

concentrate intorno ai valori della magnetizzazione spontanea $\pm m_0$, da cui $\langle m^2 \rangle \simeq m_0^2$, $\langle m^4 \rangle \simeq m_0^4$ e $B(T < T_c, L) = 1$. Le distribuzioni di probabilità della magnetizzazione del sistema, insieme con il valore asintotico $B(T, \infty)$ sono mostrate nella Figura 20.4(b).

Ogni volta che si deve studiare una transizione di fase è molto importante saperne individuare la posizione con precisione. In tutti questi casi è fondamentale trovare un parametro analogo a quello di Binder per il modello di Ising. Ad esempio, nello studio della percolazione, un buon parametro è la probabilità che una configurazione di densità ρ sia percolante: questa probabilità varia tra 0 e 1 all'aumentare di ρ e le curve per diversi valori della taglia del reticolo si incrociano al punto critico percolativo. Nel Paragrafo 20.3.4 analizziamo meglio l'importanza di un parametro di questo tipo nello studio della regione critica intorno al punto di transizione.

Laboratorio 20.3 *Modello di Ising in 3 dimensioni*



Ripetete per il modello di Ising in tre dimensioni spaziali (usando un reticolo cubico con condizioni periodiche al bordo) tutti gli studi che abbiamo fatto in questo paragrafo sul modello di Ising bidimensionale. In particolare cercate di individuare il punto critico.

20.3 Analisi dei dati

Nel Paragrafo 20.2.2 studiamo il problema delle correlazioni temporali nelle misure di un'osservabile e come queste riducano la precisione con cui possiamo conoscerne la media. Purtroppo la procedura studiata nel Paragrafo 20.2.2 per stimare il corretto errore statistico è laboriosa: richiede prima di calcolare la funzione di correlazione e quindi di stimare il tempo di correlazione con un'interpolazione o un integrale. Un metodo alternativo, e generalmente più efficace, è la cosiddetta *media a blocchi*.

20.3.1 Media a blocchi

Consideriamo una serie temporale di \mathcal{T} misure $\{A_t\}_{t=0, \dots, \mathcal{T}-1}$ (osservate che usiamo gli indici come in un array del linguaggio C, cominciando da 0) con la quale vogliamo stimare la vera media del processo che l'ha generata. È ben noto che il miglior stimatore della vera media è il valor medio dei dati

$$\langle\langle A \rangle\rangle \equiv \frac{1}{\mathcal{T}} \sum_{t=0}^{\mathcal{T}-1} A_t, \quad \text{con errore} \quad \sigma_{\langle\langle A \rangle\rangle} = \sqrt{\frac{\text{var}(A_t)}{\mathcal{T}-1}}. \quad (20.11)$$

L'espressione per l'indeterminazione sulla stima della media $\sigma_{\langle\langle A \rangle\rangle}$ è corretta solo se le \mathcal{T} misure sono tra di loro completamente scorrelate. Con il simbolo $\text{var}(A_t)$ indichiamo la

varianza della serie di dati definita da

$$\text{var}(A_t) \equiv \frac{1}{\mathcal{T}} \sum_{t=0}^{\mathcal{T}-1} (A_t - \langle\langle A \rangle\rangle)^2 = \frac{1}{\mathcal{T}} \sum_{t=0}^{\mathcal{T}-1} A_t^2 - \langle\langle A \rangle\rangle^2 = \langle\langle A^2 \rangle\rangle - \langle\langle A \rangle\rangle^2 .$$

Nel caso in cui le misure siano invece correlate l'espressione (20.11) sottostima l'incertezza statistica nella determinazione della media ed è necessaria un'analisi piú raffinata.

Supponiamo che le misure siano correlate con un tempo di correlazione τ_A e mostriamo come si può procedere per una corretta stima dell'errore sulla media. Raggruppiamo le misure in blocchi contenenti ciascuno b misure consecutive e chiamiamo $A_t^{(b)}$ la media delle misure nel t -esimo blocco,

$$A_t^{(b)} = \frac{1}{b} \sum_{t'=bt}^{b(t+1)-1} A_{t'} .$$

Il numero dei blocchi e quindi la lunghezza della nuova serie di dati $\{A_t^{(b)}\}$ è \mathcal{T}/b . In particolare notiamo che gli $A_t^{(1)}$ sono i dati originari e che il valore medio dei dati, $\langle\langle A \rangle\rangle$, non cambia quando questi sono raggruppati in blocchi di grandezza uguale (il valor medio rimane comunque la migliore stima della vera media).

