

Computer simulations in statistical physics

Homework 2: data analysis

Code of statistical analysis program

Analysis of data sets

set 1

set 2

set 3

set 4

set 5

set 6

Code of statistical analysis program

```
#define PROGNAME "stats"
#define VERSION "1.4"
#define DATE "06.11.2006"
#define AUTHOR "Nils Bluemer"

/* program stats (successor of statnew), computes average, error etc. */
/* expects input on stdin in 1st column */
/* compile with cc -lm -o stats stats.c */

#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#include <math.h>
#define BUFSIZE 100
#define NUMHIST 40
#define MAXDATA 100001
#define MAXKORR 1000
#define MIN(a,b) (a<b?a:b)
```

```

void histogram ()
{
    double data[MAXDATA];
    double hist[NUMHIST];
    char buf[BUFSIZE];
    int i,n,numdata;
    double min, max, x;

    numdata=0;
    do {
        if (fgets(buf,BUFSIZE,stdin)!=NULL){
            numdata++;
            sscanf(buf,"%lf",&data[numdata]);
            /*      printf("Nr %d: %f\n",nr,data[numdata]); */
        }
    } while (feof(stdin)==0);
    /*      printf ("stat/histogram: read %d lines\n",numdata); */
    if (numdata>MAXDATA) error ("too many data points");
    if (numdata<2) error ("not enough data");
    min = data[1];
    max = data[1];
}

```

```

for (i=2;i<=numdata;i++){
    if (data[i]<min) min=data[i];
    if (data[i]>max) max = data[i];
}
/*  printf ("min: %lf, max: %lf\n",min,max); */
for (i=0;i<NUMHIST;i++) hist[i]==0.0;
for (i=1;i<=numdata;i++){
    x = NUMHIST/(max-min)*(data[i]-min)*0.999999;
    hist[(int)floor(x)] += 1.0*NUMHIST/((max-min)*numdata);
/*  printf("x: %lf\n",x); */
}
for (i=0;i<NUMHIST;i++)
    printf ("%8.6lf    %8.6lf\n", (i+0.5)/NUMHIST*(max-min)+min,
hist[i]);

}

```

```

void average(char form){
    double data[MAXDATA];
    double korr[MAXDATA];
    char buf[BUFSIZE];
    int i,n,numdata,numkorr;
    double sum, sqsum,nsum,av, var,korrtime,trans;

    numdata=0;
    sum=0.0;
    sqsum=0.0;
    nsum=0;
    do {
        if (fgets(buf,BUFSIZE,stdin)!=NULL){
            numdata++;
            sscanf(buf,"%lf",&data[numdata]);
            sum +=data[numdata];
            sqsum += data[numdata]*data[numdata];
            nsum +=numdata*data[numdata];
        }
    } while (feof(stdin)==0);

```

```

if (numdata>MAXDATA) error ("too many data points");
if (numdata<2) error ("not enough data");
av=sum/numdata;
var=(sqsum-numdata*av*av)/(numdata-1);
/* compute transient: slope of best linear fit on data[n] */
/* a = (<xy>-<x><y>)/(<x^2>-<x>^2), here: x=numdata,
y=data[numdata] */
trans=(nsum/numdata-0.5*(numdata+1)*av)/
((numdata+.5)*(numdata+1)/3-0.25*(numdata+1)*(numdata+1));
if (form=='l')
    printf("Average: %10.8g   variance: %10.8g   error:
%10.8g\n", av, var, sqrt(var/numdata));

numkorr=numdata/3;

korr[0]=1.0;
if (form=='c'){
    printf("# Autocorrelation function: (i,c(i))\n");
    printf ("%4d %10lf\n",0,korr[0]);
}

```

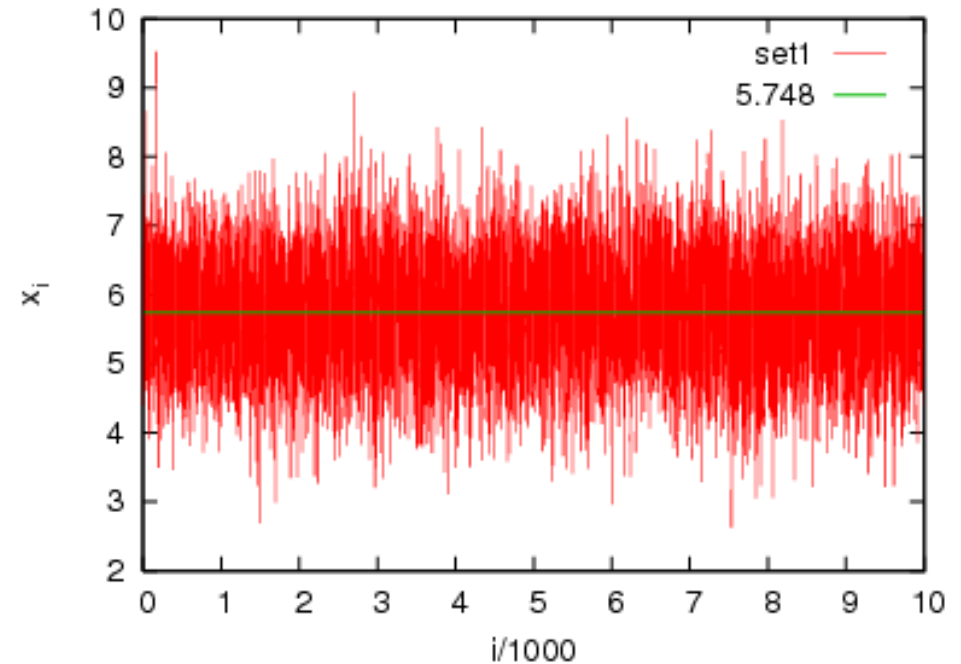
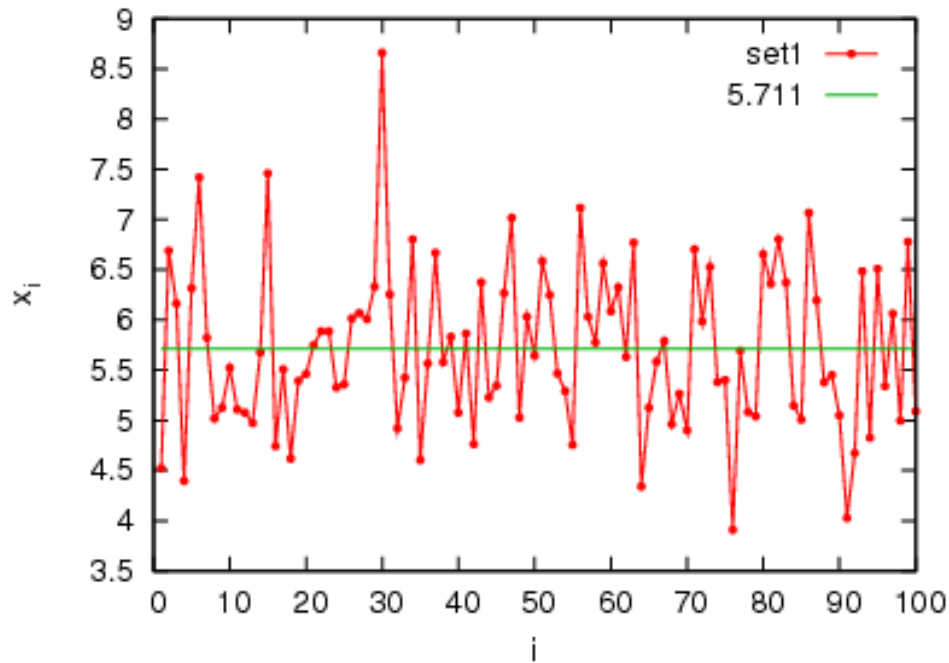
```

for (i=1;i<=numkorr;i++){
    korr[i]=0.0;
    for (n=1;n<=numdata-i;n++)
        korr[i]+=(data[n]-av)*(data[n+i]-av);
    korr[i]=(korr[i]/(var*(numdata-i))+1.0/numdata)/
        ((1-1.0/numdata)*(1-1.0/numdata));
    if (form=='c')
        printf ("%4d %10lf\n",i,korr[i]);
}
korrrtime=1;
n =1;
while ((korr[n]>0)&&(n<=numkorr)){
    korrrtime += 2*korr[n];
    n++;
}
if (form=='l')
    printf ("Korrelation time: %10.8g   corrected error: %10.8g
transient: %10.8g\n",
    korrrtime, sqrt(var/(numdata-1))*sqrt(korrrtime),trans);

