Self-averaging in finite random copolymers

N. Madras[†] and S.G. Whittington[‡]

[†] Department of Mathematics and Statistics, York University,
Toronto, Canada M3J 1P3

[‡]Department of Chemistry, University of Toronto,
Toronto, Canada M5S 3H6

June 12, 2002

Abstract

We investigate how martingale techniques can be used to derive information on the extent of self-averaging of the free energy for some lattice models of finite random copolymers.

Random copolymers are an interesting example of systems with quenched randomness (Brout 1959). The sequence of n comonomers in the random copolymer is determined by some random process but this sequence is then fixed in each particular polymer molecule. The quenched average free energy is the average of the free energy over all such randomly chosen sequences. One is interested in the extent to which properties depend on the particular sequence of comonomers and one expects that thermodynamic properties such as the free energy will self-average in the infinite n limit. That is, the free energy in this limit will be equal to the limiting quenched average free energy, for almost all comonomer sequences.

Recently there has been some interest (Chuang *et al* 2001, Naidenov and Nechaev 2001, James and Whittington 2002) in the extent of self-averaging at finite n. To what extent does a property self-average when the copolymer is finite? That is, how narrow is the distribution at finite n? In this Letter we show that this question can be answered, to some extent, using martingale methods (see eg Karlin and Taylor 1975, Varadhan 2001). The general idea is to consider the free energy of the finite n system, and to construct a martingale difference sequence. Provided that one can show that the differences are bounded one can then use Azuma's inequality (Azuma 1967) to establish a bound on the rate of convergence.

Here is an informal description of our results. We shall use χ to denote a particular random sequence of monomers. For a polymer with n monomers, the simplest case has each monomer randomly chosen to be one of two types, resulting in 2^n possibilities. Once such a sequence of monomers χ is chosen, we define the free energy per site $\kappa_n(\beta|\chi)$, obtained

by considering all conformations of an n-site polymer and their associated energies. The conformations do not depend on χ, but their energies do. For given n and β, the quantity $\kappa_n(\beta|\chi)$ is a random variable. Our main result (Equation (17)) is that the tails of this distribution cannot be heavier than Gaussian with standard deviation $O(n^{-1/2})$. In particular, let b be any number less than $1/2$, and consider the probability that $\kappa_n(\beta|\chi)$ differs from its expected value by more than n^{-b}. Then this probability converges to 0 as $n \to \infty$ (see Equation (16)). In contrast, the limit self-averaging property (which has been proven for various models in other papers) says that the random variables $\kappa_n(\beta|\chi)$ *converge* as $n \to \infty$ to a *nonrandom* limit.

We begin by giving a general description of the approach in the context of random linear copolymers, modelled by randomly coloured self-avoiding walks on a lattice, and then apply the idea to obtain bounds on the extent of self-averaging for several specific problems. We also observe that the same idea can be used to handle problems involving branched copolymers.

Consider the set of n-edge self-avoiding walks on the d-dimensional hypercubic lattice. Each vertex of the walk can be coloured A or B independently, with a vertex being coloured A with probability p. We write $\chi_i = 1$ if the i'th vertex is coloured A and $\chi_i = 0$ if the i'th vertex is coloured B. Given a sequence of colours $\chi_1, \chi_2, \ldots \chi_n$, we associate the i'th vertex of each walk with the colour χ_i. That is, the colours correspond to the sequence of comonomers in the copolymer. By convention, the 0'th vertex is uncoloured although this has essentially no effect on the argument.

Given a sequence of colours $\chi_1, \chi_2, \ldots \chi_n$ each n-edge self-avoiding walk ω has an associated energy $H(\omega|\chi_1, \chi_2, \ldots \chi_n)$. This energy will of course depend on the model being considered. The partition function is defined as

$$Z_n(\beta|\chi_1, \chi_2, \ldots \chi_n) = \sum_{\omega} e^{\beta H(\omega|\chi_1, \chi_2, \ldots \chi_n)} \tag{1}$$

and the corresponding free energy is

$$\kappa_n(\beta|\chi_1, \chi_2, \ldots \chi_n) = n^{-1} \log Z_n(\beta|\chi_1, \chi_2, \ldots \chi_n). \tag{2}$$