Per $b < \tau_A$, i dati della serie $\{A_t^{(b)}\}$ sono ancora correlati (con un tempo di correlazione circa pari a τ_A/b), mentre, per $b > \tau_A$, i dati diventano scorrelati, perché il grosso delle correlazioni temporali è finito dentro i blocchi. Sfruttiamo questa proprietà per stimare correttamente l'errore statistico, $\sigma_{\langle\langle A \rangle\rangle}$, sulla stima della media. Definiamo

$$\sigma_{\langle\langle A \rangle\rangle}^{(b)} \equiv \sqrt{\frac{\text{var}(A_t^{(b)})}{(\mathcal{T}/b) - 1}} ,$$

dove la varianza della serie $\{A_t^{(b)}\}$ è definita come prima

$$\text{var}(A_t^{(b)}) \equiv \frac{b}{\mathcal{T}} \sum_{t=0}^{\mathcal{T}/b-1} (A_t^{(b)})^2 - \langle\langle A \rangle\rangle^2 .$$

La $\sigma_{\langle\langle A \rangle\rangle}^{(b)}$ corrisponde all'indeterminazione sulla media che si ottiene assumendo scorrelati i dati raggruppati in blocchi di grandezza b . Per il discorso di prima ci aspettiamo che $\sigma_{\langle\langle A \rangle\rangle}^{(b)}$ sottostimi l'errore sulla media per $b < \tau_A$ e fornisca, invece, la giusta indeterminazione per $b > \tau_A$.

Nella Figura 20.5(a) mostriamo le indeterminazioni sulla stima delle medie $\langle e \rangle_T$ (in basso) e $\langle m \rangle_T$ (in alto) per un modello di Ising bidimensionale di lato $L = 20$ a temperatura $T = 2.4$. I dati sono stati ottenuti da una serie di 10^6 misure fatte all'equilibrio. Notiamo che i dati seguono bene l'andamento teorico previsto: crescono fino a un certo

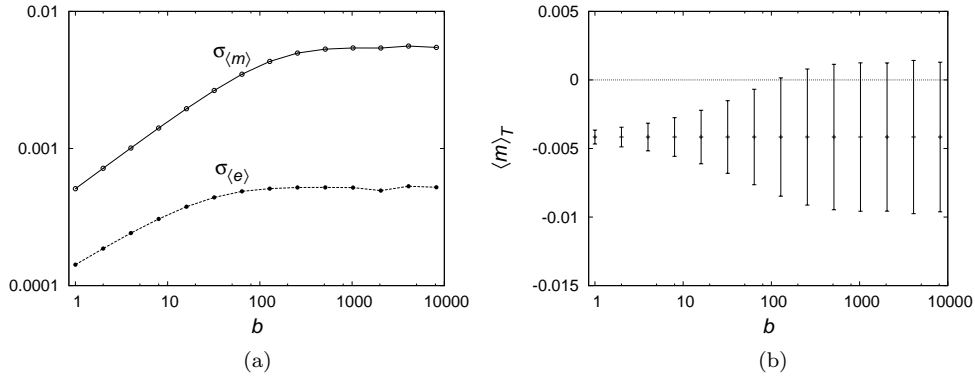


Figura 20.5 Indeterminazione sulla stima di $\langle e \rangle_T$ e $\langle m \rangle_T$ (a) e la magnetizzazione media con il suo errore stimato (b) in funzione del numero b di misure per blocco.

valore di b , corrispondente al tempo di correlazione, e quindi si stabilizzano. Il valore del *plateau* è il giusto errore da assegnare alla misura della media. Osserviamo anche che i dati riportati nella Figura 20.5 presentano fluttuazioni, dato che le varianze $\sigma^{(b)}$ sono variabili aleatorie.

Avendo a disposizione 10^6 dati avremmo potuto raggrupparli con valori di b maggiori di quelli mostrati nella Figura 20.5, ma ciò avrebbe potuto rendere l'analisi meno chiara. Infatti quando il numero di blocchi diventa piccolo il valore di $\sigma_{\langle A \rangle}^{(b)}$ può fluttuare molto; consigliamo di usare un numero minimo di blocchi dell'ordine di un centinaio.