```

Set 1

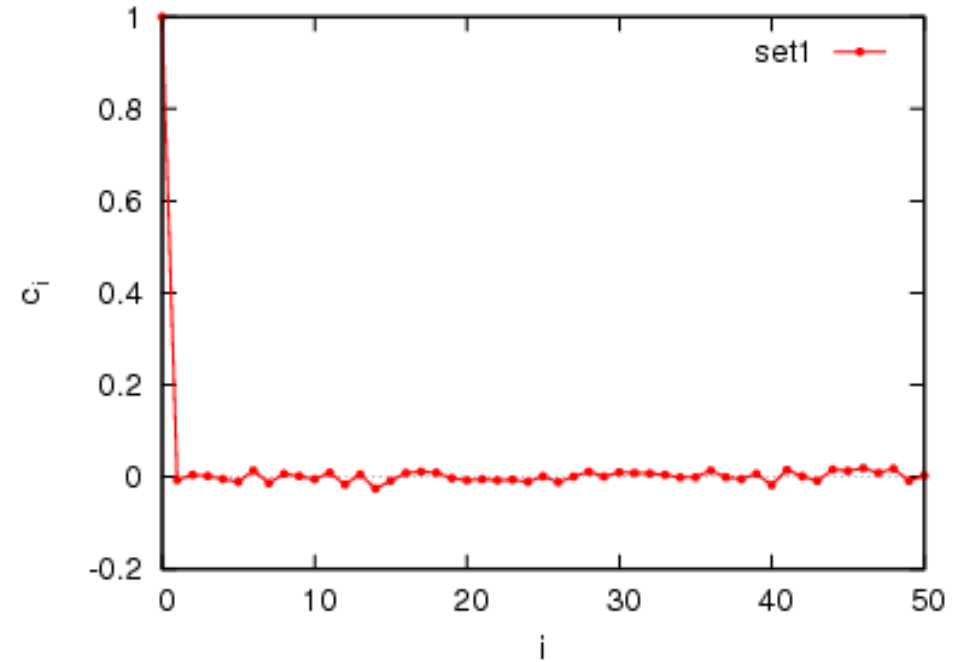
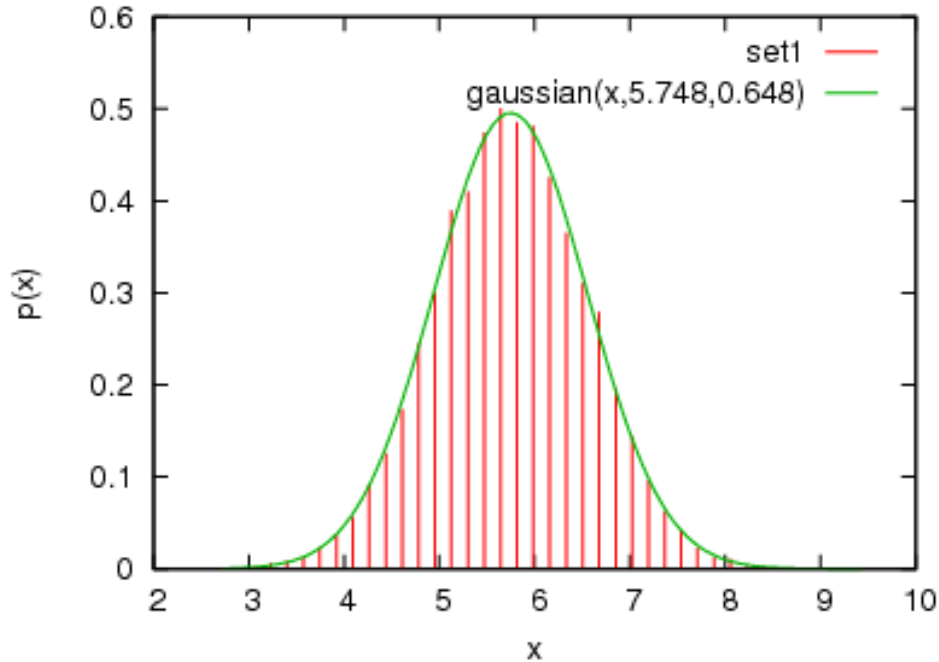
Trace and averages



Estimates of mean / initial error analysis

- $\langle x \rangle \approx 5.748 \pm 0.008$ (naive estimate)
- no transient
- autocorrelation?

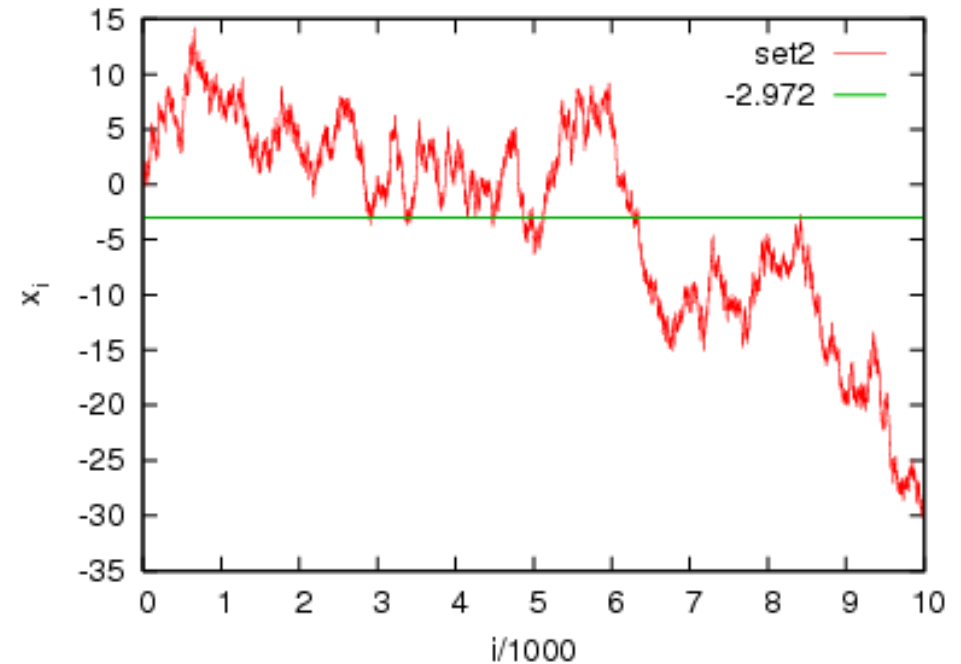
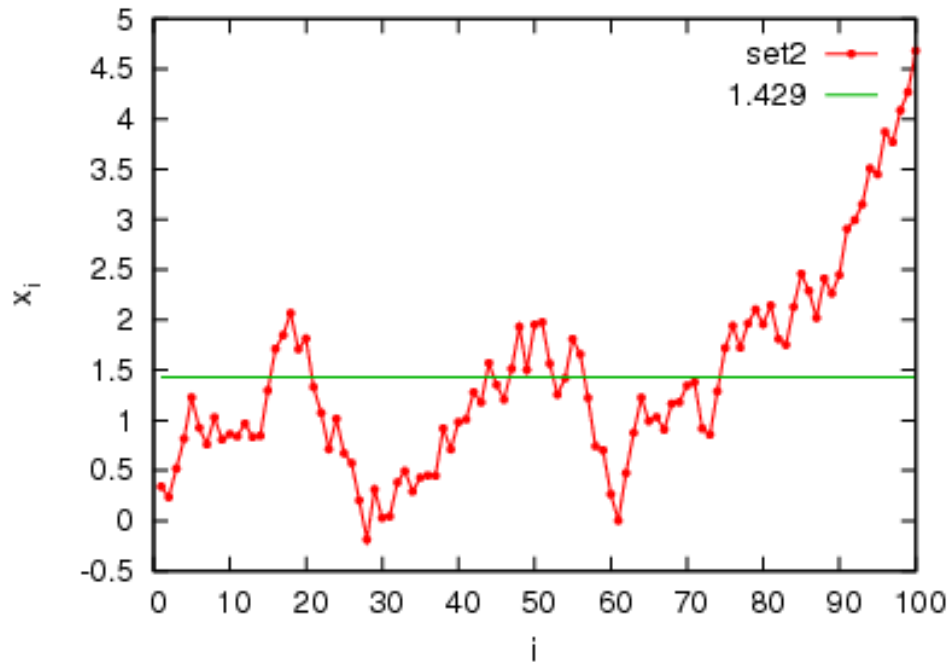
Histogram, autocorrelation function, full error estimate



