)\frac{W(Z)}{W(M,K)} , \tag{3}$$

where $W(\cdot)$ are unconditional probabilities. Hence the relative conditional probability of Z' to Z is

$$\frac{P(Z'\,|\,M,K)}{P(Z\,|\,M,K)}=\frac{P(M,K\,|\,Z')}{P(M,K\,|\,Z)}\frac{W(Z')}{W(Z)} . \tag{4}$$

If we now interpret $W(Z)$ as an "*a priori* probability," and, further (as seems reasonable), assume it to be independent of Z for $0\leq Z\leq Z_c$, where Z_c is much larger than any value that might be expected for the problem at hand, we see that the desired probability $P(Z\,|\,M,K)$ is just proportional to the calculable probability $P(M,K\,|\,Z)$, for $0\leq Z\leq Z_c$.[9]

The probability $P(M,K\,|\,Z)$ obviously may be expressed as follows. This probability, that M random trials (thrown balls) will result in K distinct local minima (balls landing in K different boxes of equal size), is the number of ways of picking K boxes out of Z boxes times the number $\mathcal{N}(M,K)$ of ways that M balls can be put into K boxes such that every box is occupied, i.e., contains at least one ball, divided by the total number of ways Z^M of putting M balls into Z

boxes. That is,

$$P(M,K|Z)=\binom{Z}{K}\frac{\mathscr{N}(M,K)}{Z^M} \quad . \tag{5}$$

According to the above assumption then, the desired probability

$$P(Z|M,K)=\frac{f(Z;M,K)}{\sum_{Z'} f(Z';M,K)} \quad , \tag{6}$$

where

$$f(Z;M,K)=\binom{Z}{K}Z^{-M} \quad . \tag{7}$$

Equation (5) was derived in a slightly different form, using a somewhat more complicated argument, by Feller.[10] Pleasantly, (7) is independent of the mathematically more complicated function[10] $\mathscr{N}(M,K)$, which is needed in (5), and which is discussed below. Equations (6) and (7) give

$$P(Z|M,K)=0 \text{ for } Z<K \quad , \tag{8}$$

and, asymptotically for large Z,

$$f(Z;M,K)\approx\frac{Z^{-(M-K)}}{K!} \quad . \tag{9}$$

Hence for $M-K>1$, the normalization sum (on Z) converges. This insures that predictions of this theory will be essentially independent of the cutoff Z_c for large enough values; thus we will take $Z_c\to\infty$.

Using (6) and (7), we have for the ratio of the probabilities of finding $K+1$ and K boxes occupied on throwing M balls

$$\frac{P(K+1|M,K)}{P(K|M,K)}=(K+1)\left[\frac{K}{K+1}\right]^M \quad . \tag{10}$$

Similarly,

$$\frac{P(K+2|M,K)}{P(K|M,K)}=\frac{(K+2)(K+1)}{2}\left[\frac{K}{K+2}\right]^M \quad . \tag{11}$$

Putting $K=Z-1$ in (10) and $Z-2$ in (11) yields (1) and (2), respectively, and applying this to the case $M=70$, $K=7$, we obtain $8\times(\frac{7}{8})^{70}\cong 7\times10^{-4}$ and $(9\times\frac{8}{2})(\frac{7}{9})^{70}\cong 8\times10^{-7}$, in agreement with Walker and Walstedt.[3(a)]

Thus we have derived the right-hand sides of Eqs. (1) and (2), each as a ratio of conditional probabilities of Z for two different values—this is quite different from the probability meaning attached previously.[3] To accomplish this we needed an important additional assumption, over and above that stated,[3] namely, the Bayesian-type assumption needed to deal with Eq. (4). Finally, we unsuccessfully tried to derive (1) and (2) by simpler interpretations, namely (i) the authors' words[3] and (ii) replacement of those words by a reasonable conditional probability. That (i) is impossible is seen from (3) and (5): The (joint) probability of having Z boxes and $Z-1$ occupied ones occurring on throwing M balls is $P(M,Z-1|Z)W(Z)$, whereas the right-hand side of (1) is

$$P(M,Z-1|Z)\frac{1}{P(M,Z-1|Z-1)} \quad ,$$

which is not of the correct form (the second factor depending on M as well as Z). The trial (ii) is to try to reinterpret the meaning of (1) to be the conditional probability that if there are Z boxes, then $Z-1$ will occur on throwing M balls. However, the latter is given by (5):

$$P(M,Z-1|Z)=Z\frac{\mathscr{N}(M,Z-1)}{Z^M} \quad ;$$

the special case $M=4$, $Z=3$ gives $P(4,2|3)=42/3^4$ [see below for $\mathscr{N}(4,2)=14$], whereas Eq. (1) gives $P_1(4)=46/3^4$.

