

# Self-averaging sequences in the statistical mechanics of random copolymers

E J Janse van Rensburg<sup>1</sup>, A Rechnitzer<sup>1</sup>, M S Causo<sup>2</sup> and S G Whittington<sup>3</sup>

<sup>1</sup> Department of Mathematics, York University, Toronto, Canada M3J 1P3

<sup>2</sup> INFN and Scuola Normale Superiore, I-56100 Pisa, Italy

<sup>3</sup> Department of Chemistry, University of Toronto, Toronto, Canada M5S 3H6

Received 29 May 2001

Published 10 August 2001

Online at [stacks.iop.org/JPhysA/34/6381](http://stacks.iop.org/JPhysA/34/6381)

## Abstract

A constructive proof is given that certain problems in the statistical mechanics of random copolymers are thermodynamically self-averaging. The proof relies on the use of normal numbers, and potentially gives a method for calculating expectations in the quenched averaged ensemble. In particular, we give a new proof that adsorbing random copolymers are self-averaging.

PACS numbers: 82.35.Jk, 02.30.Lt, 36.20.Fz, 87.15.Cc, 02.10.De, 05.70.-a

## 1. Introduction

When a system has quenched randomness (Brout 1959) an important question arises regarding self-averaging in the system. The idea is most easily understood by way of an example. Consider a polymer, produced by a random copolymerization process of two comonomers  $A$  and  $B$ . Different polymer molecules produced by the process will have different sequences of  $A$  and  $B$ , and these sequences are produced by a random process. However, each molecule (once it has been produced) has a definite and fixed sequence of comonomers. For a polymer with  $n$  monomers and a sequence  $\chi \equiv \{\chi_1, \chi_2, \dots, \chi_n\}$ , where  $\chi_i = 1$  if the  $i$ th monomer is an  $A$  and 0 otherwise, properties will usually depend on the particular sequence  $\chi$ , and we can write  $P_n(\chi)$  for the corresponding value of property  $P$ . If we average over all sequences  $\chi$  we obtain the *quenched average*  $\langle P_n(\chi) \rangle$ . Property  $P$  is said to be *self-averaging* if

$$\lim_{n \rightarrow \infty} P_n(\chi) = \lim_{n \rightarrow \infty} \langle P_n(\chi) \rangle \quad (1.1)$$

for almost all  $\chi$ . If  $P$  is the free energy the system is said to be *thermodynamically self-averaging*. Note that self-averaging is not a trivial property of a model since examples of non-self-averaging properties are known (Derrida and Hilhorst 1981, Sourlas 1987).

Several models in statistical mechanics have been proved to be thermodynamically self-averaging, including random spin models with short-range (van Hemmen and Palmer 1982)

and long-range interactions (van Enter and van Hemmen 1983), self-avoiding walk models of polymer adsorption (Orlandini *et al* 1999) and localization (Martin *et al* 2000), a lattice tree model of copolymer adsorption (You and Janse van Rensburg 2000) and some simplified models of random self-interacting copolymers (Orlandini *et al* 2000, Janse van Rensburg *et al* 2001).

For these models we know, by an existence proof, that for almost all sequences of monomers the free energy converges to the quenched average free energy as the number of monomers goes to infinity, though nothing is known about the rate of convergence. It is a curious fact that we do not know any single sequence of comonomers which has this property! The aim of this paper is to construct examples of such sequences, based on normal numbers, and at the same time to give a constructive proof of self-averaging in these models. We shall focus on a particular model (adsorption of self-avoiding walks), but the approach can be applied to a variety of other models.

## 2. Normal numbers

A number is said to be normal if every finite subsequence of digits of a given length in base  $g$  occurs with uniform density. It is known that almost all numbers are normal (Borel 1909). In base 2 the sequence of some 0 and 1 will be used as colours of vertices in a self-avoiding walk model of copolymers.

We first give a brief review of normal numbers. For further details, see Hardy and Wright (1938).