- gaussian distribution of ‘‘measurements’’ - width $\sqrt{0.648}$
- no autocorrelation ($\tau = 1.0$)
- final error estimate: $\langle x \rangle \approx 5.748 \pm 0.008$
- **true distribution:** gaussian, mean 5.740, variance 0.650

Set 2

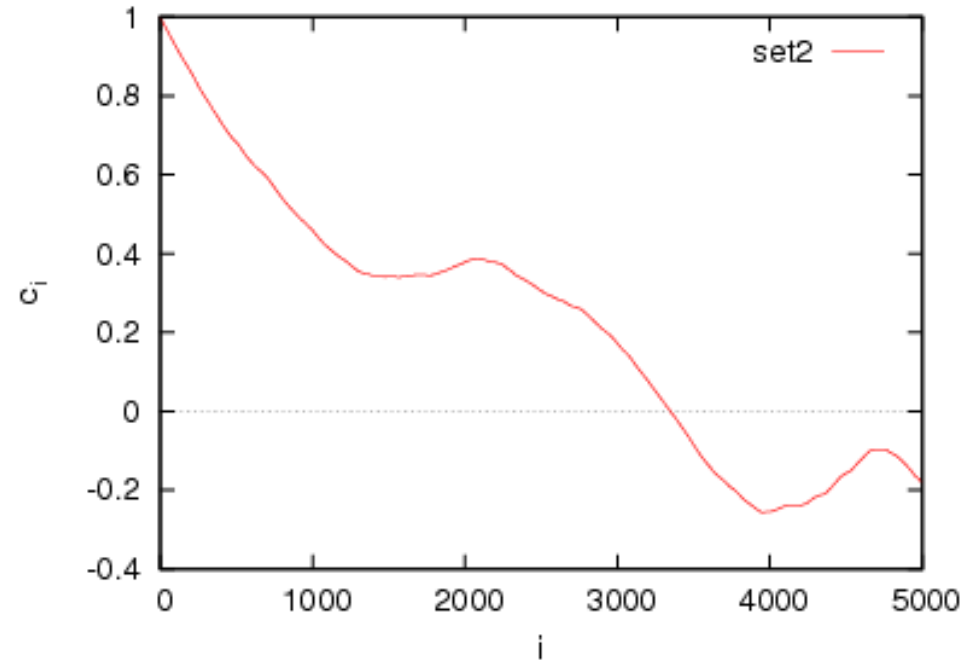
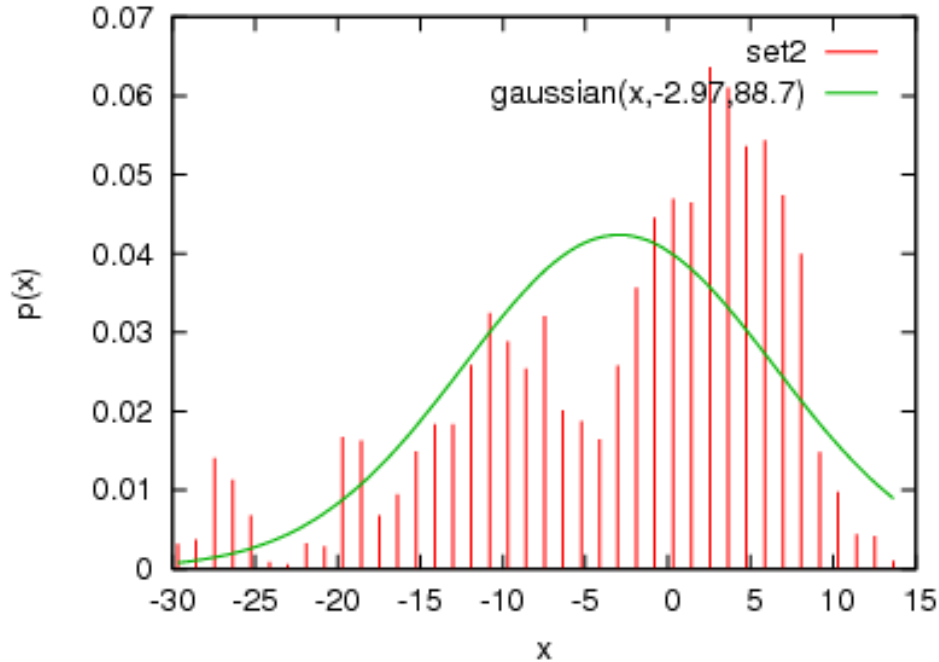
Trace and averages



Estimates of mean / initial error analysis

- $\langle x \rangle \approx -2.972 \pm 0.094$ (naive estimate)
- transient?
- long autocorrelation: of order of run length!

