Chapter 3
The battle against noise in the social sciences

Whenever the signal to noise ratio is smaller than one identification is likely to be unclear, unconvincing and open to discussion. For instance, it is because the signal in the Werther effect is of the same magnitude as the noise that the very existence of this effect is still a matter of debate thirty years after Phillips's pioneering paper. Similarly, many variables which are of central importance in economics, e.g. the elasticities of commodity prices with respect to supply or demand, are not known with a precision better than 30% or 50% [1]. Other figures regarding the accuracy of economic data can be found in Morgenstern (1950). Raising the signal to noise ratio is a crucial challenge for the social sciences. In this chapter we describe three methods for improving signal identification and we illustrate them through specific social phenomena.

Before we begin an additional remark is in order. Signal detection is an important topic in mathematical statistics. Here however, we propose upstream solutions to be used in the design phase of an experiment. Once the data have been recorded the fate of the battle against noise is largely settled. Using one statistical technique rather than another will improve matters only marginally.

[1] The fact that the elasticities are fluctuating in the course of time shows that many *different* phenomena are at work simultaneously. One would face the same situation when observing the movements of a pendulum in a train; instead of revealing the characteristics of the pendulum the observations would largely reflect the curves and bends of the railroad line.

1 The extreme value technique

Suppose we wish to study how the period of a pendulum depends upon the initial angular deviation θ_0. If we try initial amplitudes of 5, 10, 15, 20, 25 and 30 degrees the corresponding periods will differ by less than 5%. Thus, unless our measures are very precise it will not be clear if the period depends upon θ_0 or not. On the contrary if we try initial angles of 10, 90 and 179 degrees the corresponding periods will be sufficiently different to show unambiguously that the period increases with θ_0. This, in a nutshell, is the rationale of the extreme value technique. We will now show how it can be used in social phenomena.

It has been known since the late nineteenth century that the suicide rate is higher for unmarried than for married people. The ratio is about three for men and two for women (more details will be given in a subsequent chapter). Obviously, in a population in which there are more men than women all men will not be able to get married and one would therefore expect a higher suicide rate than in a population whose sex ratio is closer to one. Can this prediction be tested?

For most populations the sex ratio is confined in a narrow interval around 1, usually between 0.95 and 1.05. In such a population the expected effect will be very small. If the suicide rate of married men is s, the suicide rate of males in a population with a sex ratio male/female of 1.05 will be[2]: $(95s + 5 \times 3s)/100 = 1.15s$. For this effect to be detectable the background noise should be substantially smaller than 15% which is not the case. Thus, the effect will not be observable in this way.

The challenge is to find populations whose sex ratio is strongly different from one. It is well known that immigrants usually have an gender proportion involving more males than females. For instance:

- In 1890 the sex ratio of Chinese immigrants in the United States was 27
- In 1901 the sex ratio of Chinese immigrants in Australia was equal to 75.

[2]For the sake of simplicity we assume that the male to female ratio is the only limiting factor which limits the number of of marriages.

Of course, there are two additional requirements (i) these populations must be large enough to produce a sizable number of suicides (ii) the suicide rates for these specific populations must have been statistically recorded. In answer to the first question, there were 107,000 Chinese people in the United States in 1890 and 30,000 in Australia in 1901. Although not very large these populations are sufficient to produce fairly stable suicide rates especially if one performs averages over several successive years. This leads us to the second question: are there available data for these populations? The yearly *Mortality Statistics* volumes gives suicide numbers for Chinese immigrants in the U.S. but these data are being published only after 1923. By this time the sex ratio had dropped to 6.5, a level which, fortunately, is still high enough to produce an observable effect. Fig. 3.1 shows the suicide rate of people of Chinese descent in the continental part of the United States over the period 1923-1960, a time interval during which the sex ratio fell steadily from 6.5 to about one. As expected the suicide rate decreases along with the sex ratio. Over this time interval of 37 years the suicide rate was divided by 3. Compared to such a big change the fluctuations of 10% to 20% due to the background noise are of little importance. The Chinese immigrants are the parallel of the deviation of 179 degrees in the pendulum experiment. Of course, for social phenomena the success of the method depends on data availability, but thanks to the Internet revolution the availability of statistical data has improved tremendously[3].

The next section explains how signal identification can be improved if one has some knowledge about the date of occurrence, magnitude or shape of the expected signal.

[3] For instance all the volumes in the series of *Mortality Statistics* and the subsequent (after 1938) series of *Vital Statistics of the United States* are available on the website of the National Center for Health Statistics. More details on this question will be given in a subsequent chapter. Incidentally, it can be noted that we have not yet been able to find suicide data for Chinese immigrants in Australia.

2 Pattern matching: knowing when and what to observe

It turns out that young widowers under the age of 30 experience suicide rates which are almost 10 times higher than in the rest of the population. However, we do not know what is the average time interval η between the deaths of the wives and the suicide occurrences. The only information that we have is that η is less than 3 years. It would be very useful to know if η is of the order of one week, one month or one year. There is a similar uncertainty about the time constant of many economic mechanisms. For instance, the theory of international trade tells us that when a country has a permanent trade and current account deficit the exchange rate of its currency should fall until the deficits are brought down. However, we do not know whether this is supposed to happen after one, five or ten years[4]. Not knowing the time lags puts us in a fairly awkward position. To point out how weird such a situation would appear in physics let us return to the parallel with the pendulum. Suppose that the pendulum is at rest in vertical position and that an impulse force is applied to its mass. The existence of a time lag would mean that the mass begins to move only after, say, a few seconds or minutes; this would seem very surprising and clearly shows that lagged responses are fairly uncommon in physics. In this section we present an example in which the reaction time of the system is well known; as a result one knows precisely at which point in time the signal will appear. It will be seen that this knowledge greatly improves signal identification.

The population pyramid of Japan based on the census of 2000 presents a mysterious discontinuity for people aged 34 that is to say who are born in 1966 (Fig. 3.2 c) The number of people (both males and females) born in this year is much smaller than in 1965 or in 1967. The difference is of the order of 30%. Further investigation reveals

[4]For instance, Australia has had a trade and current account deficit for several years in the early 2000s, but at time of writing (June 2006) these deficits had not brought about a fall in the exchange rate of the Australian dollar. On the contrary, between January 2002 and January 2006, the Australian dollar progressed against the U.S. dollar (from 0.52 to 0.75) as well as against the euro (from 0.58 to 0.61). One should add that interest rates in Australia were not substantially higher than in the U.S. Even if the exchange rate of the Australian dollar eventually falls it will be difficult to say if it was indeed the deficit which was the crucial factor. In a general way, the longer the time lag beween cause and effect, the more difficult it is to demonstrate the existence of a causal link (because of the larger number of exogenous shocks).

that 1966 was a Hinoeuma year, which means a Fire Horse year in the Chinese calendar[5]. Girls born in that year grow up to be known as “Fire Horse women” and are reputed to be headstrong and to bring bad luck to their families and to their husbands. In 1966, as a baby’s sex could not be reliably identified before birth, there was a big increase in the number of abortions which brought about the sharp fall in birth rate observed on the population pyramid. According to the Chinese calendar, Fire Horse years occur every 60 years; thus, the three previous one were in 1906, 1846 and 1786. It is a natural question to see if the effects in those years were similar to the one in 1966.

Fig. 3.2a shows the curve of the male/female sex ratio in 1888 which is the first year for which such data are available. We see a distinct spike which corresponds to people aged 42 that is to say born in 1888-42=1846. It corresponds to a sex ratio about 15% greater than its normal level of about 1.05 [6]. The lower curve shows the total, male plus female, population. One would expect a dip of (at least) 15%/2=7.5%; it happens to be somewhat larger at 11%.

Is it possible to detect an after effect due to the Fire Horse year of 1786? People born in this year would have 1888-1786=102 years in 1888; unfortunately there are no data available for age groups over 84 years which means that no direct observation is possible[7]. However, if there was a deficit in girls of the same magnitude as in 1846, this should have lead to a reduction in the number of marriages about 20 years later; thus, the generation born around 1786+20=1806 should be somewhat smaller. In 1888, the people born in 1806 are 82 years old; can we identify an indentation in the total population curve in the vicinity of 82 years? The answer is no. Even if the last part of the curve is greatly magnified no trough can be detected. It is probable

[5] Because the Chinese New Year occurs in late January, the Fire Horse Year does not exactly coincide with 1966; in fact, it started on 21 January 1966 and ended on 8 February 1967.