**Definition 1.** Consider a number,  $\gamma$ , written in base  $g$ , and with some 0 appended after the digit to give it an infinite decimal expansion, if necessary. If the digit  $b$  occurs  $N_b$  times in the first  $N$  digits of  $\gamma$  and if  $N_b/N \rightarrow \rho(b)$  as  $N \rightarrow \infty$ , then we say that the digit  $b$  has frequency  $\rho(b)$ . A number is simply normal in base  $g$  if  $\rho(b) = 1/g$  for each of the  $g$  possible values of  $b$ .

**Definition 2.** A number is normal in base  $g$  if every combination  $b_1 b_2 \dots b_k$  of digits occurs with frequency  $1/g^k$  for every positive integer  $k$ .

Consider the number  $2/3$ . Expanding this number in base 2 gives  $0.101010, \dots$ , and so it is *simply normal*, since ‘0’ and ‘1’ both occur with a frequency of  $1/2$ . The sequence ‘11’ does not occur at all, and so  $2/3$  in base 2 is not *normal*. If we then rewrite  $2/3$  in base 4 we get  $0.222, \dots$ , and we see that it is no longer simply normal. Hence if a number is simply normal when written in one base, it is not necessarily simply normal in another. Lemma 1 below shows that if a number  $\gamma$  is normal in base  $g$ , then it must be simply normal in base  $g^m$ .

Perhaps the most well-known normal number is the Champernowne number (Champernowne 1933), defined (in base 2) by

$$0 \cdot 1, 10, 11, 100, 101, 110, 111, 1000, 1001, \dots \quad (2.1)$$

The only function of the commas is to separate the digits into easily recognizable groups, and the definition of this number in an arbitrary base  $g$  is apparent. (Notice that this number has different numerical values in different bases; however, our interest is not in its numerical value, but in the sequence of digits it presents.) Generalizations of the Champernowne number have also been studied: for example, if we define (in base 2)

$$S_r = 0 \dots 00, 0 \dots 001, 0 \dots 010, 0 \dots 011, \dots, 1 \dots 11 \quad (2.2)$$

where all  $2^r$  possible arrangements of  $r$  digits occur in ascending order, then the number

$$\cdot S_1, S_2, S_3, \dots, S_r, \dots$$

is normal in base 2, and, if  ${}_{\mu}S_r$  is the sequence formed by repeating  $S_r$   $\mu$  times in sequence, then

$${}_{\mu}S_1, {}_{\mu}S_2, {}_{\mu}S_3, \dots, {}_{\mu}S_r, \dots$$

is also normal, provided that  $\mu$  is a fixed positive integer. Other classes of normal numbers can be generated by using theorems due to Copeland and Erdős (1946), and to Szűsz and Volkmann (1994). For instance, if  $p_n$  is the  $n$ th prime then the concatenation of the primes  $p_1, p_2, \dots$  is normal (Copeland and Erdős 1946). Similarly (Szűsz and Volkmann 1994) the concatenation of the  $p_n$ th primes  $p_{p_1}, p_{p_2}, p_{p_3}, \dots$  and the related number  $[p_{p_1}]^2, [p_{p_2}]^2, \dots$  are normal in any base. (Note that these numbers have different numerical values in different bases. They are recipes for writing down examples of normal numbers in different bases.)

The next lemma shows that a number normal in base  $g$  is simply normal in base  $g^m$ , and this will eventually be a critical point in studying the limiting free energy of a self-avoiding walk coloured according to the digits of a normal number. The result can be obtained from a more general theorem in Kuipers and Niederreiter (1974) but we give a short, simple proof for completeness.

**Lemma 1.** *If a number is normal in base  $g$  then it is simply normal in base  $g^m$  for any positive integer  $m$ .*

**Proof.** Suppose that  $\gamma$  is not simply normal in base  $g^m$  for some positive integer  $m$ . In base  $g^m$   $\gamma$  is a sequence of digits from the set  $\{B_1, B_2, \dots, B_{g^m}\}$ . If  $\gamma$  is not simply normal in base  $g^m$  then there is at least one digit,  $B_1$  say, such that either

- (1) The limit  $\lim_{N \rightarrow \infty} N_{B_1}/N = \rho(B_1)$  exists and  $\rho(B_1) > g^{-m}$ , or
- (2) the limit  $\lim_{N \rightarrow \infty} N_{B_1}/N = \rho(B_1)$  exists and  $\rho(B_1) < g^{-m}$ , or
- (3)  $\lim_{N \rightarrow \infty} N_{B_1}/N$  does not exist.