For given β and n, the quantity κ_n is a function of the n independent random variables $\chi_1, \ldots, \chi_n$. It turns out that this allows us to show that the random variable κ_n is close to its mean with high probability, and that this can be quantified in the mathematical framework of martingale theory, as we now explain. For each $k = 0, 1, \ldots, n$, define the conditional expectation

$$M_k = E[\kappa_n(\beta|\chi_1, \chi_2, \ldots \chi_n)|\chi_1, \chi_2, \ldots \chi_k] \tag{3}$$

which is the conditional expectation of the free energy given the colours $\chi_1, \chi_2, \ldots \chi_k$. Observe that M_k is a function of $\chi_1, \chi_2, \ldots \chi_k$ and that the conditional expectation (3) amounts to an integration over the colours $\chi_{k+1}, \chi_{k+2}, \ldots \chi_n$. Note that

$$M_0 = E[\kappa_n(\beta|\chi_1, \chi_2, \ldots \chi_n)] \tag{4}$$

is the quenched average free energy, in which the expectation is taken over all colourings. It is easy to check that $M_0, M_1, \ldots M_n$ is a martingale, i.e. $E(M_k|\chi_1, \ldots, \chi_{k-1}) = M_{k-1}$ for

each k. Define the *martingale difference sequence*

$$d_k = M_k - M_{k-1} \tag{5}$$

so that

$$\begin{aligned} \sum_{k=1}^{n} d_k &= (M_n - M_{n-1}) + (M_{n-1} - M_{n-2}) + \ldots (M_1 - M_0) \\ &= M_n - M_0 \\ &= \kappa_n(\beta|\chi_1, \chi_2, \ldots \chi_n) - E[\kappa_n(\beta|\chi_1, \chi_2, \ldots \chi_n)]. \end{aligned} \tag{6}$$

This is the extent to which the free energy at size n with a particular sequence of colours differs from the quenched average free energy.

Martingales can be viewed as generalizations of partial sums of independent random variables. In general, the terms of a martingale difference sequence are not independent, but they are uncorrelated and have other appealing properties (see for instance Section 1.6 of Hall and Heyde 1980). There is a body of theory that shows that martingales under certain conditions behave like sums of independent random variables, eg they often obey central limit theorems (Hall and Heyde 1980). We cannot derive central limit theorems in our models, but we are able to apply other martingale machinery to give Gaussian bounds on the tails of the distributions. The main technical tool is the following, which is a special case of Azuma's inequality (Azuma 1967, Steele 1997).

Proposition 1 *Assume that $M_0, \ldots, M_n$ is a martingale, and that K is a number (possibly depending on n) such that*

$$\|d_k\|_\infty \leq K, \tag{7}$$

for $k = 1, \ldots, n$. Then for every $\lambda > 0$

$$Pr(|M_n - M_0| \geq \lambda) \leq 2e^{-\lambda^2/2nK^2}. \tag{8}$$

As we show below, in many problems we can take K of the form A/n for some constant A. Thus if we put $\lambda = Bn^{-1/2+\epsilon}$ for any positive B and positive ϵ we have

$$Pr\left(|M_n - M_0| \geq Bn^{-1/2+\epsilon}\right) \leq 2e^{-(B^2/2A^2)n^{2\epsilon}} \tag{9}$$

which goes to zero as n goes to infinity. Indeed, for any $\epsilon > 0$ and for any $B > 0$ we can find an integer $N = N(\epsilon, B)$ such that for $n > N$ the right hand side of (9) can be made as small as we wish. Alternatively, we could take $\lambda = t/\sqrt{n}$, which leads to

$$Pr\left(\sqrt{n}|M_n - M_0| \geq t\right) \leq 2e^{-t^2/2A^2} \quad \text{for all } t > 0, \tag{10}$$

which highlights the sub-Gaussian nature of the tails.

The following lemma describes conditions under which the hypotheses of Proposition 1 hold in our models. This approach has been used in various contexts (see for example Section 1.3 of Steele (1997), Section 3 of McDiarmid (1998), Equation (1.10) of Talagrand (1998)).