[6] This corresponds approximately to a deficit of 38,000 girls.

[7] The fact that the sex ratio spikes have a good persistence in time is shown by the population pyramide of 1925 in which the generation born in 1846 is 79 years old; the spike in the sex ratio is still clearly visible even though it has been somewhat eroded by the higher male than female mortality that prevails in old age.

that the dip which may have existed at birth was fairly broad and was further leveled off and smoothed during the 82 years of life time of these people[8].

After 1786 and 1846 we come to 1906. This effect is described in Fig. 3.2b. This figure gives the population pyramid in 1913. There is a spike for people who are 7 years old, that is to say who are born in 1913-7=1906. The sex ratio spike of people born in 1906 has an amplitude of 4.0% and the trough in the male plus female curve has an amplitude of 11%. As statistical birth data are available for 1906, it is possible to check our previous conclusions. It turns out that the total births in 1906 are 11% smaller than the average for the other years of the decade 1901-1910. The sex ratio at birth is 4.3% higher than the average of the decade. Thus, these figures indeed confirm those that we read on the population pyramid. For people who are 67 years old (i.e. born in 1846) the sex ratio spike is still clearly visible but the small trough in the male plus female curve would be indiscernible if one did not know where to look for it.

After 1906 we come to 1966. The population pyramid of 2000 shows the sharp trough already evoked. One may wonder if, as in 1846 and 1906, there was an anomalous sex ratio at birth. Of course, one expect it to be much smaller than in 1906 and this is already visible on the sex ratio curve. More detailed birth data show that in 1966 the sex ratio was 1.3% larger than over the other years of the decade 1961-1970. This small excess sex ratio can no longer be detected in the population pyramid of 2000 when the Fire Horse generation is 34 years old.

The sex ratio at birth for the three Fire Horse years for which data are available are summarized in Table 3.1. There is a last question that Fig. 3.2c can help to answer: what effect has a sudden drop (or increase) in birth rate one generation later? In accordance with the extreme value technique exposed in the previous section,

[8]Naturally one could try to repeat this reasoning a second time. As the generation born in 1806 was reduced, the generation born 20 years later should also have been smaller. These people would be 62 years old in 1888. There is indeed a small indentation in this age group. However, it would be hazardous to draw any definite conclusion. For one thing our argument rests on the assumption that all people get married and have their first child at the age of 20 which is of course a rough approximation.

we first examine the effect of the huge peak (P) that occurred between 1947 and 1952 (visible on Fig. 3.2c (ages between 48 and 53). There is indeed a subsequent peak (P') in the male plus female curve around the age of 28. Three additional observations can be made.

- The magnitude of P' is only half the magnitude of P.
- P' is about twice as wide as P.
- The time lag between P and P' is about 25 years.

Can we make a similar observation for sharp troughs? The answer seems to be yes. There is a sharp trough at age 54 that corresponds to people born in 1946; these people were 20 year old in 1966 that is to say in coincidence with the Fire Horse year. Therefore, it is quite plausible that a fraction of the birth rate trough of 1966 should be attributed to the after effect of the birth rate trough of 1946.

In conclusion, we have seen that the 60-year periodicity in the occurrence of the Fire Horse years was of considerable value in helping us to decipher the fluctuations in the population pyramids. In this specific case, we relied on a recurring time pattern, but it is clear that any pattern, whether in time, space or any other variable will be of great usefulness.

We now turn to a third method of signal identification which can be seen as an extension of the law of large numbers.

3 Reducing noise by adding up several realizations

Suppose that for the purpose of a class-room experiment one wishes to measure the period T of a pendulum. A standard procedure is to measure the time t_{10} of 10 oscillations and get the period by dividing by 10: $T = t_{10}/10$. Such a procedure makes sense because the main uncertainty comes from the operation of starting and stopping the chronometer. If γ designates each of these uncertainties, the relative error on the measurement of T will be $2\gamma/t_{10} = (1/10)(2\gamma/T)$ whereas it would be $2\gamma/T$ if only one period had been measured. The only reason why we made this rather trivial point is because it shows that in itself the rationale of the procedure has little to do with the law of large numbers. It is only if we assume that all periods are in fact slightly different (due to vibrations, friction, draft) that the addition of random variables will play a role. As one knows, the argument relies on the two following rules.

- The variance of a sum of two independent random variables X, Y is the sum of the variances: $\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y)$
- The variance of the variable λX is given by: $\sigma^2(\lambda X) = \lambda^2\sigma^2$

Combining these results and assuming in addition that X and Y have the same standard deviation σ one gets:

$$\sigma^2\left(\frac{X+Y}{2}\right) = 2\frac{\sigma^2}{2^2} = \frac{\sigma^2}{2}$$

Similarly for the average of 10 random variables one gets:

$$\sigma\left(\frac{1}{10}\sum_{i=1}^{10} X_i\right) = \frac{1}{\sqrt{10}}\sigma \simeq \frac{1}{3.2}\sigma$$

Thus by measuring 10 vibrations the measurement will be roughly 3 times more accurate[9]. Unfortunately, this argument is of little usefulness in the social sciences for two main reasons (i) usually, the different realizations are not independent, (ii) very often one cannot repeat the experiment as often as one would like. Let us illustrate these difficulties by a few examples.

In the previous chapter we mentioned that the Werther experiment can be easily repeated. Obviously, however, there is a limiting factor which is the number of suicides which are announced on the front page of the *New York Times*. In the 1950s and 1960s there were on average between one and two suicides every year which

[9]Naturally, this argument does not apply if the system is not stationary for instance if a window has been left open which provokes an increasing amount of draft.

were announced on the front page. Thus, over a period of 20 years the number of realizations will be of the order of 40; the square root of 40 is 6.3. It remains to be seen if this factor is large enough. We come back to this point later on.

Sometimes data are available for a large number of realizations but they are not independent. For instance, in 2004 there were 2,768 stocks and 1,059 bonds listed on the New York Stock Exchange. For each of these stocks, data are available in the form of time series. Unfortunately, these time series are not independent. This is fairly obvious for companies which belong to the same economic sector. Thus, in 2004-2005 the stock prices of Exxon Mobil and of Chevron (two companies involved in the production of oil) have been highly correlated with a correlation of about 0.95. Stock prices of companies in different economic sectors may be correlated as well because some variables (e.g. the number of mergers and acquisitions, the number of buybacks, the inflation rate) affect almost all stocks globally. As is fairly obvious intuitively (and will be shown later on), when two time series are highly correlated taking their average does not reduce the standard deviation.

This discussion suggests that in order to be helpful in the social sciences the standard argument must be extended in two directions.

- Because the number n of realizations is often limited one would like the standard deviation of the average to decrease faster than $1/\sqrt{n}$.
- Because the realization are often interdependent the probabilistic argument should be extended to include correlated random variables.

As will be seen, the second requirement will help us to fulfill the first condition as well.

For the sake of simplicity we consider the average of a sum of three correlated random variables X_1, X_2, X_3 of mean zero and identical standard deviation σ. The conditions on the mean and standard deviation are not a limitation because if initially the variables X'_i do not satisfy them it is always possible to carry out the transformation:

$X_i = (X'_i - m'_i)/(\sigma'_i/\sigma)$ where m'_i, σ'_i denote the mean and standard deviation of X'_i. Our objective is to compute the standard deviation of:

$$Y_3 = S_3/3 \quad S_3 = X_1 + X_2 + X_3$$

In accordance with the rules stated previously:

$$\sigma^2(Y_3) = (1/3^2)\sigma^2(S_3)$$

By definition of the variance and due to the fact that the expectation of S_3 is equal to 0 [10] one gets:

$$\sigma^2(S_3) = E\left[S_3^2 - E^2(S_3)\right] = E\left[\sum_{i=1}^{3} X_i^2 + 2(X_2X_3 + X_3X_1 + X_1X_2)\right]$$

$$\sigma^2(S_3) = \sum_{i=1}^{3} E(X_i^2) + 2\left[E(X_2X_3) + E(X_3X_1) + E(X_1X_2)\right]$$