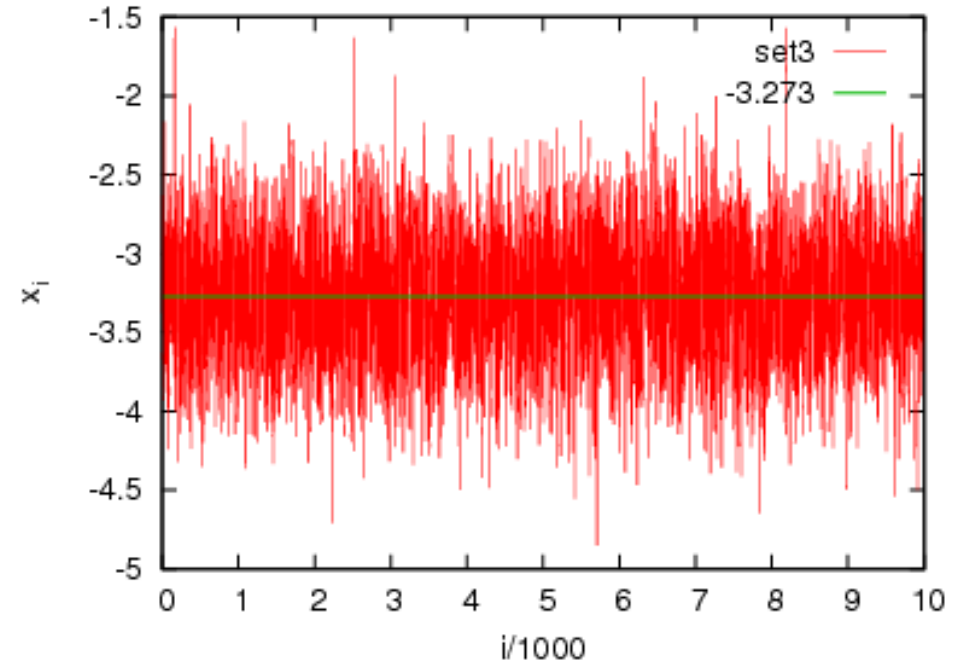
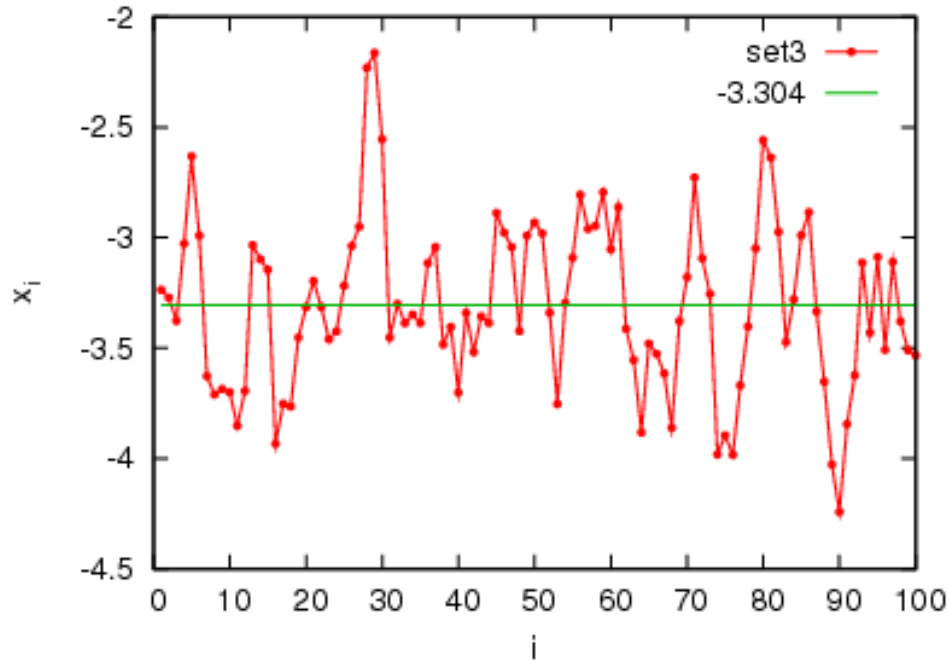
Histogram, autocorrelation function, full error estimate



- nongaussian distribution of ‘‘measurements’’
- strong autocorrelation ($\tau = 2797$ for $N = 10000$)
- final error estimate: $\langle x \rangle \approx -2.97 \pm 4.98$
- **true distribution:** gaussian (from random walk), mean 0, non-uniform variance $\sigma(X_k) = k$

Set 3

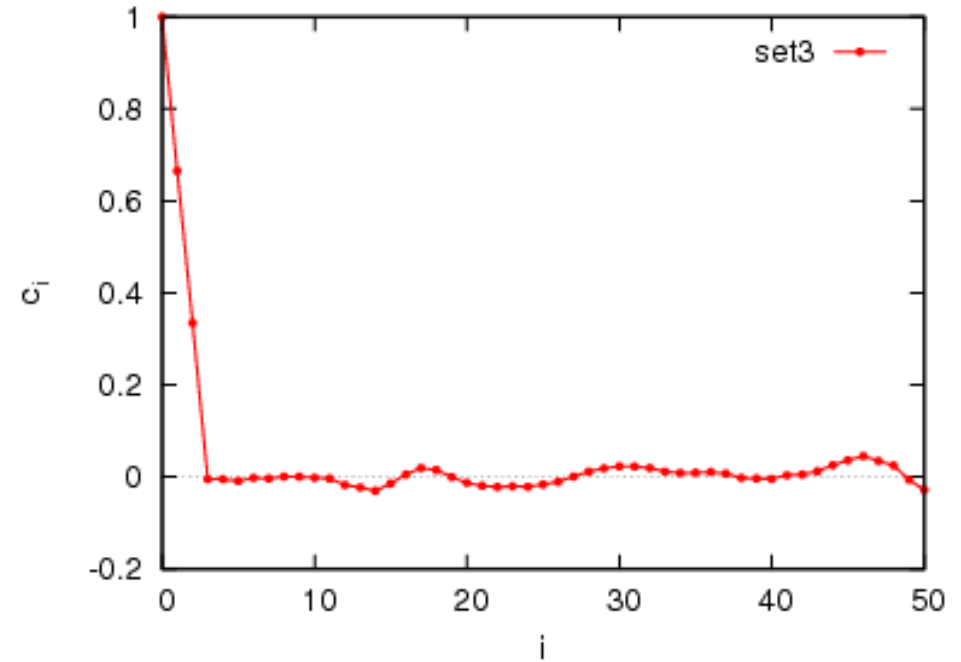
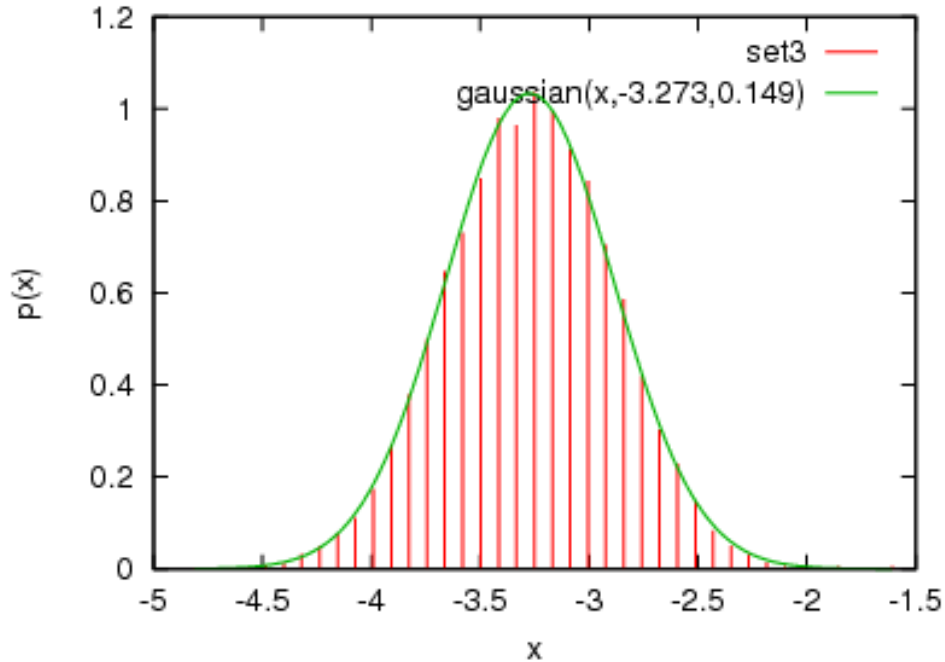
Trace and averages



Estimates of mean / initial error analysis

- $\langle x \rangle \approx -3.273 \pm 0.004$ (naive estimate)
- no transient
- short-time autocorrelation?