I dati riportati nella Figura 20.5(b) dovrebbero convincerci di quanto sia importante una buona stima degli errori. Nella figura è riportato, per diversi valori della grandezza b dei blocchi, il valore medio della magnetizzazione con il suo errore (il modello è sempre quello di Ising bidimensionale con $L = 20$ e $T = 2.4$). L'osservazione importante è che, se non avessimo stimato correttamente l'errore statistico, cioè se avessimo usato la media corrispondente a $b = 1$, avremmo dovuto concludere che il valor medio della magnetizzazione non era compatibile con zero (risultato sbagliato che ci dovrebbe far ripensare l'intera simulazione numerica!). Invece, una volta stimato il corretto errore, la magnetizzazione media è compatibile con zero, come dev'essere.

I dati riportati nella Figura 20.5 sono stati ottenuti con la funzione riportata nel Listato 20.1, in cui si calcola la media e la varianza della serie di dati per ottenere $\sigma_{\langle A \rangle}^{(b)}$ (righe 6-8) e contemporaneamente si genera la prossima serie di dati raggruppando quelli attuali a due a due (riga 9). La funzione `binning` riceve in ingresso il puntatore `data` all'inizio dell'array contenente i `numMeas` dati di cui vogliamo determinare la media a blocchi. Considerando che la funzione scrive le medie così ottenute nello stesso array (più precisamente nella sua prima metà), è importante passare un puntatore a una *copia* dei dati se non vogliamo perdere quelli originali.

ESTRATTO DAL LIBRO **PROGRAMMAZIONE SCIENTIFICA** DI
L. M. BARONE, E. MARINARI, G. ORGANTINI E F. RICCI-TERSENGHI

I diritti di riproduzione e memorizzazione elettronica degli estratti appartengono a Pearson Education Italia S.r.l. e sono riservati per tutti i paesi. La stampa di essi è concessa gratuitamente solo a titolo personale. Ogni altro uso è vietato.

```

1 double binning(double *data, int numMeas) {
2   int i, tmp = numMeas / 2;
3   double mean = 0.0, variance = 0.0;
4
5   for (i = 0; i < tmp; i++) {
6     mean += data[2 * i] + data[2 * i + 1];
7     variance += data[2 * i] * data[2 * i] +
8               data[2 * i + 1] * data[2 * i + 1];
9     data[i] = 0.5 * (data[2 * i] + data[2 * i + 1]);
10  }
11  if (2 * i < numMeas) {
12    mean += data[2 * i];
13    variance += data[2 * i] * data[2 * i];
14  }
15  mean /= numMeas; variance /= numMeas;
16  return sqrt((variance - mean * mean) / (numMeas - 1));
17 }

```

Listato 20.1 La funzione che esegue il raggruppamento dei dati per calcolare il giusto errore sulla media.

Le istruzioni alle righe 11-14 sono eseguite solo quando `numMeas` è dispari, perché in questo caso il ciclo alle righe 5-10 scorre su tutti i dati tranne l'ultimo. In questo caso, nella nuova serie perdiamo un dato, ma ciò non è assolutamente un problema, visto che generalmente si parte da serie di dati molto lunghe.

Le righe di codice che seguono illustrano un possibile uso della funzione `binning`.

```

binSize = 1;
while (numMeas >= 100) {
  printf("%i %lg %lg %lg %lg %i\n",
         binSize, aveM, binning(m, numMeas),
         aveE, binning(e, numMeas), numMeas);
  binSize *= 2;
  numMeas /= 2;
}

```

Prima dell'esecuzione di queste righe di codice gli array `m` ed `e` sono stati riempiti con `numMeas` valori di magnetizzazione ed energia; le variabili `aveM` e `aveE` contengono le medie di queste due serie di dati. La variabile `binSize` è la grandezza del blocco (*bin* in inglese) e `numMeas` rappresenta la lunghezza dell'attuale serie: dopo ogni chiamata alla funzione `binning` queste variabili sono aggiornate, rispettivamente moltiplicandole e dividendole per 2. Il ciclo in queste righe di codice è eseguito finché il numero di dati nella serie è non minore di 100, affinché la misura della varianza sia stabile. Provando a diminuire questo limite, ci si convince facilmente che non si guadagna nessuna informazione in più perché le misure di $\sigma_{\langle A \rangle}^{(b)}$ cominciano a fluttuare molto.

20.3.2 Jackknife

Con il metodo descritto nel Paragrafo 20.3.1 possiamo sempre produrre serie di dati praticamente scorrelati a partire da una serie di misure all'equilibrio. Ad esempio, possiamo determinare le medie delle 10^6 misure della magnetizzazione usate per produrre la Figura 20.5 a blocchi di 10^3 (è quello il valore di b dove inizia il plateau), ottenendo così 10^3 dati scorrelati. Per le misure dell'energia possiamo anche usare un valore minore di b , vedi Figura 20.5(a), e ottenere una serie più lunga di dati scorrelati. Quindi d'ora in avanti consideriamo solo serie di dati scorrelati.