We express the expectations of the products by introducing the coefficient of correlation of the X_i:

$$r_{12} = \frac{E\left[(X_1 - E(X_1))\,(X_2 - E(X_2))\right]}{\sigma(X_1)\sigma(X_2)} = \frac{E(X_1X_2)}{\sigma^2}$$

Thus:

$$\sigma^2(S_3) = 3\sigma^2 + 2\sigma^2(r_{23} + r_{31} + r_{12})$$

Introducing the mean of the r_{ij}, $\overline{r} = (r_{23} + r_{31} + r_{12})/3$ we obtain:

$$\sigma^2(S_3) = 3\sigma^2[1 + 2\overline{r}]$$

and finally:

$$\sigma(Y_3) = \frac{\sigma}{\sqrt{3}}\sqrt{1 + 2\overline{r}} \tag{3.1a}$$

This formula has an obvious generalization to an arbitrary number n of random variables:

$$\sigma(Y_n) = \frac{\sigma}{\sqrt{n}}g \quad g = \sqrt{1 + (n-1)\overline{r}} \tag{3.1b}$$

[10] It should be recalled in this respect that the expectation of a sum of random variables is always equal to the sum of the expectations whether the variables are independent or not.

where:

$$\overline{r} = \frac{1}{[n(n-1)/2]} \sum_{i \neq j}^{n} r_{ij}$$

When the r_{ij} are all equal to zero $g = 1$ and we get the standard result for independent variables. On the other hand, $\overline{r} = 1$ implies that all the r_{ij} are equal to 1; in this case the three variables are identical (with probability 1) and one gets: $\sigma(Y_3) = \sigma(X_1) = \sigma$ in agreement with formula (3.1).

Formula (3.1) has interesting implications when $\overline{r}$ becomes negative. First we observe that for $n = 3$ the smallest value that $\overline{r}$ can take is $\overline{r} = -1/2$. In other words, for three random variables, it is impossible that $r_{23} = r_{31} = r_{12} = -1$. This makes sense intuitively because $r_{12} = -1$ and $r_{13} = -1$ imply that $X_2 = -X_1$ and $X_3 = -X_1$ which implies of course that $X_2 = X_3$ and $r_{23} = 1$, hence $\overline{r} = (-1 - 1 + 1)/3 = -1/3$.

When $\overline{r}$ is equal to -0.5 the standard deviation of Y_3 is equal to zero. In other words, the background noise represented by σ is completely eliminated. Just to show how dramatic this effect can be we set up a simulation in which a small deterministic signal, a lightly damped vibration, has been added to white noise. As the amplitude of the deterministic signal is only a fraction of the amplitude of the noise, it is completely hidden as can be seen in the first line of Fig. 3.3. The two series have zero mean, but they do not have the same standard deviation: $\sigma_1 = 0.97$ and $\sigma_2 = 0.49$. The correlation of the two series is -0.96. The average of the two series (panel 3) is almost as noisy as the initial series and the deterministic signal is still invisible. However, if we normalize the series by dividing them by their standard deviation before taking their average, the level of noise is drastically reduced and the deterministic signal becomes clearly visible. Of course, this is a simulation and such a dramatic effect is not likely to be observed with real series. Nonetheless, we will show by two illustrations that the method can indeed be helpful.

Application 1 We build a data set of daily stock prices for 24 stocks: 20 stocks from the Dow Jones Industrial plus the first 4 stocks of the Standard and Poor's 500 sample. All series cover the period from January 1999 to the end of 2002. As these series are mostly for large corporations they follow more or less the price evolution of the market as described for instance by the evolution of the Standard and Poor's 500 index. In order to get rid of this common trend we divided the series by the S&P500 index. Then, we normalized the series to reduce their mean to zero and their standard deviation to 1. With these 24 series, $(24 \times 23)/2 = 276$ pairs can be formed; for each of these pairs we computed the correlation coefficient.

- It turns out that the pair which gives the most negative correlation is (Hewlett-Packard, Altria)[11]. The correlation is -0.84 which gives a coefficient g equal to $\sqrt{1-0.84} = 0.40$; this means that, due to the negative correlation, the standard deviation is divided by 2.5 with respect to what would be obtained with two independent time series.

- With the 24 series it is possible to form $(24 \times 23 \times 22)/(1 \times 2 \times 3) = 2024$ triplets. For each of these triplets (i, j, k) one can compute the average $\overline{r}$ of the 3 cross-correlations r_{jk}, r_{ki}, r_{ij}. The triplet which turns out to have the most negative $\overline{r}$ ($\overline{r} = -0.40$) is (American International Group, Du Pont de Nemours, Hewlett-Packard)[12]. In this case, $g = 0.45$ which means that the standard deviation is 2.2 times smaller than the standard deviation of three non-correlated series.

- The same operation can be performed for the 10,626 quadruplets, the 42,504 quintuplets, the 134,596 sextuplets, etc. The only practical limitation is the computing time which rises very rapidly. Thus, it took about 40 hours of computing time to identify the quintuplet which gives the most negative correlation. It corresponds to the following companies: Du Pont, Hewlett Packard, Intel, Altria, Merck. The average correlation is $\overline{r} = -0.21$ which gives $g = 0.40$. It would require $5/0.40^2 = 31$ non-correlated companies to achieve the same reduction in standard deviation. The main advantage of achieving the same reduction with only 5 companies lies in the

[11] Hewlett-Packard is a computer company, while Altria is (since 2003) the new name of the Philip Morris company.
[12] American International Group is a consortium of insurance companies, Du Pont is a chemical company

fact that a common deterministic signal is more likely to be found in 5 companies than in 31.

Of course, once noise reduction has been carried out, the main question is how to interpret the resulting average. For instance, in the 200 days after the shock of September 11, 2001 the average of the previous five stocks displays a succession of sinusoidal oscillations of increasing period and increasing amplitude:

$$(25 \text{ days},\ 0.06),\quad (45 \text{ days},\ 0.10),\quad (77 \text{ days},\ 0.14),\quad (92 \text{ days},\ 0.20)$$

This kind of pattern may give us an insight into the main vibrating modes of the New York stock market but it must of course be confirmed by the observation of a similar pattern in the wake of other major shocks.

Application 2 This second application is not strictly speaking about signal identification, but it has much to do with noise, addition of random variables and inter-correlations. It will help us gain an understanding of what determines the standard deviation of suicide rates at county, state and nation level. This knowledge will be of great importance for the discussions of the Werther effect at the end of the chapter.

First, we describe the procedure which leads to the curves in Fig. 3.4. We selected (fairly randomly) 6 counties whose populations are comprised between $80,000$ and $100,000$ inhabitants. The average population of these counties , namely 0.09 million, defines the x-value in the graph of Fig. 3.4. From the database of the Center for Diseases Control, we get the numbers of suicides in each year of the 20-year long time interval $1979-1998$. This allows us to compute the standard deviations of the 6 time series; their average, namely $\sigma = 3.76$ defines the y-coordinates of the squares in the graph of Fig. 3.4. In order to estimate how much these series are interdependent we compute the pair correlations of the 15 pairs that can be formed with the 6 series. The average of these correlations, namely -0.03 defines the y-value of the circle (the corresponding scale is on the right-hand side). Then, we repeat this procedure for population units of increasing sizes. The last square on the far right corresponds to the United States. No correlation can be computed in this case because there are no other series in this size group. While the two solid curves correspond to observations, the dashed line shows the theoretical function $y = 1/\sqrt{x}$ that would be expected if all series were uncorrelated.

How should Fig. 3.4 be interpreted? It shows that the empirical curve remains close to the $1/\sqrt{x}$ curve until the population reaches a threshold size of about 1 million. For population units over one million, the decrease in the standard deviation is slower than $1/\sqrt{x}$ and at the same time the intercorrelations increase. At this point we must say a little bit more about the curve $y = 1/\sqrt{x}$. Unlike stock prices for which there are no standard models, suicide rates can be described in a natural way as the increment of a Poisson process. A Poisson process $X(t)$ is defined by the assumption that during each time interval Δt there is a probability $\lambda\Delta t$ that a new suicide occurs. This process models the cumulative number of suicides which is an increasing function of time. The number of suicides in a given time interval is so to say the derivative of the Poisson process; for instance the annual number of suicides $Y(t)$ is defined as the increment of $X(t)$:

$$Y(t) = (X(t+\theta) - X(t)/\theta$$

where θ represents one year. It can be shown that according to this simple model, the standard deviation of the suicide rates is given by: $\sigma = \sqrt{t_m/N\theta}$, where t_m denotes the average suicide rate and N the size of the population unit under consideration (Papoulis 1965, p. 287). The fact that σ is proportional to $1/\sqrt{N}$ as in the addition of independent random variables does not come as a surprise because it is known that a Poisson process can alternatively be defined as a sum of an increasing number of random variables (Papoulis 1965, p. 558). Why then, does the curve of the standard deviation break away from $1/\sqrt{x}$ in the population range over one million?