Consider these three possibilities in turn.

- (1) This implies that the sequence of  $m$  digits  $b_1 b_2 \dots b_m$  corresponding to the digit  $B_1$  in base  $g^m$  occurs with frequency greater than  $g^{-m}$  which is impossible since  $\gamma$  is normal in base  $g$ .
- (2) This implies that there is a digit,  $B_2$  say, such that

$$\liminf_{N \rightarrow \infty} N_{B_2}/N > g^{-m}. \tag{2.3}$$

We show that this gives a contradiction. Place a comma between each of the digits of  $\gamma$  and then replace each digit by its corresponding sequence of  $m$  digits in base  $g$ , and denote the number obtained (in base  $g$  now) by  $\gamma_2$ . There is a comma between successive blocks of  $m$  digits in  $\gamma_2$  so that each such block corresponds to a single digit in  $\gamma$ . A sequence of  $m$  digits,  $b_1 b_2 \dots b_m$ , corresponding to the digit  $B_2$  in base  $g^m$ , either occurs between a pair of commas or crosses one of them<sup>4</sup>. Hence the total number of occurrences in  $\gamma_2$ ,  $n_{b_1 b_2 \dots b_m}$ , of the sequence of  $m$  digits  $b_1 b_2 \dots b_m$  corresponding to the digit  $B_2$  is at least as large as the number of occurrences,  $N_{B_2}$ , of  $B_2$  in  $\gamma$ . Hence

$$\liminf_{n \rightarrow \infty} \frac{n_{b_1 b_2 \dots b_m}}{n} \geq \limsup_{N \rightarrow \infty} \frac{N_{B_2}}{N} \geq \liminf_{N \rightarrow \infty} \frac{N_{B_2}}{N} > g^{-m}. \tag{2.4}$$

This implies that the sequence  $b_1 b_2 \dots b_m$  occurs with too great a frequency, which contradicts the fact that  $\gamma$  is normal in base  $g$ .

<sup>4</sup> Consider, for example, the number  $\gamma = 0.132$  written in base 4; the digit '3' occurs only once. Writing this in base 2 we have  $\gamma_2 = 0.01, 11, 10$  and we see that the sequence '11' occurs three times.

- (3) For case 3 we have three possibilities. Either
- (a)  $\limsup_{N \rightarrow \infty} N_{B_1}/N > \liminf_{N \rightarrow \infty} N_{B_1}/N \geq g^{-m}$  or
  - (b)  $g^{-m} \geq \limsup_{N \rightarrow \infty} N_{B_1}/N > \liminf_{N \rightarrow \infty} N_{B_1}/N$  or
  - (c)  $\limsup_{N \rightarrow \infty} N_{B_1}/N > g^{-m} > \liminf_{N \rightarrow \infty} N_{B_1}/N$ .

Since  $n_{b_1 b_2 \dots b_m} / n \geq N_{B_1} / N$ , it follows that

$$\limsup_{N \rightarrow \infty} N_{B_1} / N \leq \rho(b_1 b_2 \dots b_m) = g^{-m} \quad (2.5)$$

so that 3(a) and 3(c) are impossible. Condition 3(b) implies that there is a digit  $B_2$  satisfying

$$\limsup_{N \rightarrow \infty} N_{B_2} / N > g^{-m} \quad (2.6)$$

and we can apply an argument essentially identical to that used for condition 2. This proves the lemma.  $\square$