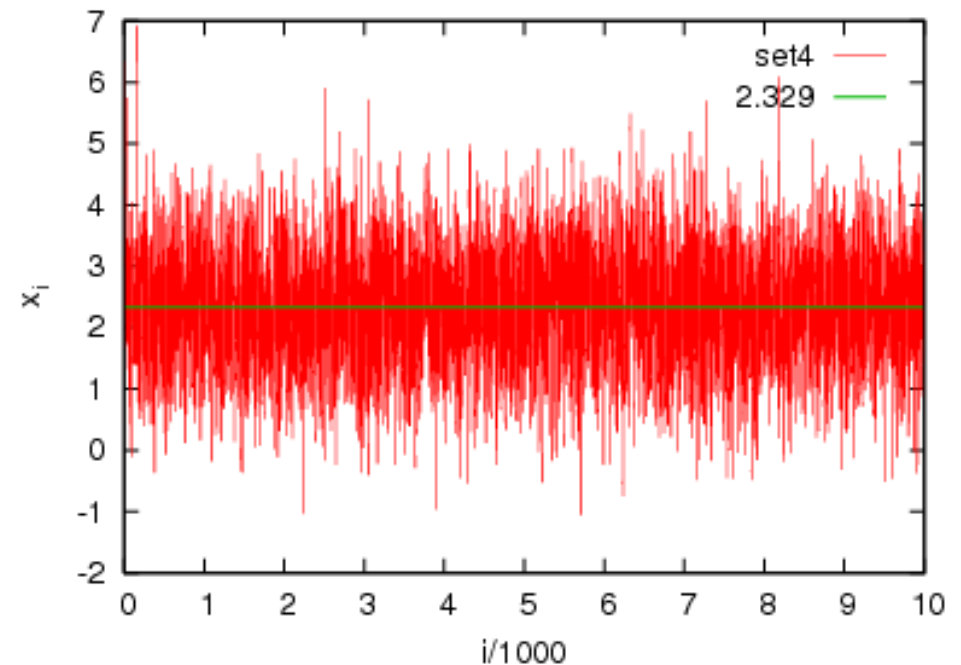
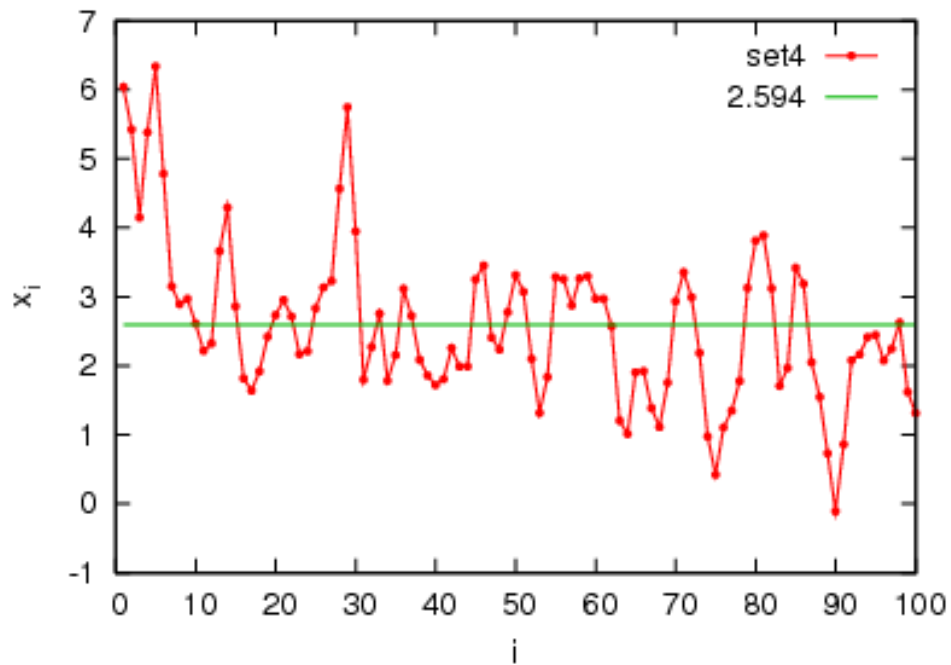
Histogram, autocorrelation function, full error estimate



- gaussian distribution of ‘‘measurements’’ - width $\sqrt{0.149}$
- finite autocorrelation ($\tau = 3.0$)
- final error estimate: $\langle x \rangle \approx -3.273 \pm 0.007$
- true distribution: gaussian, mean -3.28, variance 0.15

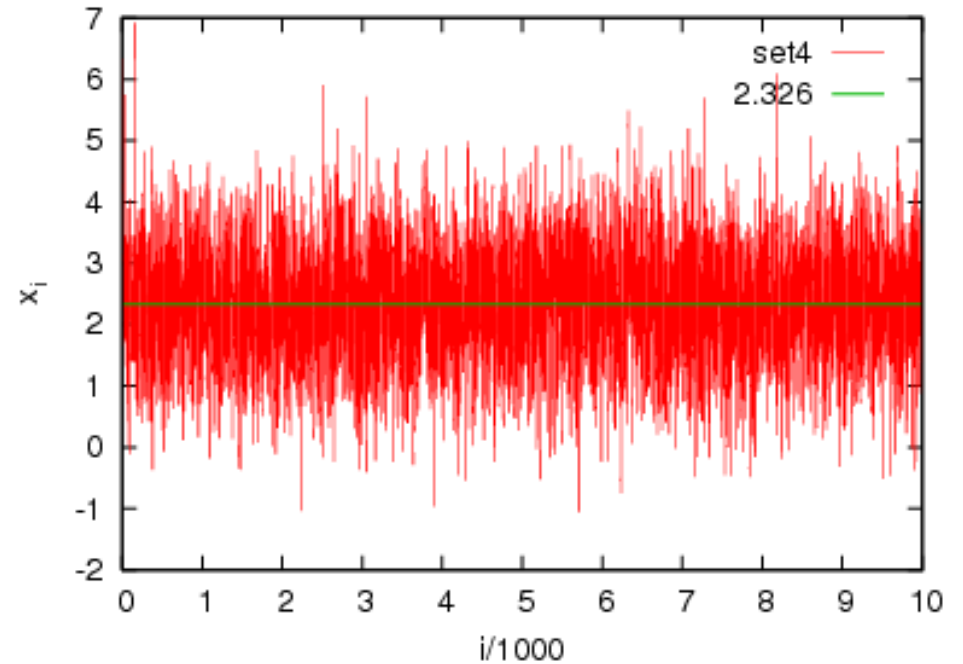
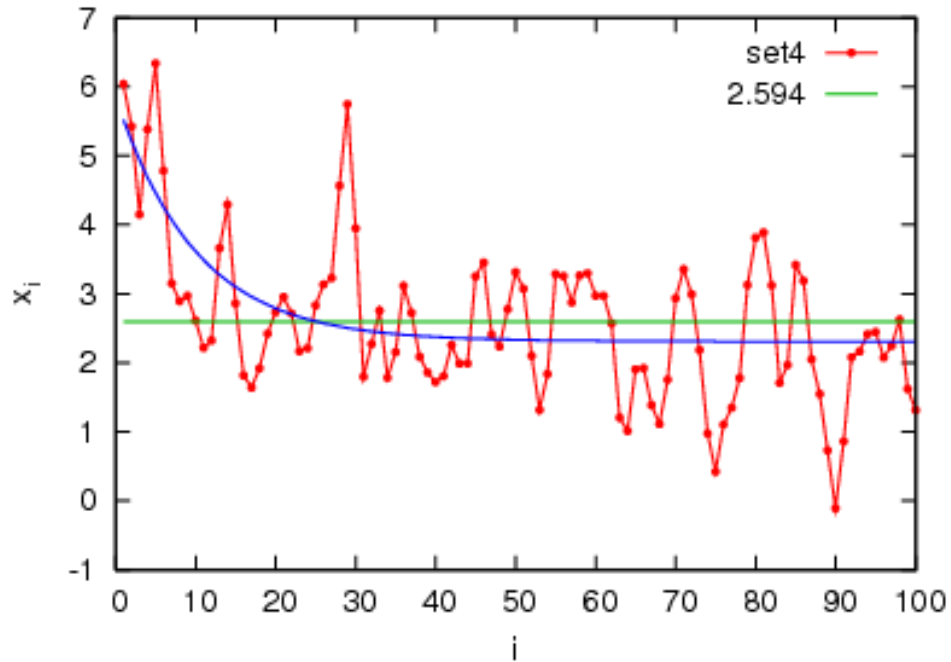
Set 4