Lemma 1 *Let $\chi_1, \ldots, \chi_n$ be any sequence of random variables. Let Y be any random variable, and let*

$$M_k \;=\; E(Y \,|\, \chi_1, \ldots, \chi_k) \quad (k = 0, 1, \ldots, n).$$

(a) It follows that the sequence $M_0, M_1, \ldots, M_n$ is a martingale.
(b) Assume that $\chi_1, \ldots, \chi_n$ are independent random variables, and that Y is of the form

$$Y \;=\; f(\chi_1, \ldots, \chi_n)$$

for some deterministic function f that satisfies

$$|f(x) - f(x')| \;\le\; K \tag{11}$$

for every two n-component vectors x and x' that differ in exactly one component. Then $||d_k||_\infty \le K$ for every k.

For a proof, see for instance Section 1.3 of Steele (1997).

We shall apply the above lemma with

$$f(\chi_1, \ldots, \chi_n) \;=\; \kappa_n(\beta|\chi_1, \ldots, \chi_n).$$

The basic step is to establish the inequality (11). We first consider the problem of adsorption of a random copolymer at an impenetrable surface. We write $(x, y, \ldots z)$ for the coordinates of a vertex on the d-dimensional hypercubic lattice and $(x_i, y_i, \ldots z_i)$ for the coordinates of the i'th vertex of a self-avoiding walk on this lattice. We fix the 0'th vertex at the origin and consider the subset of walks for which $z_i \ge 0$ for all $i \le n$, so that the walk is confined to a half-space. For any particular fixed colouring $\chi \equiv \{\chi_1, \chi_2, \ldots \chi_n\}$ let $c_n(v_A|\chi)$ be the number of n-edge self-avoiding walks with colouring χ which have v_A vertices coloured A in the hyperplane $z = 0$. The partition function with this fixed colouring is

$$Z_n(\beta|\chi) = \sum_{v_A} c_n(v_A|\chi) e^{\beta v_A} \tag{12}$$

and the corresponding free energy is

$$\kappa_n(\beta|\chi) = n^{-1} \log Z_n(\beta|\chi). \tag{13}$$

If the colouring χ' differs from χ by changing the colour at (say) the k'th vertex then

$$e^{-|\beta|} Z_n(\beta|\chi) \;\le\; Z_n(\beta|\chi') \;\le\; e^{|\beta|} Z_n(\beta|\chi). \tag{14}$$

Hence

$$|\kappa_n(\beta|\chi) - \kappa_n(\beta|\chi')| \;\le\; \frac{|\beta|}{n} \tag{15}$$

which establishes (11) with $K = |\beta|/n$. For this problem it is already known (Orlandini *et al* 1999) that the free energy self-averages in the infinite n limit, but (15) together with Azuma's inequality imply that

$$Pr\left(\left|\kappa_n(\beta|\chi) - E[\kappa_n(\beta|\chi)]\right| \geq Bn^{-1/2+\epsilon}\right) \leq 2e^{-(B^2/2\beta^2)n^{2\epsilon}} \tag{16}$$

for any positive B and positive ϵ, which improves the result given in James and Whittington (2002) for the extent of self-averaging at finite n. Equation (10) becomes

$$Pr\left(\sqrt{n}\left|\kappa_n(\beta|\chi) - E[\kappa_n(\beta|\chi)]\right| \geq t\right) \leq 2e^{-t^2/2\beta^2} \quad \text{for all } t > 0. \tag{17}$$

Taking $\lambda = C\sqrt{\log n/n}$ in Equation (8) gives

$$Pr\left(\left|\kappa_n(\beta|\chi) - E[\kappa_n(\beta|\chi)]\right| \geq C\sqrt{\frac{\log n}{n}}\right) \leq 2n^{-(C^2/2\beta^2)}. \tag{18}$$

When $C > \sqrt{2}\,|\beta|$, the series $\sum_n n^{-(C^2/2\beta^2)}$ converges. Therefore, the Borel-Cantelli lemma tells us that

$$\limsup_{n\to\infty} \sqrt{\frac{n}{\log n}}\,\left|\kappa_n(\beta|\chi) - E[\kappa_n(\beta|\chi)]\right| \leq \sqrt{2}\,|\beta| \tag{19}$$

with probability 1.