A possible interpretation consists in assuming that there are two different sources of noise which affect suicide rates (i) A purely random component which leads to a standard deviation σ_r and is well described by the Poisson increment process; this random component corresponds to a large number of factors at individual or local level. (ii) A deterministic source of noise which leads to a standard deviation σ_d and which corresponds to the response of the system to a few macro-factors at regional or national level. As examples of such macro factors one can mention the marriage and divorce rates or the unemployment rate[13]. In short, $\sigma = \sigma_r + \sigma_d \quad \sigma_r \sim 1/\sqrt{N}, \quad \sigma_d \simeq 0.5$

Of course, the σ_d component exists at all population levels, but for small units σ_r is large enough to make σ_d almost invisible in relative terms.

The nature of the deterministic factors can be confirmed by the correlation curve. Suppose for a moment that these factors are local factors at county level. In this case there would be no reason for the intercorrelation of suicides to increase with unit size. In contrast if the deterministic factors are indeed macro-factors at regional or national level it is not surprising that they may bring about increased inter-correlation[14].

We close this chapter by a brief discussion of the questions of statistical significance and confidence intervals.

4 Confidence intervals and statistical significance

The concepts of signal to noise ratio and of error bars that we used in this chapter are commonly used in physics and in electrical engineering. Econometricians as well as the social scientists who have adopted the language of econometrics rather rely on the notion of test of significance. In this section we give precisions about the notions

[13]Empirical studies show that the correlation between unemployment and suicide is fairly low but there can be an indirect relationship. A possible connection may be through the disruption of the family unit occasioned by unemployment. If this interpretation is correct, the connection between unemployment and suicide should be very dependent upon the level of social protection; thus, one would expect the connection to be lower in Scandinavian countries than elsewhere.

[14]The fact that marriage or unemployment rates are not necessarily the same in nearby counties is irrelevant for our argument. What matters is the rate of change of these factors over a period of 20 years. It turns out that these changes are slow and smooth enough to be called deterministic.

of confidence intervals, test of significance and how they are related[15]. First we will recall the definition of statistical significance; then we give some typical orders of magnitude; finally, we discuss relevant applications to the social sciences. However, before we begin, it should be recalled that when the signal to noise ratio is high, say higher than ten, the results are clear-cut no matter what statistical tests are used to analyze them; on the contrary, if the signal to noise ratio is low the conclusions will be uncertain unless one has additional information about the expected signal which allows pattern matching (or similar procedures) to be used.

If a random variable X has a Gaussian frequency distribution of mean m (which we take equal to 0 for the sake of simplicity), and standard deviation σ, X will fall into the interval $(-1.96\sigma, 1.96\sigma)$ in 95% of the drawings (Ventsel 1977, p. 307). Equivalently, we can say that there is a likelihood of 0.05 for random drawings of $|X|$ to be larger than 1.96σ. Thus, if a spike that one believes to be a signal has an amplitude around 1.96σ where σ is the standard deviation of the background noise, there is one chance in 20 that the "signal" is in fact a random noise fluctuation[16].

Naturally, if the amplitude of the signal is larger than 1.96σ one would expect the likelihood that it is noise to be smaller. Table 3.2 summarizes some typical values. It should be emphasized that from a scientific perspective there is no threshold of significance that is better than another. In other words a probability threshold of 0.1 is just as "good" as one of 0.001. Whether one adopts one particular threshold or another is an issue which depends upon the context of the experiment. If the number of events produced by the experiment is large one can be more selective, if the number

[15]To statisticians and econometricians the approach that we use may appear rather unsophisticated. However, one should keep in mind in a general way that the more sophisticated a statistical test, the larger is the set of assumptions on which it relies and the greater the difficulty of checking that the dataset under consideration indeed satisfy these requirements. Very often this proof is simply omitted either because the dataset is too limited to permit the appropriate verifications or because it would be too time consuming to do so.

[16]Strictly speaking, 0.05 is the probability $P\{|X|/\sigma > 1.96\}$, but due to the very rapid decrease of the Gaussian function the probability that $|X|/\sigma$ is substantially larger than 1.96 is so small that in fact:

$$P\{|X|/\sigma > 1.96\} \simeq P\{|X|/\sigma \simeq 1.96\}$$

of events is small, being very selective would make most results non-significant and would therefore be a fairly unproductive procedure. To make this discussion more concrete we consider an application to the specific case of the Werther effect.

Statistical significance of the Werther effect Is there really an increase in the suicide rate after the suicide of a celebrity or is the effect too small to be clearly detected? We examine how the previous notion can help us to answer this question. In his paper of 1974, Phillips analyzes $n = 33$ events. Each event is a suicide story published on the front page of the *New York Times*. In each case, Phillips compares two variables: (i) an observed number of suicides s and (ii) what he calls an expected number of suicides e. For definiteness, let us consider the case of Marilyn Monroe. She died on 6 August 1962, so s will be the number of suicides which occurred in the United States during the month of August 1962[17]. From the *Vital Statistics of the United States* (Vol. II, part A, section 1, p. 77) we learn that $s = 1838$. In July, there were only 1659 suicides, but this increase is of little significance because of the seasonal pattern described in the previous chapter. To make the test independent of the seasonal pattern, Phillips used the following procedure. He computes the average number of suicides in August 1961 and August 1963 which he calls the expected number of suicides: $e = (1579 + 1801)/2 = 1690$. The test variable is defined as the difference between s and e: $p = (s - e)/e = 8.8\%$ [18].

How significant is a value $p = 8.8\%$? The answer is given in table 3.2 under the condition that the distribution of the numbers p is Gaussian, and provided we know the standard deviation of the p. In order to determine the distribution of the p we need far more years than just 1961, 1962 and 1963. How many years do we need? As each year gives 12 values of p, a sample of 5 years will give 60 values which is

[17]Clearly if the suicide of a celebrity occurs late in the month, it makes more sense to consider the suicides in the subsequent month. Phillips chose the 23rd day as the cut-off point; thus if Marilyn Monroe's death had occurred on August 24 one would rather consider the suicides that occurred in September 1962.

[18]Phillips's paper (1974, p. 344) incorrectly gives $e = 1640.5$ which leads to the higher value $p = 12.0$; probably the mistake comes from a confusion with the following line which contains exactly the same number 1640.5. However, this mistake does not substantially affect the overall conclusion for the whole sample of 33 suicides.

sufficient to determine the standard deviation but is too small a sample to study the shape of the distribution. Such a determination requires at least 250 data points that is to say 20 years. When the test is performed over the period 1935-1960 (shown in Fig. 2.2) the distribution of p is indeed found to be reasonably Gaussian with zero mean and with a standard deviation $\sigma(p)$ equal to 4.92% [19] . This leads to a ratio $p/\sigma(p) = 8.8/4.92 = 1.79$. By using a table of the Gaussian integral (e.g. Ventsel 1973, p. 543), we find that there is one chance in 28 that such a deviation is merely a random fluctuation[20]. In other words, fluctuations of a magnitude of 8.8% occur every 28 months that is to say almost every two years[21]. As already mentioned, whether or not this signal should be considered as being "significant" is a subjective rather than a scientific question. From a scientific point of view the real question is whether the previous identification procedure can be improved. Three methods will be proposed but before that we would like to discuss the second case which is mentioned in Table 3.3a.

The death of Lady Diana on 31 August 1997 was not a suicide but an accident. The results in Table 3.3a show that the p number that is to say in the deseasonalized number of suicides in September 1997 reached 10%. As England has a smaller population than the United States, it is not surprising that the standard deviation $\sigma(p)$ is larger. As a result the signal to noise ratio $p/\sigma(p)$ is lower than in the Monroe case. Perhaps of greater interest is the fact that there seems to be a fundamental difference in the reactions of males and females. For men the signal to noise ratio is 1.1 whereas it is almost zero for women.