### 3. Adsorption of random linear copolymers

We now turn our attention to a self-avoiding walk model of random copolymer adsorption. Consider the simple cubic lattice  $\mathbb{Z}^3$ . We write  $(x, y, z)$  for the coordinates of a vertex of  $\mathbb{Z}^3$ . We shall consider a self-avoiding walk with  $n$  edges on  $\mathbb{Z}^3$ , which starts at the origin and is confined to the half-space  $z \geq 0$ . Such a walk is called a *positive walk*. We write  $(x_i, y_i, z_i)$  for the coordinates of the  $i$ th vertex,  $i = 0, 1, \dots, n$ , so that  $z_0 = 0$  and  $z_i \geq 0$  for all  $i > 0$ . The zeroth vertex is uncoloured and the remaining vertices of the walk are coloured independently and uniformly by a random variable belonging to a probability space  $Y$ . A sequence of colours  $\chi = (\chi_1, \chi_2, \chi_3, \dots, \chi_n)$  can be sampled from the product space  $X = Y \times Y \times Y \times \dots \times Y$ . We shall consider colourings by only two colours  $A$  and  $B$ , but this can easily be generalized to cases with any finite number of colours. We shall be primarily concerned with the case  $n \rightarrow \infty$ , so that the sequences of colours are infinite. If colours are assigned to a walk of size  $n$ , then only the first  $n$  colours in an infinite sequence are used.

A (product) measure  $\mu$  can be defined on the space  $X$ , and if  $X$  is interpreted as a probability space, then  $\mu(X) = 1$ . Let  $X'$  be that subspace of  $X$  which contains all the sequences of colours which are normal. Then it is also known that  $\mu(X') = 1$ , since almost every sequence is normal (Borel 1909).

Let  $c_n(v|\chi)$  be the number of positive walks with  $n$  edges, with vertices  $1, 2, \dots, n$  coloured  $\chi_1, \chi_2, \chi_3, \dots, \chi_n \equiv \chi$ , having  $v$  vertices coloured  $A$  in the surface  $z = 0$ . We define the partition function

$$Z_n(\alpha|\chi) = \sum_v c_n(v|\chi) e^{\alpha v} \quad (3.1)$$

and the reduced free energy

$$\kappa_n(\alpha|\chi) = n^{-1} \log Z_n(\alpha|\chi). \quad (3.2)$$

We define a *loop* to be a positive walk which satisfies the inequalities

$$0 = x_0 < x_i \leq x_n \quad 0 < i \leq n \quad (3.3)$$

and the condition

$$0 = z_0 = z_n \leq z_i \quad 0 < i \leq n. \quad (3.4)$$

We write  $l_n(v|\chi)$  for the number of loops with  $n$  edges and colouring  $\chi$ , having  $v$  vertices coloured  $A$  in the plane  $z = 0$ . Define the partition function

$$L_n(\alpha|\chi) = \sum_v l_n(v|\chi) e^{\alpha v}. \quad (3.5)$$

Let  $c_n^h(v|\chi)$  be the number of  $n$ -edge self-avoiding walks with  $z_0 = h$ , no vertices with negative  $z$ -coordinate, colouring  $\chi$ , and having  $v$  vertices coloured  $A$  in  $z = 0$ . Define

$$Z_n^*(\alpha|\chi) = \max_h \sum_v c_n^h(v|\chi)e^{\alpha v}. \tag{3.6}$$

Orlandini *et al* (1999) proved the existence of the quenched average free energy

$$\lim_{n \rightarrow \infty} \langle \kappa_n(\alpha|\chi) \rangle \equiv \bar{\kappa}(\alpha) \tag{3.7}$$

where the angular brackets denote an average over colourings  $\chi$ , and that

$$\lim_{n \rightarrow \infty} \langle n^{-1} \log L_n(\alpha|\chi) \rangle = \lim_{n \rightarrow \infty} \langle n^{-1} \log Z_n^*(\alpha|\chi) \rangle = \bar{\kappa}(\alpha). \tag{3.8}$$

**Lemma 2.** *If  $\chi_0$  is a number normal in base 2, and  $\chi(1)\chi(2) \dots \chi(p)$  is the first  $mp$  digits of  $\chi_0$ , with each  $\chi(i)$  having  $m$  digits, then*

$$\lim_{p \rightarrow \infty} p^{-1} \sum_{i=1}^p m^{-1} \log L_m(\alpha|\chi(i)) = \langle m^{-1} \log L_m(\alpha|\chi) \rangle \tag{3.9}$$

where the angular brackets denote an average over the possible colourings  $\chi$  of length  $m$ . Similarly,

$$\lim_{p \rightarrow \infty} p^{-1} \sum_{i=1}^p m^{-1} \log Z_m^*(\alpha|\chi(i)) = \langle m^{-1} \log Z_m^*(\alpha|\chi) \rangle. \tag{3.10}$$

**Proof.** Since  $\chi_0$  is normal in base 2 it is simply normal in base  $2^m$ . Therefore each of the possible colourings of length  $m$  will occur with frequency  $2^{-m}$  in the set of  $p$  colourings  $\chi(1), \chi(2), \dots, \chi(p)$ , as  $p \rightarrow \infty$ .  $\square$

We now turn to the main theorem of this paper.