The same approach works for other random copolymer problems. Consider self-interacting, randomly coloured self-avoiding polygons where vertices are coloured A or B independently and where $A-A$, $B-B$ and $A-B$ nearest neighbour pairs (ie contacts) have different interaction potentials. Suppose k is a vector with elements k_{AA}, k_{BB}, k_{AB} where k_{AA} is the number of AA contacts in the polygon, etc, and suppose $p_n(k|\chi)$ is the number of polygons with colouring χ and contact vector k. The partition function can be written as

$$P_n(\beta|\chi) = \sum_k p_n(k|\chi)e^{\beta g(k)} \tag{20}$$

where $g(k)$ is a linear function of the elements of the vector k, and where $|g(k)| \leq \gamma(k_{AA} + k_{BB} + k_{AB})$ for some fixed finite γ. For this problem the free energy is known to self-average in the infinite n limit (Janse van Rensburg *et al* 2001) but nothing is known about the extent of self-averaging for finite values of n. To use Azuma's inequality we need to establish the bound (11). If χ' differs from χ only by the colouring of one vertex then at most $2d-2$ contacts can change so $g(k)$ can change by at most $2\gamma(2d-2)$. Hence

$$\left|n^{-1}\log P_n(\beta|\chi') - n^{-1}\log P_n(\beta|\chi)\right| \leq 4\beta\gamma(d-1)/n. \tag{21}$$

Therefore the inequality (7) is satisfied with $K = 4\beta\gamma(d-1)/n$ and

$$Pr\left(\left|n^{-1}\log P_n(\beta|\chi) - E[n^{-1}\log P_n(\beta|\chi)]\right| \geq Bn^{-1/2+\epsilon}\right) \leq 2e^{-(B^2/32\beta^2\gamma^2(d-1)^2)n^{2\epsilon}} \tag{22}$$

for any positive B and ϵ. Analogues of (17), (18) and (19) also hold.

We note that a similar argument works for self-interacting randomly coloured self-avoiding walks, although in this case there is no proof that the limiting quenched average free energy exists. Nevertheless this method gives a bound on the extent of self-averaging for finite n.

Finally we point out that the same approach works for a model of random copolymer localization at a surface between two immiscible liquids (Martin *et al* 2000), for a randomly coloured lattice tree model of copolymer adsorption (You and Janse van Rensburg 2000) and, as well, for models of branched copolymer collapse. The method has the advantage that no concatenation is necessary and the colouring of the vertices can be carried out according to the underlying graph theoretic structure, so that it does not depend on the particular embedding. The argument, and final result, are almost identical to the arguments given above.

Acknowledgments: The authors wish to thank NSERC for financial support. The first author also thanks CITO (Ontario) for support.

References

Azuma K 1967 *Tohoku Math. J.* **19** 357-367

Brout R 1959 *Phys. Rev.* **115** 824-835

Chuang J, Grosberg A and Kardar M 2001 *Phys. Rev. Lett.* **87** 078104

Hall P and Heyde C C 1980 *Martingale Limit Theory and Its Application* (New York: Acadamic Press)

James E W and Whittington S G 2002 *J. Phys. A: Math. Gen.* **35** 3203–3211

Janse van Rensburg E J, Orlandini E, Tesi M C and Whittington S G 2001 *J. Phys. A: Math. Gen.* **34** L37–44

Karlin S and Taylor H M 1975 *A First Course in Stochastic Processes* (New York: Academic Press)

Martin R, Causo M S and Whittington S G 2000 *J. Phys. A: Math. Gen.* **33** 7903–7918

McDiarmid C 1998, in *Probabilistic Methods for Algorithmic Discrete Mathematics*, Habib M, McDiarmid C, Ramirez-Alfonsin J, and Reed B (eds) (Berlin: Springer).

Naidenov A and Nechaev S 2001 *J. Phys. A: Math. Gen.* **34** 5625–5634

Orlandini E, Tesi M C and Whittington S G 1999 *J. Phys. A: Math. Gen.* **32** 469–477

Steele J M 1997 *Probability Theory and Combinatorial Optimization* (Philadelphia: SIAM)

Talagrand M 1998 *Probab. Theory Relat. Fields* **110** 109–176

Varadhan S R S 2001 *Probability Theory* (Providence: AMS)

You S and Janse van Rensburg E J 2000 *J. Phys. A: Math. Gen.* **33** 1171–1186