[19] In fact $\sigma(p)$ was computed over the period 1945-1960 during which the series is more stationary than during the whole period; over the whole period one gets $\sigma(p) = 5.74$, this higher estimate should certainly be attributed to the non-stationary.

[20] This result is consistent with the first line of Table 3.2: $28 \in (6, 40)$.

[21] What makes the Monroe case important is the fact that it is the event in Phillips's list which leads to the highest number of excess-suicides. Thus, in line with the extreme value technique, it is reasonable to begin to investigate this event in some detail.

5 Upgrading statistical tests

We discuss three methods for improving the previous results[22]. In each case we will need additional data. This illustrates our previous statement that identification can be improved in a substantial way only by including additional information.

- The first method consists in observing the effects of several deaths instead of just one. It is the same method that when one measures 10 swings of a pendulum instead of just one.

- The second method consists in observing the effect of a death in several places instead of just one. The detection of a gravitational waves by several detectors located in different countries relies on the same approach.

- The third method consists in observing the effect of a death over several months instead of just one in order to apply a pattern-matching procedure.

We now discuss each of these methods in more detail.

Several events This is the method that Phillips (1974) used in his paper. As we already mentioned he selected a sample of 33 suicides publicized on the front page of the *New York Times*. In each case he computed the variable $p = (s - e)/e$. The first question which arises is how many of the p values are positive. One finds that p is positive in 26 of the 33 cases which represents a percentage of 79%. The second question is: "Are the p values close to zero or markedly different from zero. One finds that their average is 2.51%.

Next comes the crucial question: "Does 2.51% represent a deviation which is significantly different from zero?". From the previous section we know that for each of the events the standard deviation of p is equal to 4.92%. If the 33 events are uncorrelated

[22] In the literature one can find numerous procedures of peak identification. Many of them are context dependent, such as for instance identification in radar detection or in astronomy. Many others use specific mathematical tools such as Fourier or wavelet analysis. In the present section, we focus on basic ideas rather than on specific techniques.

the standard deviation of their average will be

$$\sigma\left(\frac{1}{33}\sum_{i=1}^{33} p_i\right) = \sigma(p)/\sqrt{33} = 0.86$$

But are the events really uncorrelated? The answer is yes. The argument goes as follows: on average the time interval between two events is 20 years$/33$ =7.3 months. The events will be uncorrelated if the autocorrelation function of p falls to zero in a time which is shorter than 7.3 months. It turns out that the autocorrelation function falls to zero within 2 or 3 months. Thus, the signal to noise ratio is $2.51/0.86 = 2.92$. By using the formulas given in Table 3.2 we find that there is one chance in 526 for this fluctuation of the average to be merely a random fluctuation. If one recalls that in the case of Marilyn Monroe, the result was one chance in 28, we see that we have been able to greatly improve the level of significance.

Several places In the second method one explores the effect of one event in different places. Until 1950 the *Vital Statistics of the United States* provided monthly suicide data not only for the whole country but also for each state. Can we use these data to improve the significance of the test? To take advantage of this additional information we must select a suicide that occurred before 1950. In Phillips's list there are only 5 events which occur before 1950. To be in the most favorable position we select the event which gives the largest p value (i.e. 1.44%), namely the suicide of James Forrestal which occurred on 22 May 1949[23]. We consider the effect of this suicide in a sample of 17 states[24]. Table 3.3 b shows that the average of the p values for the 17 states is 2.32%. This value must be compared with the standard deviation of the average of p over the 17 states. Each state is characterized by a specific $\sigma(p)$ and their mean is 27%. If the series are uncorrelated[25] the standard deviation of the average is obtained by computing the sum of their variances divided by the number

[23]In Phillips (1974, p. 344) the expected number of suicides e is incorrectly given as 1493.5 instead of 1527.5; 1493.5 is identical with the figure in the line immediately below which suggests that the mistake is probably of the same kind as in the Monroe case, i.e. a confusion between two successive lines.

[24]These states were selected through the following criterion: they are the most populous states which belong to the registration area in 1912.

[25]When one computes the average intercorrelation over a sample of 10 pairs one gets indeed a correlation which is close to zero.

of states which gives approximately $27/\sqrt{17} = 6.5\%$. This result makes sense because it is of the same magnitude as the figure of 4.92 corresponding to the United States but slightly greater due to the fact that the 17 states are smaller than the U.S. Thus, we get a signal to noise ratio equal to $2.32/6.5 = 0.35$. This figure should be compared with the result obtained for Forrestal at the level of the whole country, i.e.: $1.44/4.92 = 0.29$. In other words, the signal to noise ratio was indeed improved but only marginally. This is due to two circumstances (i) In 9 of the 17 states p was negative which shows that in a majority of the states the signal was smaller than the background noise (ii) Furthermore, many states have relatively small populations and therefore their $\sigma(p)$ are fairly large. For instance, $\sigma(p)$ is equal to 11% in California but it becomes as high as 45% in Colorado whose population is seven times smaller.

In the previous attempt we used what can be called an indiscriminate, systematic statistical analysis. As we had no reason to expect the effect to be greater in one state than in another we treated them all in the same way. However, this is not the only possible strategy and in fact it may not be the most appropriate in the exploratory phase of the investigation. An alternative strategy is to adopt a case-study approach which focuses on specific states in order to discover underlying determinants. Let us illustrate this approach by the example of the suicide of filmstar Carole Landis (second entry in Phillips's list, 1974, p. 344). Born in Wisconsin, she died in California on 5 July 1948 at age 29. At the global level of the U.S. the p value of the suicides in July 1948 is positive but fairly small $p = 1.72\%$. However, one may expect the impact of her death to be stronger in states in which she was well known, for instance Wisconsin or California. In Wisconsin p takes on a negative value, but in California p is not only positive but fairly large, $p = 18\%$. For a proper interpretation of this result one would need to know the amounts of media coverage of her career before her death and of her suicide in the month following her death [26].

The same technique could be applied also to the death of Lady Diana. In this case

[26] An approach of the Werther effect based on the amount of media coverage was tried by Steven Stack (1987).

we can use separate data for England, Wales and Scotland. By the same reasoning we get a signal to noise ratio equal to 2.1 for males and 0.38 for females. For males the signal to noise ratio was doubled which confirms that males reacted much more than females.

Pattern identification The rationale of the third method can be explained as follows. We do not yet have any knowledge of the time dependence of the Werther or Diana effects. However, if these effects really exist one would expect the number of excess-suicides to decrease progressively after the peak instead of falling abruptly to zero. In other words it should be possible, at least for the largest peaks, to detect an excess number of suicides not just in the month following the death of the celebrity but in several subsequent months. This is indeed what is observed in the Monroe and Diana cases (Table 3.4).

Intuitively, one would expect the occurrence of such a relaxation pattern to be relatively rare in a random series. The question is how rare is it exactly? The probability of such a pattern can be estimated through the following reasoning. We consider a time series Y_i of Gaussian white noise of mean zero and standard deviation σ. The expression “white noise” means that values at different times are uncorrelated, a property which can be checked by verifying that the autocorrelation function is almost equal to zero for all non-zero time lags. For definiteness we consider a pattern for which:

$$Y_i \geq 2\sigma \quad \text{and} \quad Y_{i+1} \geq \sigma \tag{3.2}$$

What is the probability of such a pattern? From Table 3.2 we know that:

$$P\{Y \geq 2\sigma\} = 1/40 \quad \text{and} \quad P\{Y \geq \sigma\} = 1/6$$

If the time series has 10^6 points, the first assertion means that $10^6/40 = 25000$ points will be above the 2σ level; we call these points 2σ-points. Now consider the points which follow immediately the 2σ-points. They constitute a subsample of points whose values are independent of the 2σ-points because of the white noise assumption. Thus, we can apply the same reasoning to this subsample which leads to the result that there are:

$$\left(10^6 \times \frac{1}{40} \times\right) \frac{1}{6} = 4166$$

points which fulfill the requirement (3.2). This result (which can easily be checked by running a simulation) is summarized in the following rule[27]:

> **Identification of relaxation patterns** Y_i is a Gaussian white noise time series of mean zero and standard deviation σ. We consider the probability $P\left\{R\right\}$ of observing a relaxation pattern described by the following event:
>
> $$R = \left\{ \frac{Y_i}{\sigma} \geq a_0, \frac{Y_{i+1}}{\sigma} \geq a_1, \ldots, \frac{Y_{i+n}}{\sigma} \geq a_n \right\} \qquad a_0, a_1, \ldots, a_n > 0$$
>
> $P\left\{R\right\}$ is given by the following formula:
>
> $$P\left\{R\right\} = P\left\{\frac{Y_i}{\sigma} \geq a_0\right\} P\left\{\frac{Y_i}{\sigma} \geq a_1\right\} \ldots P\left\{\frac{Y_i}{\sigma} \geq a_n\right\}$$

When this formula is applied to the Monroe case one gets:

$$P\left\{\frac{s_i}{\sigma} \geq 1.79, \frac{s_{i+1}}{\sigma} \geq 1.42, \frac{s_{i+2}}{\sigma} \geq 0.77, \frac{s_{i+3}}{\sigma} \geq 0.51\right\} = 2.10^{-4}$$

In words there is a chance in 5,000 that such a pattern will occur in a purely random series[28]. A similar calculation can be performed in the Diana case, but in order to be on firm ground one would first have to explain what produces the sharp increase which occurs in November and December.