**Theorem 1.** *If  $\chi_0$  is a normal number in base 2, then*

$$\lim_{n \rightarrow \infty} n^{-1} \log Z_n(\alpha|\chi_0) \equiv \kappa(\alpha|\chi_0) \tag{3.11}$$

exists for all  $\alpha < \infty$ , and

$$\kappa(\alpha|\chi_0) = \bar{\kappa}(\alpha). \tag{3.12}$$

**Proof.** Fix  $\alpha < \infty$ . For fixed  $m$  write  $n = mp + q$  where  $0 \leq q < m$ . Divide the colouring  $\chi_0$  into  $p$  colourings of length  $m$ ,  $\chi^{(1)}, \chi^{(2)}, \dots, \chi^{(p)}$  and a colouring of length  $q$ ,  $\chi^{(p+1)}$ . By concatenating a set of  $p$  loops each with  $m$  edges and a loop with  $q$  edges, as in Orlandini *et al* (1999), lemma 3.5, we have the inequality

$$Z_n(\alpha|\chi_0) \geq \left[ \prod_{i=1}^p L_m(\alpha|\chi^{(i)}) \right] L_q(\alpha|\chi^{(p+1)}). \tag{3.13}$$

Similarly, by a concatenation argument similar to that used in Orlandini *et al* (1999), lemma 3.6, we obtain the upper bound

$$Z_n(\alpha|\chi_0) \leq Z_m(\alpha|\chi^{(1)}) \left[ \prod_{i=2}^p Z_m^*(\alpha|\chi^{(i)}) \right] Z_q^*(\alpha|\chi^{(p+1)}). \tag{3.14}$$

Hence

$$\frac{1}{n} \log Z_n(\alpha|\chi_0) \geq \frac{1}{p + \frac{q}{m}} \sum_{i=1}^p m^{-1} \log L_m(\alpha|\chi^{(i)}) + \frac{1}{n} \log L_q(\alpha|\chi^{(p+1)}) \tag{3.15}$$

and

$$\frac{1}{n} \log Z_n(\alpha|\chi_0) \leq \frac{1}{p + \frac{q}{m}} \sum_{i=1}^p m^{-1} \log Z_m^*(\alpha|\chi^{(i)}) + \frac{1}{n} \log Z_q^*(\alpha|\chi^{(p+1)}) \quad (3.16)$$

where we have made use of the fact that  $Z_m(\alpha|\chi^{(1)}) \leq Z_m^*(\alpha|\chi^{(1)})$ . Letting  $p \rightarrow \infty$  with  $m$  fixed, and using lemma 2, we see that

$$\begin{aligned} \langle m^{-1} \log L_m(\alpha|\chi) \rangle &\leq \liminf_{n \rightarrow \infty} n^{-1} \log Z_n(\alpha|\chi_0) \\ &\leq \limsup_{n \rightarrow \infty} n^{-1} \log Z_n(\alpha|\chi_0) \\ &\leq \langle m^{-1} \log Z_m^*(\alpha|\chi) \rangle. \end{aligned} \quad (3.17)$$