It is interesting to apply (6) and (7) to the other case considered previously,[3] namely, $M=50$, $K=41$. I found the most probable Z, $Z_0=120$; the median value of Z is 138 and the mean value is 149.1. Also, the probabilities that $Z<85$ and $Z>250$ are each about 5%. These results support in a rough sense the earlier conclusion[3] $Z\approx200$.

Because of the controversial history of the Bayesian approach,[8,11] it is important to assess the reasonableness of the result of the approach. In general, one can find Z_0, the most probable Z, quite simply by considering

$$\frac{f(Z-1;M,K)}{f(Z;M,K)}=\frac{1-K/Z}{(1-1/Z)^M}=\frac{1-Kx}{(1-x)^M} \quad . \tag{12}$$

Since for $M>K>1$,

$$F(x)=1-Kx-(1-x)^M \tag{13}$$

is easily seen to have only one zero x_0 in $0<x\leq1$, it follows that as a function of Z, $f(Z;M,K)$ has only one maximum, and it occurs at x_0^{-1}, Z_0 being the nearest integer to this. It is easy to see that

$$x_0^{-1}\to K \text{ as } M\to\infty \quad . \tag{14}$$

Also if one increases M holding x_0^{-1} fixed, one can see from (12) and (13) [with $F(x_0)=0$ giving $K=K(M)$] that the probability $P[Z|M,K(M)]$ becomes more sharply peaked. Equation (14) plus this property is quite reasonable. Also, the property (8) is not only reasonable, it is an absolute necessity. Another very nice property is the unnormalizability of $f(Z;M,K)$ for $K=M$ [Eq (9)]: This says that if on M throws, M boxes are occupied, then one cannot estimate an upper limit on Z—clearly, this is as it must be. The lack of convergence for $K=M-1$ is not so obviously required, but is not unreasonable. Thus the result of the Bayesian approach here has indeed turned out to be quite reasonable.

An important assumption that needs to be relaxed is that all the minima are *a priori* equally probable. Surely for the energy functions considered[1-3] the "catchment regions"[1] or basins are most probably not of equal size. In the balls and boxes transcription, this means the box openings are not of equal area. Intuitively, it seems that any deviation from equal probabilities would cause an estimate of Z made on the assumption of equal probabilities to be too small, or, at least, never too large. I will not try to prove this here, but will be content with a few remarks.

One can easily write down appropriate generalizations of (5). For example, suppose of the Z boxes, Z_a and Z_b of them were each assigned *a priori* probabilities p_a and p_b, respectively; here

$$Z_a+Z_b=Z \quad , \qquad Z_ap_a+Z_bp_b=1 \quad . \tag{15}$$

Then the conditional probability that M balls will occupy K different boxes is, evidently,

$$P(M,K\,|\,Z_a,Z_b,p_a,p_b)=\sum_{M_a,M_b}\sum_{K_a,K_b}\begin{pmatrix}M\\M_a\end{pmatrix}\begin{pmatrix}Z_a\\K_a\end{pmatrix}\begin{pmatrix}Z_b\\K_b\end{pmatrix}\mathcal{N}(M_a,K_a)\mathcal{N}(M_b,K_b)p_a^{M_a}p_b^{M_b} \ . \tag{16}$$

Here the sums over the non-negative integers M_a, M_b, K_a, and K_b are restricted by $M_a+M_b=M$ and $K_a+K_b=K$. Now it is necessary to deal with the function $\mathcal{N}(M,K)$; this is given by Feller[10] as

$$\mathcal{N}(M,K)=K^M\sum_{l=0}^{K}(-)^l\begin{pmatrix}K\\l\end{pmatrix}\left(1-\frac{l}{K}\right)^M \ . \tag{17}$$

Note that from (5) we have an equivalent relation

$$\sum_{K=0}^{P}\begin{pmatrix}P\\K\end{pmatrix}\mathcal{N}(M,K)=P^M \ . \tag{18}$$

A few examples, using the Kronecker δ function, follow:

$$\mathcal{N}(M,K)=0 \ \text{ for } K<M \ , \quad \mathcal{N}(M,0)=\delta_{M0} \ , \tag{19}$$

and for $M\geq K$,

$$\mathcal{N}(M,2)=2^M-2 \ , \quad \mathcal{N}(M,3)=3^M-3(2^M-2) \ . \tag{20}$$

For large K it is clearly clumsy to calculate these $\mathcal{N}(M,K)$. Feller[10] gives an approximate expression which he states is good for large K; unfortunately this is not quite true: His criterion also requires $M>>K$ (strictly, his condition[10] requires $K\rightarrow\infty$ and $M/K\rightarrow\infty$).

