

TEST OF A PROBABILISTIC MODEL
OF EVOLUTIONARY SUCCESS

H. Epstein and D. Ruelle

I.H.E.S.
35, route de Chartres
91440 Bures-sur-Yvette, France

Abstract. A proposed relation between spin glasses and biological evolution is given a precise form, using a probabilistic model called Generalized Random Energy Model (GREM). Here we test this idea using the taxonomic distribution of European monocots and dicots. Although a rough agreement is observed, there are systematic deviations which may be attributable to competition between closely related species.

Disordered systems called spin glasses are currently the object of intensive study by theoretical physicists. In particular, it has been claimed that the structure of equilibrium states (valleys) of a spin glass and the tree structure of biological evolution are related ¹⁾. We present here a version of this claim which can in principle be tested. This is based on a probabilistic model - the GREM - which is supposed to represent the probabilistic structure of valleys for a spin glass. More precisely, the GREM (Generalized Random Energy Model) introduced by Derrida ²⁾ represents the solution based on Parisi's Ansatz for the Sherrington-Kirkpatrick spin glass model ³⁾. We shall now forget about spin glasses and use a recent reformulation ⁴⁾ of the GREM which emphasizes the universality of this probabilistic model.

Let $0 < x_1 < x_2 < \dots < x_n < 1$. For each level $i = 1, 2, \dots, n$, the GREM describes a population of objects $(\alpha_1, \dots, \alpha_i)$, with weights $\xi(\alpha_1, \dots, \alpha_i)$ such that $\sum_{\alpha_i} \xi(\alpha_1, \dots, \alpha_i) = \xi(\alpha_1, \dots, \alpha_{i-1})$. For fixed $\alpha_1, \dots, \alpha_{i-1}$ the weights $\xi(\alpha_1, \dots, \alpha_i)$ are independently distributed with a density proportional to

$$x_i \xi_i^{-x_i - 1} \tag{1}$$

(In principle there are infinitely many values of α_i , corresponding to the fact that $\int_0^\epsilon x \xi^{-x-1} d\xi = \infty$). Note that we need define the weights $\xi(\alpha_1, \dots, \alpha_i)$ only up to an arbitrary multiplicative constant, since this is absorbed in the proportionality factor of (1). Note also that if we allow the $\alpha_1, \dots, \alpha_{i-1}$ to vary, the distribution of the $\xi(\alpha_1, \dots, \alpha_i)$ remains described by a density proportional to (1).

In an application to biological evolution, $\xi(\alpha_1, \dots, \alpha_n)$ could measure the success (in a sense to be made more precise) of a species which at times ⁵⁾ $t_1, t_2, \dots, t_n = \text{present}$, has gone through the stages $\alpha_1, \alpha_2, \dots, \alpha_n$. Such an idea

is hard to test, and we have here tested something related but different and more amenable to analysis. With $n = 3$, we let $\xi(\alpha_1, \alpha_2, \alpha_3)$ be the number of species in the genus α_3 of family α_2 of order α_1 , of a suitably prescribed set of species. Specifically we have tested the European dicots and monocots as described in Flora Europaea ⁶⁾. (We have used blindly the numbers printed in the Flora Europaea, ignoring subtleties due to introduced species, hybrids, apomixis, etc. ; it appears that taking these into account would not alter our conclusions.) We compute the number $f_A(k)$ of orders, $f_B(k)$ of families or $f_C(k)$ of genera which contain at least k species. According to the GREM these should be proportional to

$$\int_k^\infty x_i \xi^{-x_i-1} d\xi = k^{-x_i} \quad (2)$$

In a log-log plot, f_A , f_B and f_C should appear as approximately straight lines with slopes $-x_1$, $-x_2$, $-x_3$. As the figures show, this is approximately true, and the slopes for monocots and dicots are remarkably similar : $x_1 \cong 0.2$, $x_2 \cong 0.5$, $x_3 \cong 0.8$ for the left-hand part of the graphs. There is however an unquestionable bending of the curves. This bending expresses that large taxonomic units are less large than expected on the basis of the GREM. There might of course be a failure of botanists to subdivide small genera as much as large ones, or to distinguish all the related species of a large genus. Another explanation is that competition among related organisms in a limited geographic range may prevent too many distinct species from surviving : this would cause a deviation from the assumptions of independence which are at the basis of the GREM.

In conclusion, it appears that, in the example considered, a probabilistic description of biological evolution in terms of the GREM is not totally unreasonable, but overestimates the size of large taxonomic units.

Footnotes

- 1) See R. Rammal, G. Toulouse and M.A. Virasoro. Ultrametricity for physicists. Rev. Mod. Phys., to appear. See also the earlier papers of M. Eigen and P. Schuster. The hypercycle. Naturwissenschaften 64, 541-565 (1977); 65, 7-41 (1978); 65, 341-369 (1978); D. Ruelle. A mechanism for speciation based on the theory of phase transitions. Math. Biosciences 56, 71-75 (1981); P.W. Anderson. Suggested model for prebiotic evolution : the use of chaos. Proc. Natl. Acad. Sci. USA 80, 3386-3390 (1983).
- 2) B. Derrida. A generalization of the Random Energy model which includes correlations between energies. J. Physique Lett. 46, L-401-L-407 (1985)
- 3) See M. Mézard, G. Parisi, N. Sourlas, G. Toulouse and M. Virasoro. Replica symmetry breaking and the nature of the spin glass phase, J. Physique 45, 843-854 (1984).
- 4) D. Ruelle. A mathematical reformulation of Derrida's REM and GREM. Preprint.
- 5) The t_i may have to be understood as phylogenetic stages rather than properly chronological times.
- 6) Flora Europaea (T.G. Tutin, V.H. Heywood, N.A. Burges, D.M. Moore, D.H. Valentine, S.M. Walters, D.A. Webb, editors). Cambridge University Press, 1964, 1968, 1972, 1976, 1980.