Rimane da risolvere un problema molto comune nell'analisi dei dati, quello della stima corretta dell'errore per la media di funzioni complicate delle misure. Un esempio concreto, connesso con quello che abbiamo visto nel Paragrafo 20.2.3, è quello della stima dell'incertezza nella determinazione del parametro di Binder definito nell'equazione (20.10). Possiamo valutare facilmente l'errore sulla stima di $\langle m^2 \rangle_T$ e $\langle m^4 \rangle_T$ con il metodo illustrato nel Paragrafo 20.3.1 usando come serie di dati $A_t = m_t^2$ e $A_t = m_t^4$ rispettivamente. Una volta note le incertezze $\sigma_{\langle m^2 \rangle}$ e $\sigma_{\langle m^4 \rangle}$, potremmo ottenere quella sulla stima di $\langle B(T, L) \rangle_T$ con la propagazione degli errori, ma ciò darebbe un risultato che sovrastima abbondantemente il vero errore. Il motivo è semplice e si capisce osservando la definizione (20.10) di $B(T, L)$, in cui i momenti (secondo e quarto) della distribuzione della magnetizzazione compaiono solo con il loro rapporto: se la stima del numeratore $\langle m^4 \rangle$ ha una fluttuazione che la porta sopra la vera media, anche la stima del denominatore $\langle m^2 \rangle^2$ ne ha una positiva e il rapporto non cambia molto. Quindi le fluttuazioni della stima del rapporto sono molto più piccole rispetto a quelle previste dalla semplice propagazione degli errori, che assume scorrelate le fluttuazioni al numeratore e al denominatore. Illustriamo un metodo che permette di calcolare la giusta indeterminazione anche in presenza di forti correlazioni come quelle appena viste.

Il metodo si chiama *jackknife* ed è così utile che andrebbe tenuto sempre a portata di mano quando si fa l'analisi dei dati. La traduzione letterale del nome è "coltello a serramanico", perché anche questo è un oggetto utile e da tenere a portata di mano (ad esempio in campeggio). Il metodo jackknife permette di calcolare la giusta indeterminazione sulla stima del valor medio di una funzione qualsiasi dei dati. Ad esempio, data una serie di \mathcal{T} misure scorrelate della magnetizzazione, la stima del valore medio del parametro di Binder, definito nella (20.10), è dato dalla funzione

$$B(\{m_t\}) = \frac{1}{2} \left(3 - \frac{\mathcal{T} \sum_t m_t^4}{(\sum_t m_t^2)^2} \right),$$

ma non è ovvio quale errore dobbiamo associare a questa stima.

Il metodo jackknife si basa sulla seguente idea. Supponiamo di avere a disposizione tutte le misure tranne la i -esima. Con queste $\mathcal{T} - 1$ misure possiamo ancora stimare il valore medio della funzione. Tipicamente questa stima è leggermente diversa da quella ottenuta con tutte le \mathcal{T} misure e la differenza tra queste due stime ci dà proprio l'informazione con cui calcolare l'indeterminazione sulla stima della media.

Scritto in termini leggermente piú formali, se la migliore stima, μ , di una generica quantità è data dalla seguente funzione delle \mathcal{T} misure *scorrelate*¹,

$$\mu = f(\{x_t\}_{t=1,\dots,\mathcal{T}}),$$

allora, definendo le \mathcal{T} medie mancanti di una misura come

$$\mu_j = f(\{x_t\}_{t \neq j}),$$

possiamo determinare l'incertezza su μ come

$$\sigma_\mu = \sqrt{\frac{\mathcal{T}-1}{\mathcal{T}} \sum_{j=1}^{\mathcal{T}} (\mu_j - \mu)^2}. \quad (20.12)$$

Quesito 20.1 *Calcolo dell'errore sulla media col jackknife*



Nel caso della stima della media la forma del miglior stimatore è semplicemente il valor medio delle misure

$$\mu = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} x_t.$$

Dimostrate che in questo caso l'errore fornito dal metodo jackknife, dato dall'equazione (20.12), è identico a quello che si ottiene calcolando la deviazione standard delle misure, come nell'espressione (20.11).