Trace and averages



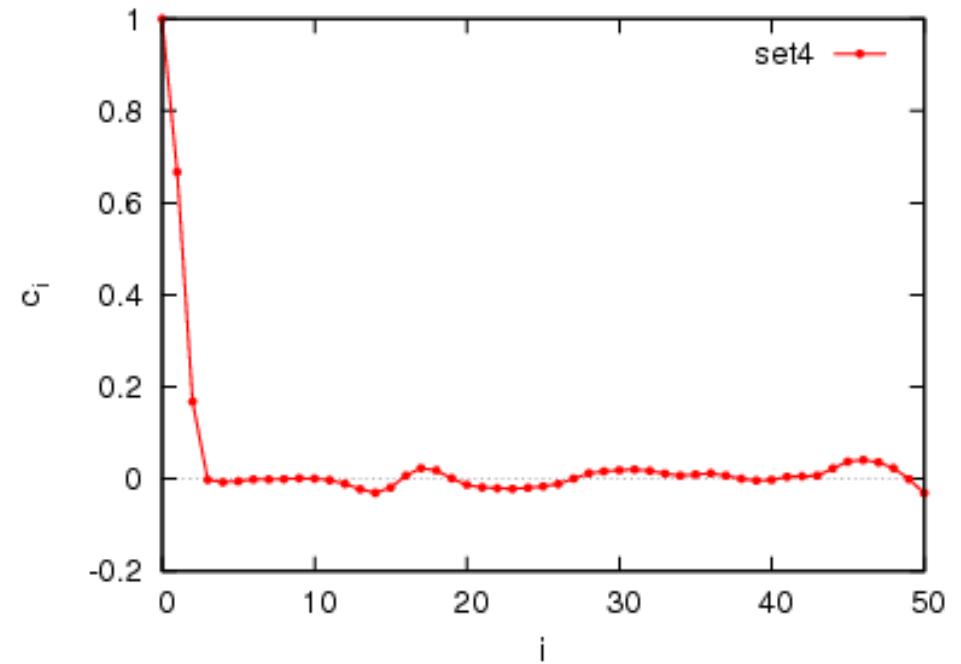
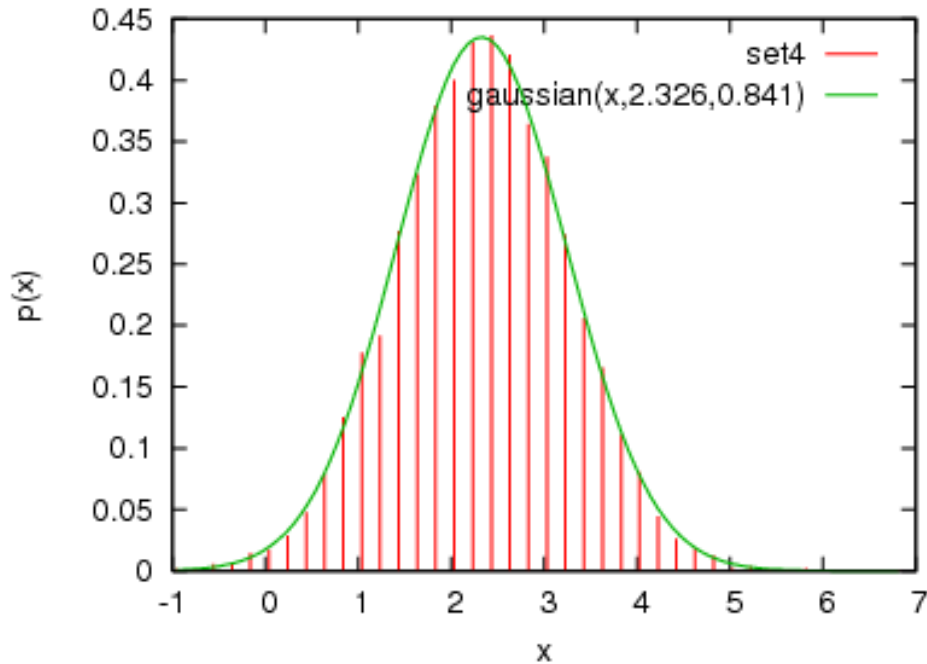
Estimates of mean / initial error analysis

- $\langle x \rangle \approx 2.329 \pm 0.009$ (naive estimate)



- $\langle x \rangle \approx 2.329 \pm 0.009$ (naive estimate)
- transient, remove first 100 points $\rightsquigarrow \langle x \rangle \approx 2.326 \pm 0.009$
- some autocorrelation?

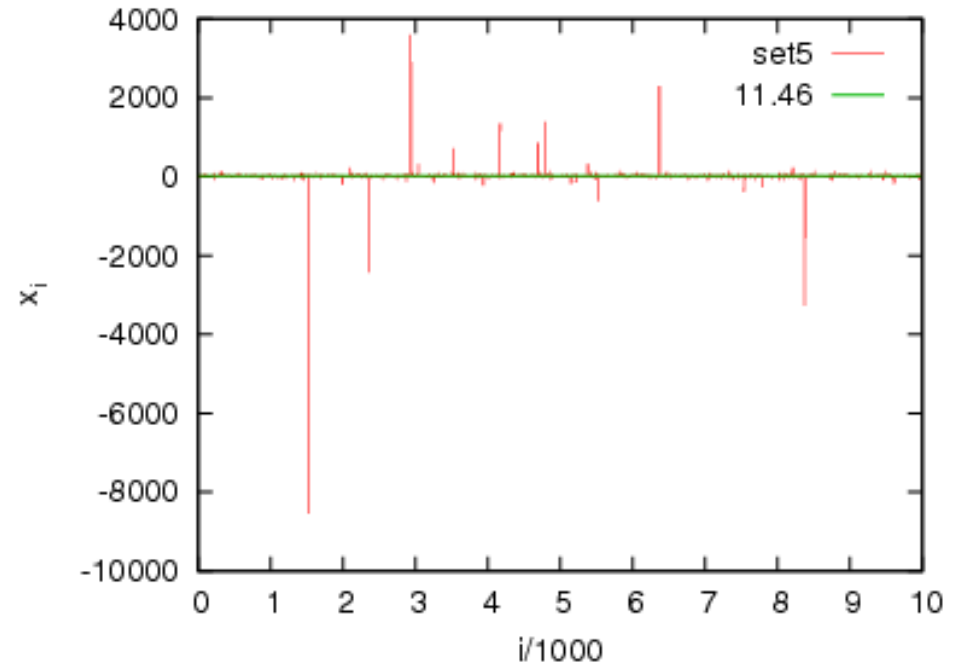
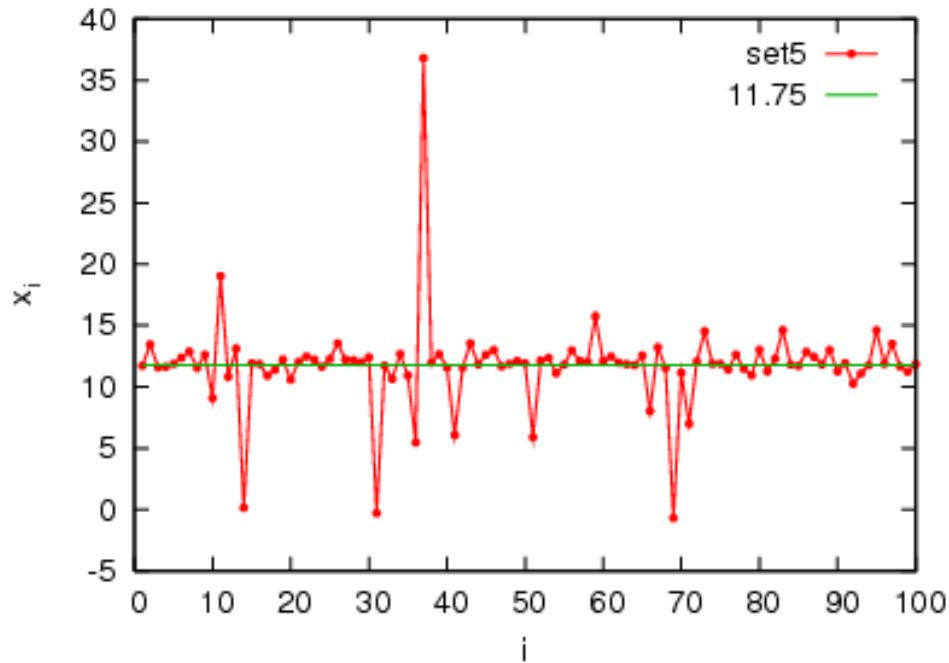
Histogram, autocorrelation function, full error estimate