Letting  $m \rightarrow \infty$  and using the result of Orlandini *et al* (1999) in equation (3.8), we have

$$\lim_{n \rightarrow \infty} n^{-1} \log Z_n(\alpha|\chi_0) = \bar{\kappa}(\alpha) \quad (3.18)$$

whenever  $\chi_0$  is normal in base 2.  $\square$

**Corollary 1.** *Since almost all numbers are normal, the system is thermodynamically self-averaging.*

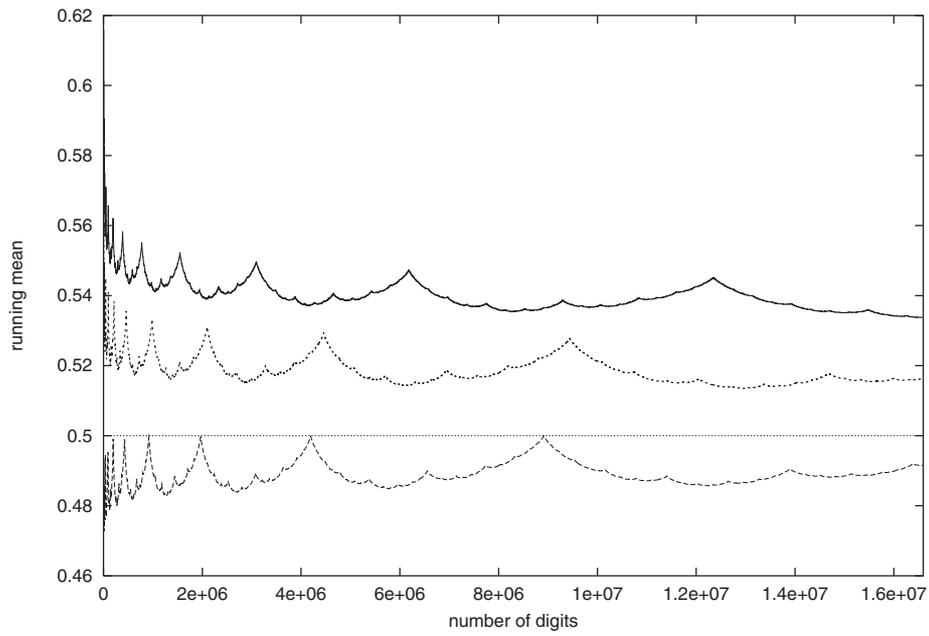
#### 4. Discussion

The primary result of this paper is that any colouring corresponding to a number which is normal in base 2 gives a limiting free energy which is equal to the limiting quenched average free energy for a randomly coloured self-avoiding walk model of copolymer adsorption. This result gives a constructive proof that this model of copolymer adsorption is thermodynamically self-averaging. Previous proofs of self-averaging in this and in related models have been existence proofs.

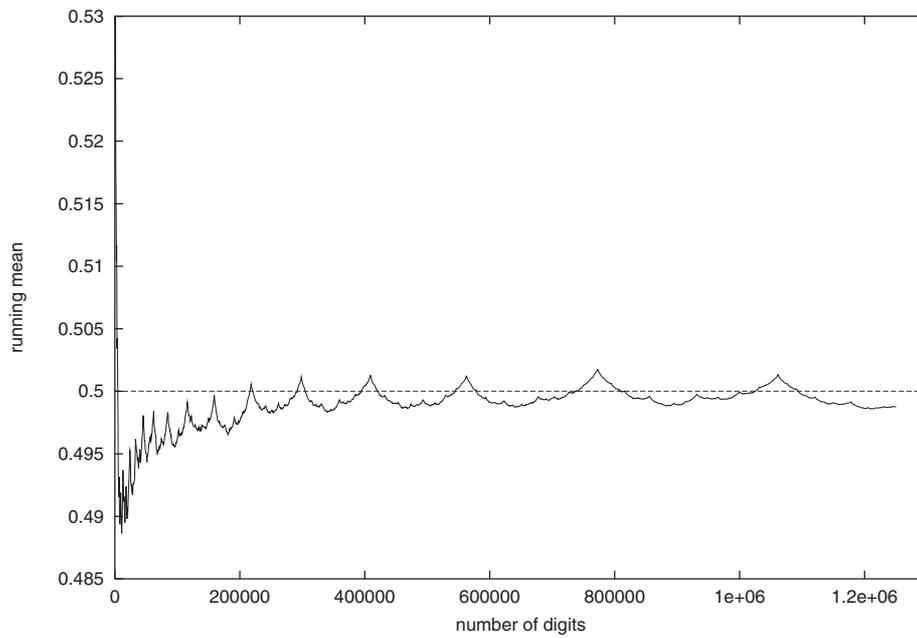
In principle this gives a means of calculating the quenched average free energy from results for a single colouring, but the convergence to the limiting (infinite  $n$ ) quenched average free energy may be very slow. In figure 1 we show the running means of the digits (i.e. the fraction of 1's) of several numbers which are known to be normal in base 2 (Champernowne 1933, Copeland and Erdős 1946 and Szűs and Volkmann 1994). It is evident that the rates of convergence of the running means of all these numbers are slow and erratic. By comparison, in figure 2 we plot the running mean of the digits of the normal number in base 2 constructed from  $[p_{p_n}]^2$ , where  $p_n$  is the  $n$ th prime (see Szűs and Volkmann (1994)). This shows more rapid and less erratic convergence. These results demonstrate that care must be taken in the selection of a normal number for the purposes of simulations. It is an open question which normal numbers are most appropriate for any given model.