Rating the quality of the data Before closing this section, we would like to emphasize (once again) that prior to caring about statistical treatment of the data it is important to assess the quality of the data. This step is often omitted by social scientists. As we already mentioned, in physics the quality of experimental data

[27] This statement is a direct consequence of the independence of the variables $Y_i, Y_{i+1}, \ldots$

[28] One reason why such an impressive figure should not be taken too seriously is precisely because we are not sure that the suicide numbers constitute a *purely* random series. For instance, it cannot be excluded that the relaxation pattern is due to a series of correlated exogenous shocks which so to say mimic a decreasing output.

is guaranteed by the fact that the same experiment is performed by several groups. Even once a phenomenon is well known it is not uncommon that new experiments are performed in order to improve the accuracy of the measurement. In the social sciences the quasi-experiments done by one researcher are almost never repeated and checked by others. As an illustration, we will try to rate the quality of the data used in Phillips's paper (1974).

The criterion that the suicides must be publicized on the front page of the *New York Times* seems to define them unambiguously. However, this definition is not as clear-cut as could seem at first sight. Consider for instance the death of Marilyn Monroe. The title of the article in the issue of 6 August 1962 reads: "Marilyn Monroe dead. Pills near. Official verdict delayed". There was no qualification of suicide in the first announcement. Yet, this case was included in the sample. On the contrary, the death of Ernest Hemingway in July 1961 was *not* included in the sample in spite of the fact that it became known subsequently that it was indeed a suicide[29]. The title of the article on the front page of the *New York Times* of 3 July 1961 reads: "Hemingway dead of shotgun wounds. Wife says he was cleaning his weapon". As in the Monroe case there is no qualification of suicide in this first announcement. Thus, it is not obvious why the two events should be treated in different ways.

More generally, one can observe that 22 of the 33 cases comprised in the sample produce only 1.7% of the total number of excess-suicides whereas there are 6 cases which produce as much as 54% of the total. This suggests that the criterion based on the *New York Times* it too wide a net in the sense that it collects a lot of irrelevant events among which, almost by chance, there are a few ones which are of greater relevance.

The main practical messages of this chapter can be summarized as follows.

1) It is important to optimize the signal to noise ratio in the early phase of the design of the quasi-experiment.

[29] In contrast to the death of Marilyn Monroe, Hemingway's death was *not* followed by an excess number of suicides.

2) The extreme value technique can be of value especially when one wishes to assert the reality and order of magnitue of the phenomenon under consideration.

3 Before embarking into statististical tests it is appropriate to check the reliability and accuracy of the date (for instance by comparing them to similar data or making internal consistency checks). The question of data reliability is the subject of the next chapter.

4) Every time one has some *a priori* knowledge about the phenomenon under study it makes sense to use pattern matching techniques.

5) In order to estimate the likelihood that a fluctuation of amplitude $Y_t = s$ is due to the noise background (rather than to a genuine signal) one must compute the standard deviation $\sigma(Y)$ of the time-series Y_t with the best accuracy possible, which means over a time interval where Y_t is stationary and which is the longest possible.

6) By repeating the experiment several times in the course of time or by resorting to spatial disaggregation (i.e. observing the phenomenon in different regions of the country under consideration) it is possible to generate many realizations. If these realizations are not positively correlated (a negative correlation is not an obstacle but on the contrary an advantage) the averaging process often allows a substantial enhancement of the level of significance.

6 Conclusion

In this chapter we explained and illustrated several methods and techniques for improving the signal to noise ratio. As several of our illustrative examples concerned the Diana-Werther effect, it may appear somewhat surprising and frustrating that we did not propose a definite conclusion regarding the existence of this effect. We are in the same position as physicists who try to detect gravitational waves (see the previous chapter) in the sense that if we had several dozens “big” events such as the Monroe and Diana cases, it would be possible to draw a fairly clear conclusion. For the deaths of less known female celebrities, the effect is just too small to be detected. However, the evidence is fairly significant a the global level of sample of thirty events. It is by accumulating an ever larger number of tests that we will be able to draw a more definite conclusion.

In this chapter we suggested that in physics (thanks to the collective validation procedure) the reality (or non reality) of a phenomenon can usually be established fairly quickly. This may be true in 90% of cases, but there are also some cases in which it takes decades or even centuries before a definite conclusion can be reached. As an example one may mention the use of a divination rod in the discovery of underground mines or springs. This effect has puzzled physicists for over three centuries. It is only in recent times, thanks to the use of highly sensitive magnetomers, that the problem has received a preliminary answer (e.g. see Chadwick and Jensen 1971, Rocard 1981). Furthermore, one should keep in mind that even once a question has for the most part been solved there may be objections which cannot be answered adequately. For instance, the fact that the Earth moves around the sun was well accepted by most scientists in the early nineteenth century, but there was still the objection that no apparent displaments of the stars could be observed as a consequence of the Earth's motion. The fact that these displacements do indeed exist but are too small to be measured except for the nearest stars, became completely clear only in 1838 when Friedrich Bessel was able to measure the displacement (the so-called stellar parallax) of the star *61 Cygni*. Such examples teach us the great virtue of patience. In conclusion, the problem of the Diana-Werther effect will perhaps provide an exciting stimulus for coming generations of sociologists and econophysicists just as the measure of the stellar parallax did in astronomy.

Table 3.1 Male and female births in three Fire Horse years

	1846	1906	1966
Total births (%)	−11	−11	−24
Female births (%)	−19	−13	−24
Sex ratio (%)	+20	+4.3	+1.3

Notes: The percentages refer to the differences between the year under consideration and the mean of the 9 other years in the same decade. For 1846, as no birth statistics are available, the percentages were derived from the population pyramid of 1888. Surprisingly, the 1966 decrease in total births was substantially higher than in previous Fire Horse years. One may wonder if this should not be attributed to the influence of an additional factor. A possible candidate is the after-effect, one generation later, of the fall in birth rate which occurred in 1946.
Sources: Historical Statistics of Japan (http://www.stat.go.jp): Population by single years of age and sex, live births by sex and sex ratio of live births; Matsumoto (1975).

Table 3.2 Probability that a signal is in fact a random fluctutation

Signal (s) to noise ratio: $x = \frac{s}{\sigma}$	1	2	3.3	4	5	6	7	8	10
$P\{X > x\sigma\}$, one chance in	6	40	2000	30000	$3.4\ 10^6$	10^9	$0.8\ 10^{12}$	$1.6\ 10^{15}$	$1.2\ 10^{23}$
$P\{\lvert X\rvert > x\sigma\}$, one chance in	3	20	1000	15000	$1.7\ 10^6$	$0.5\ 10^9$	$0.4\ 10^{12}$	$0.8\ 10^{15}$	$0.6\ 10^{23}$

Notes: The table gives the probability that a Gaussian random variable X of mean zero and standard deviation σ is greater than a given threshold $x\sigma$; the probability is expressed as one chance in n drawings, thus one chance in 40 (which corresponds to a probability of 1/40) means that a signal greater than 2σ will on average be observed once in 40 random drawings. The second line gives the same information for the variable $|X|$; it is identical to the first line except for the multiplication by a factor of 2; indeed:

$$P\{|X| > x\sigma\} = P\{X < -x\sigma \ \cup \ X > x\sigma\} = P\{X < -x\sigma\} + P\{X > x\sigma\} = 2P\{X > x\sigma\}$$

In most applications the first line is of greater interest because one is interested in deviations of a well-defined sign. In particle physics, the conventional threshold of significance is 5σ; in the social sciences it is rather 2σ. These rules are nothing but conventions which are roughly in relation with the total number of events. In an experiment producing 10^{30} events it would make sense to take 10σ as the threshold of significance. The numbers in the table which are not given in standard textbook tables (e.g. Ventsel 1973) have been computed by using an asymptotic approximation of the complementary error function, namely $\sqrt{\pi/2}x \exp x^2/2$ where x denotes the signal to noise ratio.