- gaussian distribution of ‘‘measurements’’ - width $\sqrt{0.841}$
- finite autocorrelation ($\tau = 2.67$)
- final error estimate: $\langle x \rangle \approx 2.326 \pm 0.015$
- **true distribution:** gaussian, mean 2.310, variance 0.85

Set 5

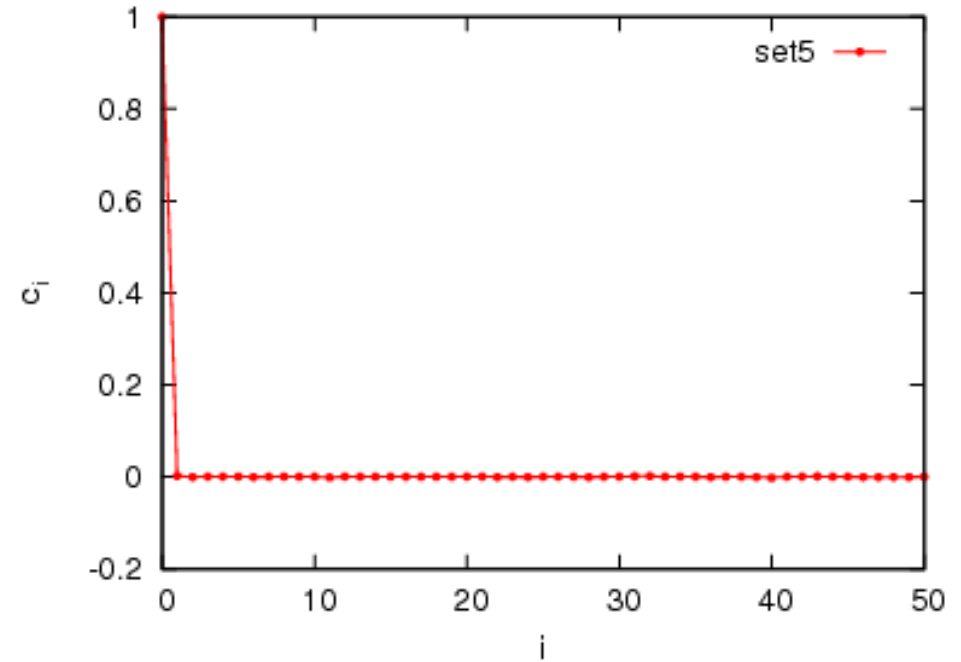
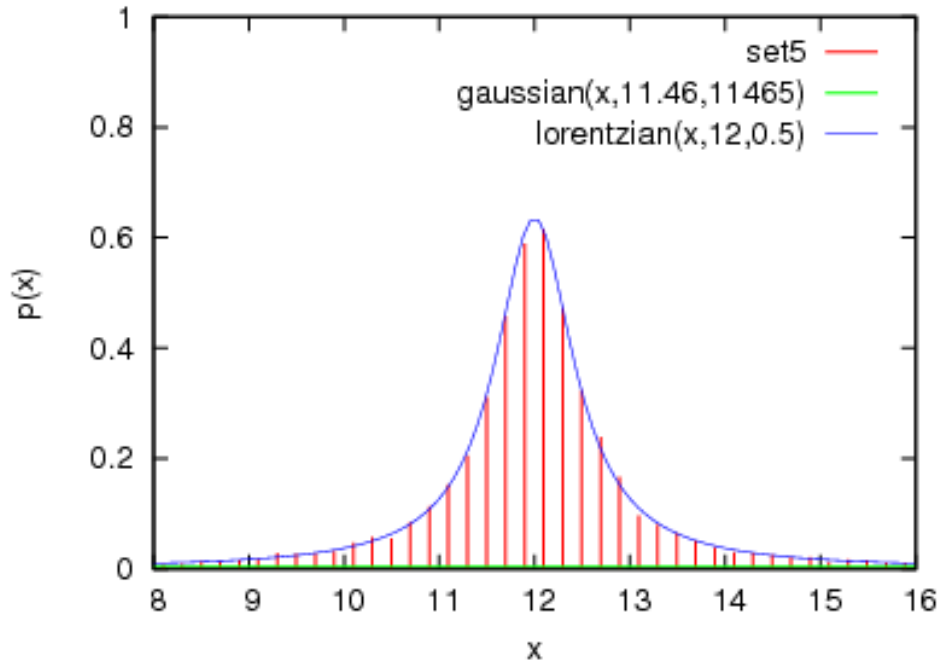
Trace and averages



Estimates of mean / initial error analysis

- $\langle x \rangle \approx 11.46 \pm 1.08$ (naive estimate)
- no transient
- no autocorrelation
- variance estimate increases with length of set!