The proof of theorem 1 relies on deriving upper and lower bounds for the free energy with a specified colouring, and then using a squeeze theorem. It is also directly applicable to a model of linear polymer localization at an interface, and to directed versions of these models. The model of localizing linear polymers has already been shown to be thermodynamically self-averaging (Martin *et al* 2000), but our methods allow us to exhibit self-averaging sequences. Again, any sequence which is normal in base 2 will give rise to self-averaging in the localization model.

The same approach can be used for problems in which only a lower bound is known and an ergodic theorem for super-additive processes (Kingman 1973, Akcoglu and Krengel 1981) is applied (Orlandini *et al* 2000, You and Janse van Rensburg 2000, Janse van Rensburg *et al*



**Figure 1.** The running mean plotted as a function of the number of digits in the sequence for (a) the Champernowne number given in (2.1) (top curve), (b) the normal number in base 2 derived from the sequence of primes  $p_1, p_2, \dots$  (centre curve) and (c) the number  $S_1, S_2, \dots, S_r, \dots$  (bottom curve).



**Figure 2.** The running mean plotted as a function of the number of digits in the sequence for the base 2 number derived from the sequence  $[p_{p_n}]^2$ .

2001). However, in these cases we have only been able to show that almost all numbers normal in base 2 lead to self-averaging, so we are unable to exhibit a self-averaging sequence for these cases.

### Acknowledgment

This research was supported, in part, by NSERC of Canada.

### References

- Akçoglu M A and Krengel U 1981 *J. Reine Angew. Math.* **323** 53–67  
Borel E 1909 *Rend. Circ. Mat. Palermo* **27** 247–71  
Brout R 1959 *Phys. Rev.* **115** 824–35  
Champernowne D G 1933 *J. Lond. Math. Soc.* **8** 254–60  
Copeland A H and Erdős P 1946 *Bull. Am. Math. Soc.* **52** 857–60  
Derrida B and Hilhorst H 1981 *J. Phys. C: Solid State Phys.* **14** L539–44  
van Enter A C D and van Hemmen J L 1983 *J. Stat. Phys.* 141–52  
Hardy G H and Wright E M 1938 *An Introduction to the Theory of Numbers* (Oxford: Oxford University Press)  
van Hemmen J L and Palmer R G 1982 *J. Phys. A: Math. Gen.* **15** 3881–90  
Janse van Rensburg, Orlandini E, Tesi M C and Whittington S G 2001 *J. Phys. A: Math. Gen.* **34** L37–44  
Kingman J F C 1973 *Ann. Prob.* **1** 883–909  
Kuipers L and Niederreiter H 1974 *Uniform Distribution of Sequences* (New York: Wiley)  
Martin R, Causo M S and Whittington S G 2000 *J. Phys. A: Math. Gen.* **33** 7903–18  
Orlandini E, Tesi M C and Whittington S G 1999 *J. Phys. A: Math. Gen.* **32** 469–77  
——— 2000 *J. Phys. A: Math. Gen.* **33** 259–66  
Sourlas N 1987 *Europhys. Lett.* **3** 1007–11  
Szűs P and Volkmann B 1994 *Forum Math.* **6** 399–414  
You S and Janse van Rensburg E J 2000 *J. Phys. A: Math. Gen.* **33** 1171–86