Table 3.3a Is there a Diana-Werther effect? Observation of single events

Name	Date of death	Observed suicides (s)	$s_{m,y-1}$	$s_{m,y+1}$	Expected suicides (e)	p $\left(\frac{s-e}{e}\right)$ $(\%)$	$\sigma(p)$ $(\%)$	Signal to noise ratio
U.S.								
Marilyn Monroe	6 Aug. 1962	1838	1579	1801	1690	8.8	4.92	1.79
James Forrestal	22 May 1949							
$m =$ May		1549	1455	1600	1527.5	1.41	4.92	0.29
$m =$ June		1567	1410	1507	1458.5	7.44	4.92	1.51
U.K.								
Lady Diana	31 Aug. 1997							
M+F		420	381	381	381	10	11	0.91
F		103	96	105	100.5	2.5	16	0.16
M		317	285	276	280.5	13	12	1.08
M,15-24		43	37	31	34	26	38	0.68

Notes: The third column of the table gives $s = s_{m,y}$, the monthly number of suicides (in month m of year y) in the month in which the death of the celebrity occurred. The 6th column gives the expected number of suicides $e = [s_{m,y-1} + s_{m,y+1}]/2$. The variable p represents the deseasonalized number of suicides. For a stationary time series the average of p over several years can be expected to be close to zero; this is indeed verified for the series under consideration; consequently, in a given month p quantifies the percentage of excess-suicides due to exceptional events. $\sigma(p)$ denotes the standard deviation of p in the time series under consideration. In the case of James Forrestal we have shown the data for two successive months because he committed suicide toward the end of May. The lines labelled F and M show the suicide figures for females and males respectively. In the U.S. the monthly suicide data do not make a distinction between males and females, but such data are available for the U.K. It turns out that the Diana effect is much stronger for men than for women. The last line shows that p is even larger for young men but in this case the smaller numbers of suicides results in a larger standard deviation with the consequence that the signal to noise is not much improved.

Sources: U.S. data: Vital Statistics of the United States, yearly volumes, Grove and Hetzel (1968); British data: personal communication from Ms. Anita Brook (U.K. Office for National Statistics) to whom I express my grateful thanks.

Table 3.3b Is there a Diana-Werther effect? Observation of several events or places

		p $\left(\frac{s-e}{e}\right)$ (%)	$\sigma(p)$ (%)	Signal/noise ratio
	Several events			
1	33 celebrities	2.44	0.86	2.84
	Several places			
2	Forrestal, 17 states	2.32	6.55	0.35
	Lady Diana, 3 regions			
3	M	21	10	2.10
4	F	8.3	22	0.38

Notes: As in previous tables, p represents the deseasonalized monthly number of suicides, a variable whose average over several years is almost equal to zero. In the Diana case the three regions are England, Wales and Scotland. The comparison of cases 3 and 4 confirms that the effect is stronger for males than for females. The biggest improvements in the signal to noise ratio are cases 1 and 3.
Sources: Same as in Table 3.3a.

Table 3.4 Excess-suicides in the months following the event

	Month -1 (%)	Month 0 (%)	Month 1 (%)	Month 2 (%)	Month 3 (%)
p					
Monroe, M+F, (%)	-2.5	8.8	7.0	3.8	2.5
Diana, M, (%)	-11	13	5.9	16	21
$p/\sigma(p)$					
Monroe, M+F	0.51	1.79	1.42	0.77	0.51
Diana, M	-0.92	1.1	0.49	1.33	1.75

Notes: The table gives the deseasonalized number of suicides $p = (s-e)/e$ and the ratios $p/\sigma(p)$ in the months before and after the death of a celebrity. Month 0 is the month in which the death of the celebrity occurs. Month -1, the month preceding the death, is shown for the purpose of verifying that there is no overall trend. The fact that in the Diana case p increases in months 2 and 3 is probably due to a yet unidentified exogenous shock.
Sources: Same as in Table 3.3a.

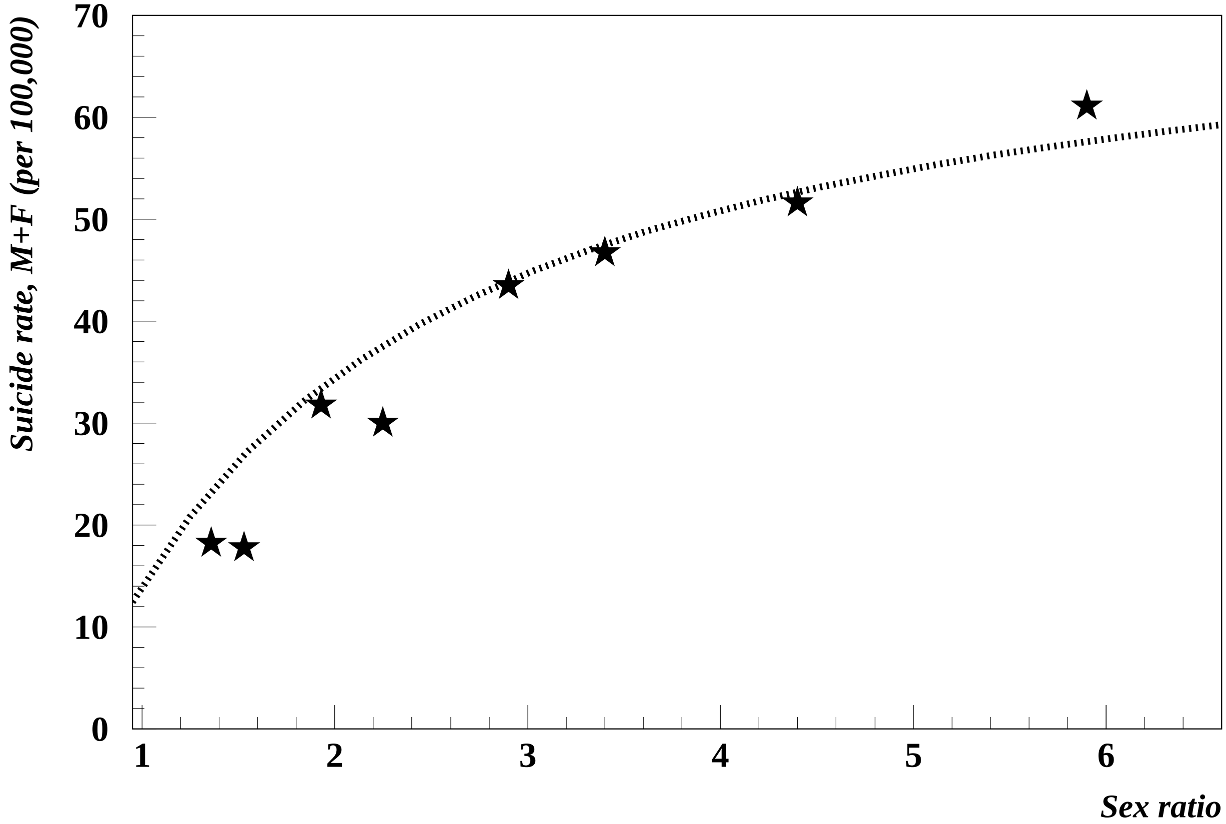

Fig. 3.1 Suicide rate in the Chinese community of the United States, 1923-1960. Horizontal scale: Male to female sex ratio; the high ratios on the right-hand side correspond to the 1920s, while the low ratios on the left-hand side correspond to the 1950s. The data refer to the continental U.S. which means that Hawaii (where there is also a substantial Chinese community) is excluded. As there were only about 30 suicides annually, we performed an average over two successive years. The suicide rate decreases with the sex ratio in the expected way (more details can be found in a subsequent chapter). Note that: (i) The rest of the U.S. did *not* experience a steady decline in suicide rate; for instance there was a strong increase between 1923 and 1932 and a smaller increase between 1945 and 1950. (ii) Most of the Chinese people who where in the U.S. in the early 1920s had been there for several decades as can be seen from the fact that the Chinese population in the U.S. reached a maximum in 1890 (107,000) and decreased steadily in subsequent decades until after 1940. *Sources: Mortality Statistics, annual reports 1923-1937, various years. Vital Statistics of the United States, 1938-1960, various years. These volumes are available on line on the website of the National Center for Health Statistics.*