For a simple example I considered the case $Z_a=Z_b=Z/2$ (Z even), $M=4$, $K=2$. Writing $p_b=\nu p_a$, I found [using (15)]

$$P\left(4,2\,|\,\frac{Z}{2},\frac{Z}{2},p_a,\nu p_a\right)f(Z;\nu)=\left(\frac{2}{Z(1+\nu)}\right)^4\left(\frac{7(1+\nu^4)+4\nu+6\nu^2+4\nu^3}{4}Z^2-\frac{7}{2}(1+\nu^4)Z\right) \ . \tag{21}$$

[This checks with (5) for $\nu=1$.] Writing the second bracket in (21) as BZ^2-CZ, one sees that the value Z_m that maximizes $f(Z;\nu)$ is

$$Z_m=\frac{3}{2}\frac{C}{B} \ ,$$

and this is the only stationary point (for finite Z). For $\nu=1$ this is $\frac{3}{2}$, implying the most probable value (of the allowed values $Z=2,4,6,\ldots$) is $Z_0=2$. Furthermore, the minimum of Z_m over ν is this $\nu=1$ value, so Z_m and therefore Z_0 cannot decrease when ν changes from 1. For $\nu\rightarrow\infty$ or 0 [note $f(Z;\nu^{-1})=f(Z;\nu)$], $C/B\rightarrow 2$, so $Z_m\rightarrow 3$. It turns out that as ν increases past $\bar{\nu}=4.68$, Z_0 switches from 2 to 4 and remains there as ν continues to increase (the latter is a very reasonable behavior). Notice that this increase in the most probable value Z_0 when ν changes from 1 is in accord with the intuitive thought mentioned above.

I thank J. Borysowicz, D. Bruce, G. Kemeny, S. de Leeuw, G. Pollack, and S. D. Mahanti for useful discussions. In particular, I am indebted to Professor Borysowicz for pointing out the equivalence of my analysis to the Bayesian approach.

[1]M. R. Hoare and J. A. McInnes, Adv. Phys. **32**, 792 (1983).

[2]F. H. Stillinger and T. A. Weber, Science **225**, 983 (1984).

[3](a) L. R. Walker and R. E. Walstedt, Phys. Rev. B **22**, 3816 (1980); (b) L. R. Walker and R. E. Walstedt, J. Magn. Magn. Mater. **31–34**, 1289 (1983); (c) R. E. Walstedt, in *Heidelberg Colloquium on Spin Glasses*, edited by J. L. Van Hemmen and J. Morganstern, Lecture Notes in Physics, Vol. 192 (Springer-Verlag, New York, 1983), p. 177.

[4]L. T. Wille and J. Vennek, J. Phys. A **18**, L419 (1985).

[5]Although in some special cases this determination is possible; see T. A. Kaplan, Bull. Acad. Sci. USSR, Phys. Ser. **28**, 328 (1964), for a review concerning spin models.

[6]There are two cases, both spin models, where the counting of such states was accomplished (at least substantially) for $N\rightarrow\infty$: T. A. Kaplan, Phys. Rev. B **24**, 319 (1981); A. J. Bray and M. A. Moore, J. Phys. C **14**, 2629 (1981).

[7]See Ref. 3 for the technical definition of distinct.

[8]See E. T. Jaynes, in *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus (MIT Press, Cambridge, 1979), p. 15; J. R. Borysowicz, in *Proceedings of the Workshop on Advanced Methods in the Evaluation of Nuclear Scattering Data, Berlin 1985*, edited by H. J. Krappe and R. Lipperheide (Springer-Verlag, New York, in press).

[9]Other assumptions for $W(Z)$ might be appropriate, particularly if some *a priori* information were available.

[10]W. Feller, *Introduction to Probability Theory and Its Applications*, 3rd ed. (Wiley, New York, 1968), Vol. I, Chap. IV, Sec. 2.

[11]See also Ref. 10, Chap. V, Sec. 2.