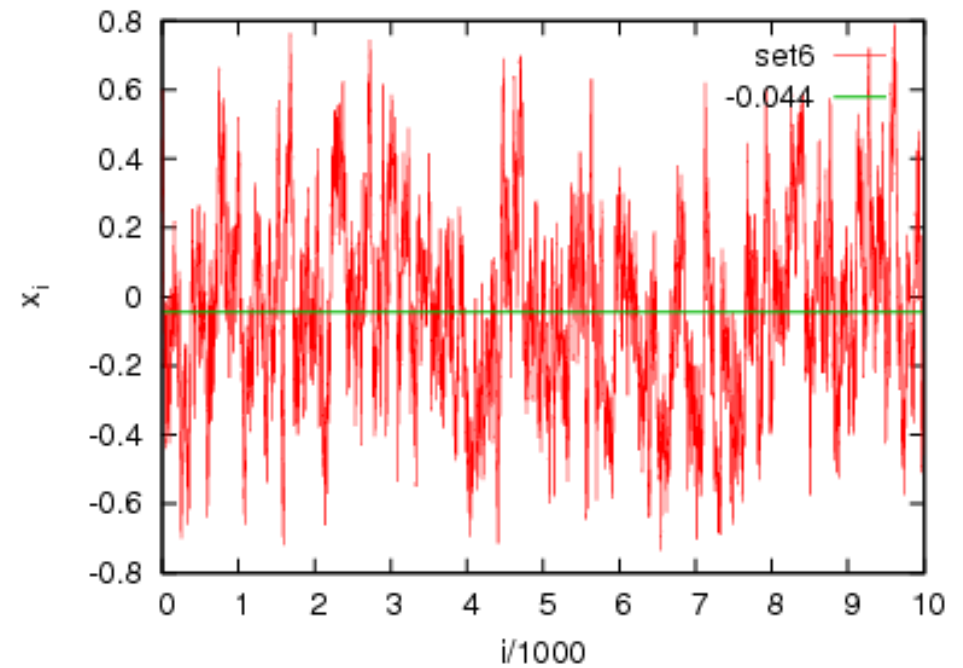
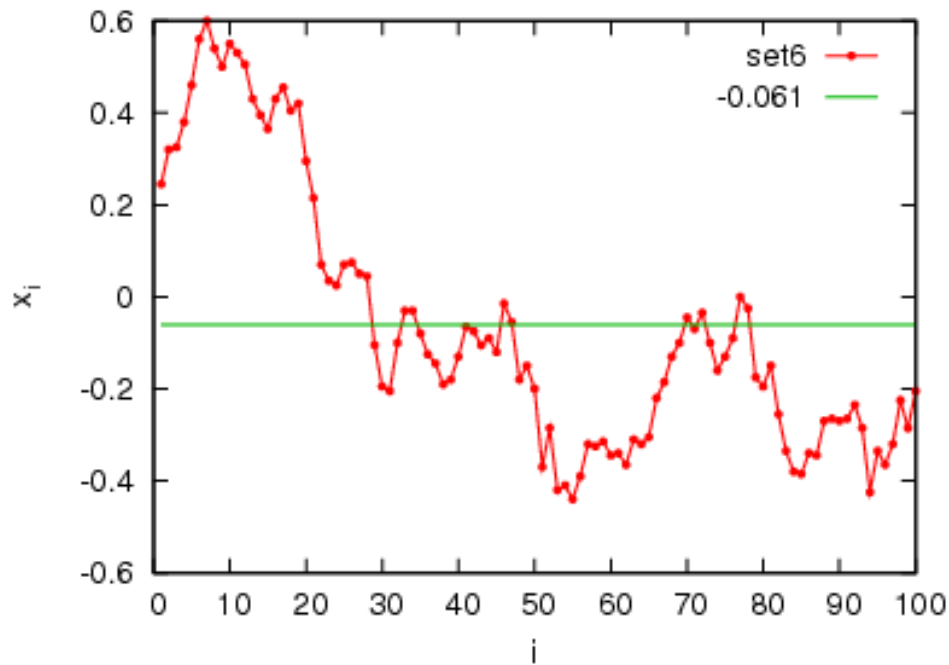
Histogram, autocorrelation function, full error estimate



- strongly nongaussian distribution of ‘‘measurements’’: lorentzian
- no autocorrelation ($\tau = 1.0$)
- no finite variance, still ‘‘consistent’’ statistics
- **true distribution:** lorentzian, median 12, ‘‘width’’ 0.5

Set 6

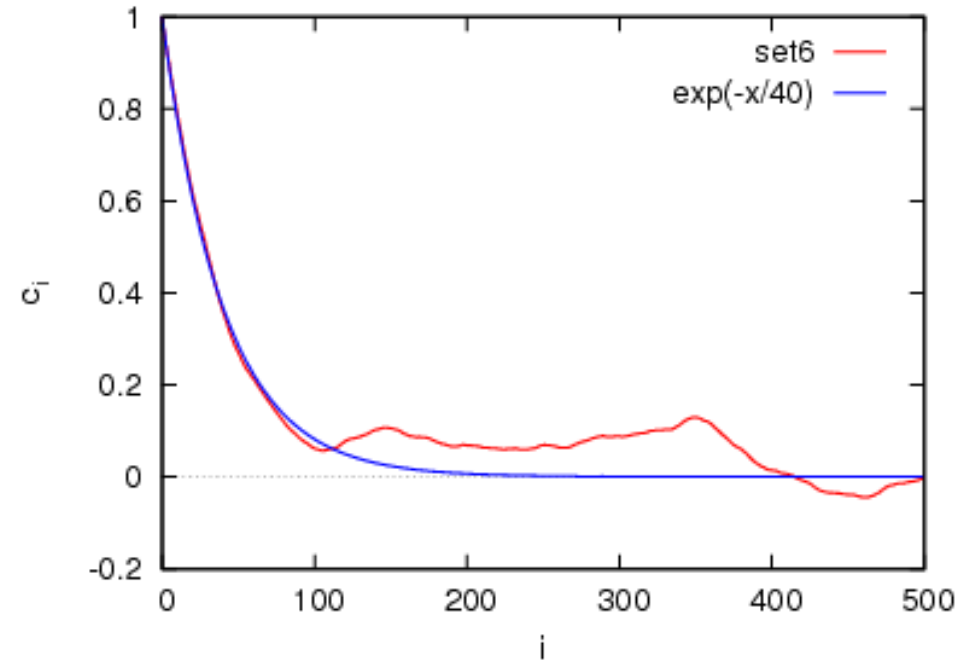
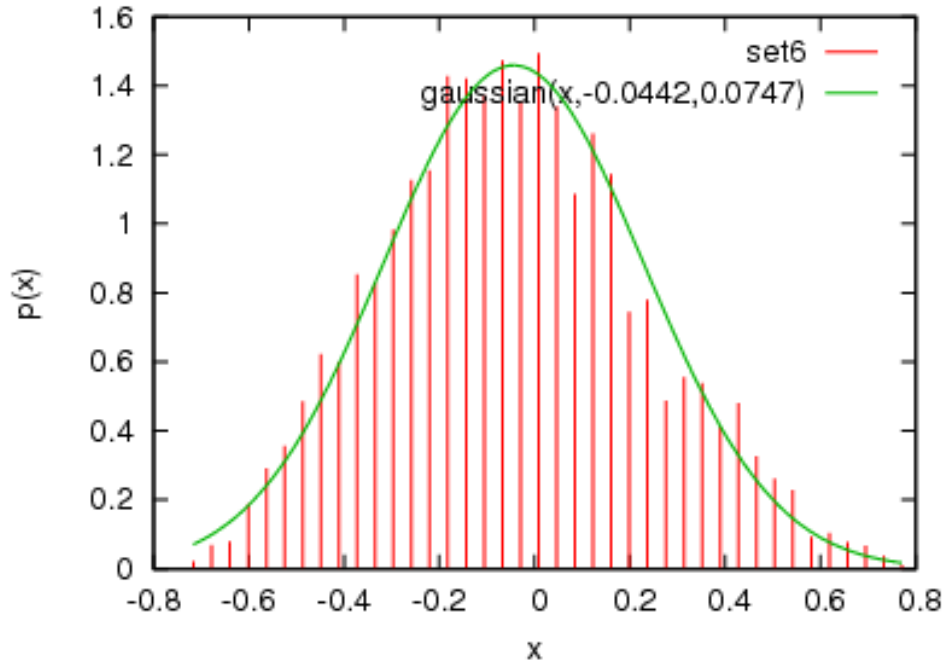
Trace and averages



Estimates of mean / initial error analysis

- $\langle x \rangle \approx -0.044 \pm 0.003$ (naive estimate)
- transient?
- long autocorrelation

Histogram, autocorrelation function, full error estimate



- essentially gaussian distribution of ‘‘measurements’’ - width $\sqrt{0.075}$
- strong autocorrelation ($\tau = 119$)
- final error estimate: $\langle x \rangle \approx -0.044 \pm 0.030$
- **true distribution:** 20x20 2D Ising model ($kT/J=2.6$), mean 0