La realizzazione di un codice per il calcolo dell'errore col metodo jackknife è immediata se si scrivono due funzioni che restituiscono rispettivamente il valore di μ e μ_j ; in realtà si può scrivere una sola funzione che, chiamata con parametro $j \in [0, \mathcal{T} - 1]$, calcola la media senza la j -esima misura e che restituisce invece la media totale se $j \notin [0, \mathcal{T} - 1]$. Il difetto di questo approccio è dato dal grande numero di operazioni necessarie: nel codice ci sarebbero $\mathcal{T} + 1$ chiamate a funzioni che eseguono $\mathcal{O}(\mathcal{T})$ operazioni ognuna. Fare un numero di operazioni $\mathcal{O}(\mathcal{T}^2)$ è talvolta molto oneroso (non è raro avere 10^8 dati e 10^{16} operazioni sono molte) e dobbiamo adottare soluzioni piú efficienti, quando sono possibili. Ad esempio, nel caso del parametro di Binder, essendo la magnetizzazione una variabile intera possiamo usare la funzione riportata nel Listato 20.2, la quale richiede solo un numero di operazioni $\mathcal{O}(2\mathcal{T})$, perché le medie μ_j si calcolano sottraendo una misura per volta dalle variabili `sumM2` e `sumM4` che contengono le somme su tutte le misure. Nel caso di variabili rappresentate in virgola mobile l'uso di espressioni quali `sumM4 - tmp * tmp`, in cui si esegue una sottrazione tra quantità molto diverse in valore assoluto, è rischioso e richiede almeno di usare l'algorithm di Kahan (descritto nel Paragrafo 4.5).

¹Nel caso la misure siano correlate è necessario raggrupparle prima.

```


1 void binder(int *M, int numMeas, double *pMean, double *pError) {
2   unsigned long long int tmp, sumM2 = 0, sumM4 = 0;
3   double reducedMean;
4   int i, *pM;
5
6   for (i = 0, pM = M; i < numMeas; i++, pM++) {
7     tmp = (*pM) * (*pM);
8     sumM2 += tmp;
9     sumM4 += tmp * tmp;
10  }
11  *pMean = 0.5 * (3.0 - (double)numMeas * sumM4 / sumM2 / sumM2);
12  *pError = 0.0;
13  for (i = 0, pM = M; i < numMeas; i++, pM++) {
14    tmp = (*pM) * (*pM);
15    reducedMean = 0.5 * (3.0 - (double)(numMeas - 1) *
16      (sumM4 - tmp * tmp) / (sumM2 - tmp) / (sumM2 - tmp));
17    *pError += (reducedMean - *pMean) * (reducedMean - *pMean);
18  }
19  *pError = sqrt(*pError * (numMeas - 1) / numMeas);
20 }

```

Listato 20.2 Una funzione che calcola l'errore sulla stima del parametro di Binder con il metodo jackknife.

La funzione `binder` riportata nel Listato 20.2 accetta come argomenti il puntatore `M` all'array contenente le misure della magnetizzazione e il numero di queste misure, `numMeas`. Gli ulteriori due argomenti, `pMean` e `pError`, sono i puntatori alle locazioni di memoria dove la funzione salva il valore del parametro di Binder e il suo errore, calcolato con il metodo jackknife, assumendo che le misure siano scorrelate.

Laboratorio 20.4 *Errore sul parametro di Binder col jackknife*

 Riconsiderate i dati della magnetizzazione che avete ottenuto nello studio del modello di Ising in 2 e/o in 3 dimensioni. Dopo averli raggruppati in blocchi della grandezza opportuna per eliminare le correlazioni, stimate il valore medio del parametro di Binder e il corrispondente errore con il metodo jackknife. Confrontate questo errore con quello che potete ottenere con la piú semplice propagazione degli errori.

20.3.3 Leggi di scala ed esponenti critici

Molto spesso una simulazione Monte Carlo è usata per trovare la risposta a problemi con un numero dato di variabili: ad esempio, l'ottimizzazione di una funzione di costo assegnata o il calcolo di un valor medio in un modello che descrive accuratamente uno

ESTRATTO DAL LIBRO **PROGRAMMAZIONE SCIENTIFICA** DI
 L. M. BARONE, E. MARINARI, G. ORGANTINI E F. RICCI-TERSENGHI
 I diritti di riproduzione e memorizzazione elettronica degli estratti appartengono a Pearson Education Italia S.r.l. e sono riservati per tutti i paesi. La stampa di essi è concessa gratuitamente solo a titolo personale. Ogni altro uso è vietato.