Captions

Figure 1 : log - log plots of the numbers $f_A(k)$, $f_B(k)$, $f_C(k)$ of (A) orders, (B) families, (C) genera with more than k species in Flora Europaea.

In each case the upper curve is dicots and the lower curve monocots. The significant points are the lower points at discontinuities. The vertical bar is at the log of the largest number of species in an order, family or genus (where log - becomes $-\infty$: large bar for dicots, small bar for monocots). If natural logs are used, the horizontal and vertical coordinates vary from 0 to 8.

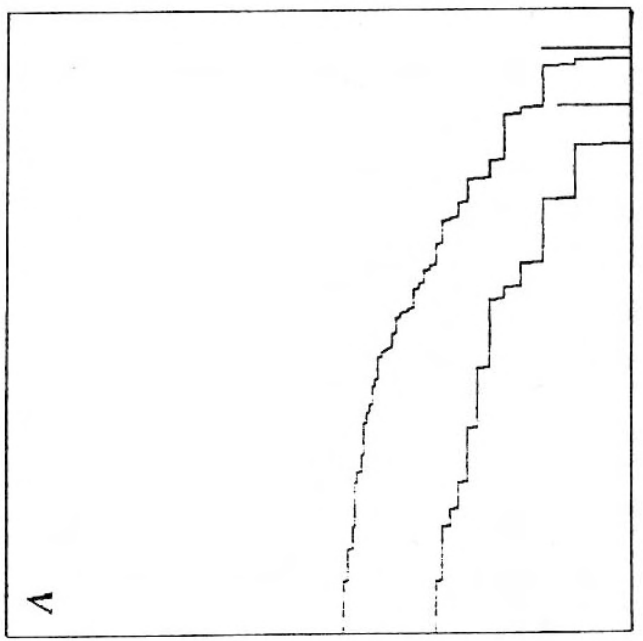
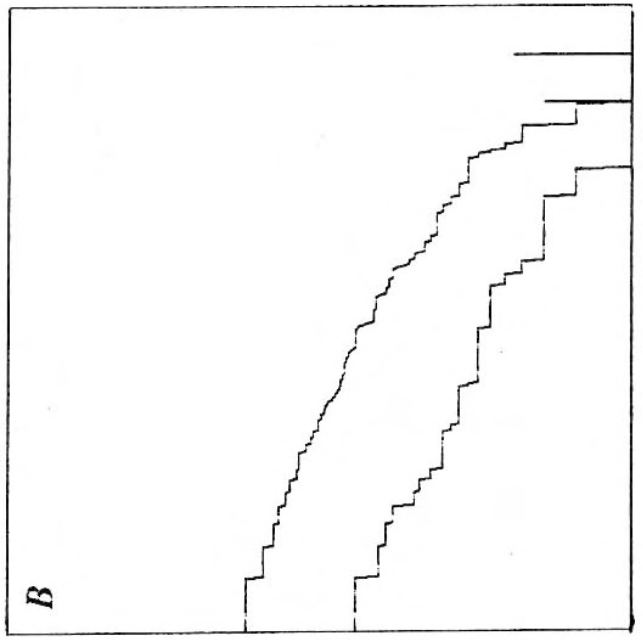
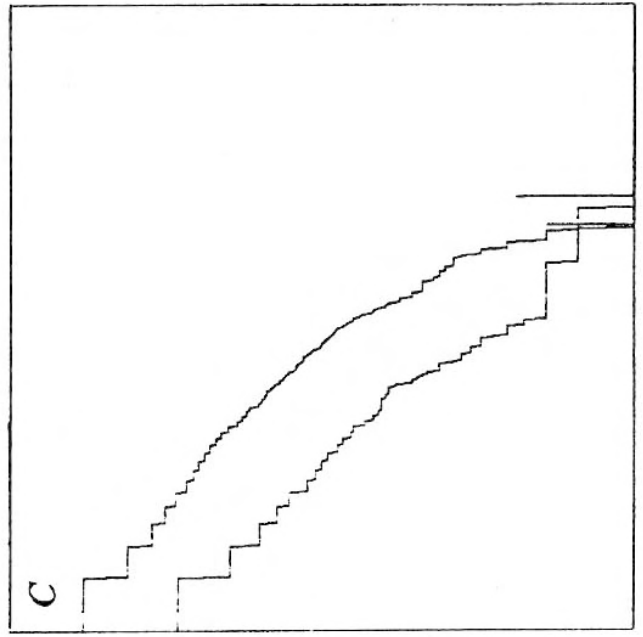


Figure 1