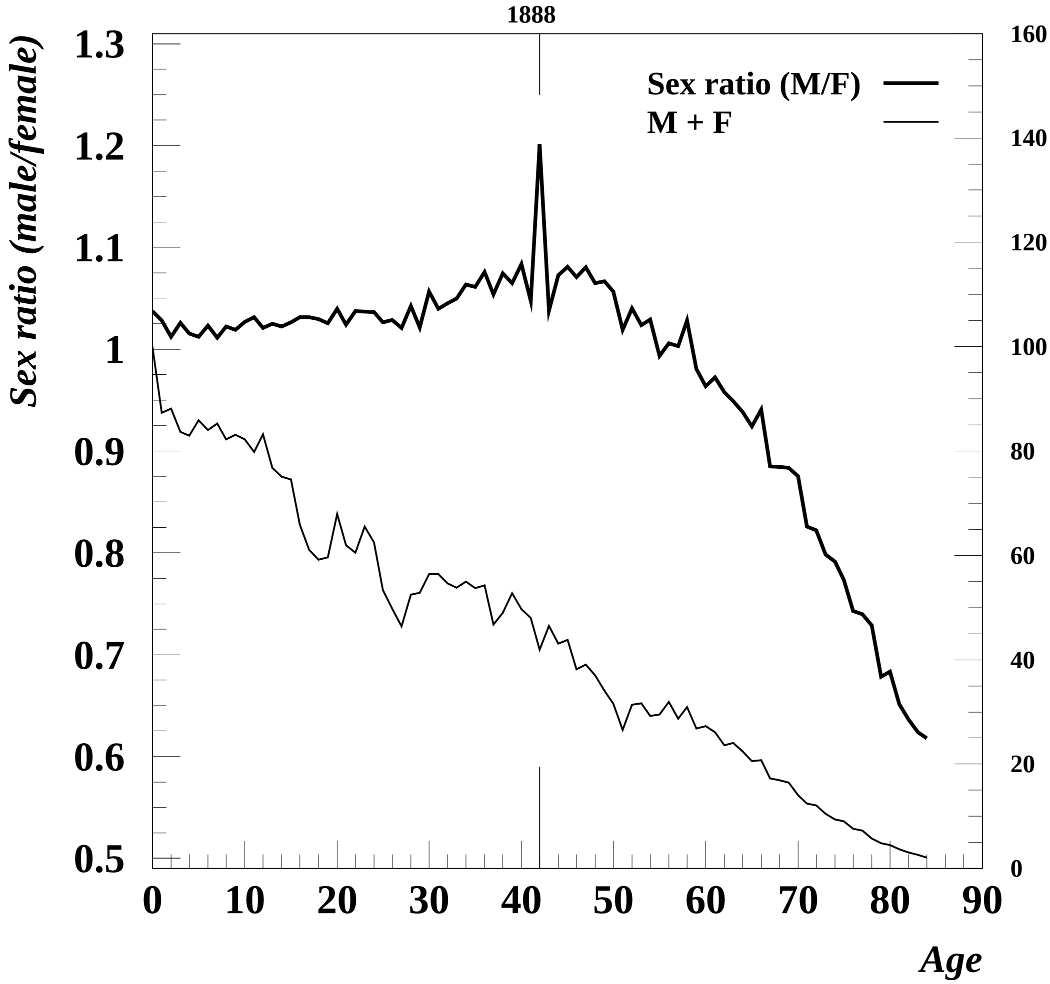

Fig. 3.2 a Sex ratio and population by age in Japan (1888). In this graph (as well as in Fig. 3.2 b,c) the thick line corresponds to the male/female sex ratio (left-hand side scale), the thin line to the total population normalized to 100 at age 0 (right-hand scale) and the thin vertical lines indicate the generations born in a Fire Horse year. Age 0 means aged less than one year. For the age group born during the Fire Horse year of 1846 there is a 20% increase in the sex ratio and a 11% fall in the total population. *Sources: Same as in Table 3.1.*

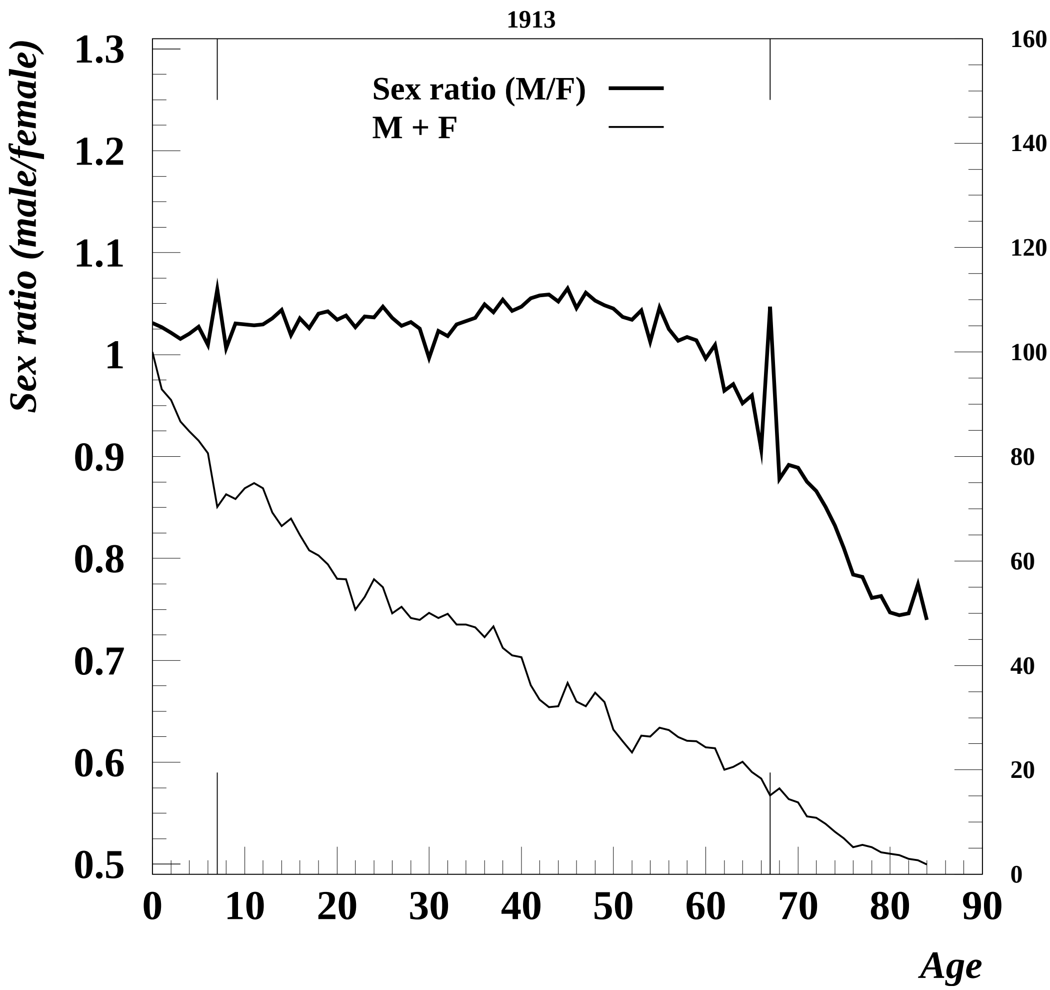

Fig. 3.2 b Sex ratio and population by age in Japan (1913). For the generation born in 1906 the sex ratio spike has an amplitude of 5% and the fall in population is 11%. *Sources: Same as in Table 3.1.*

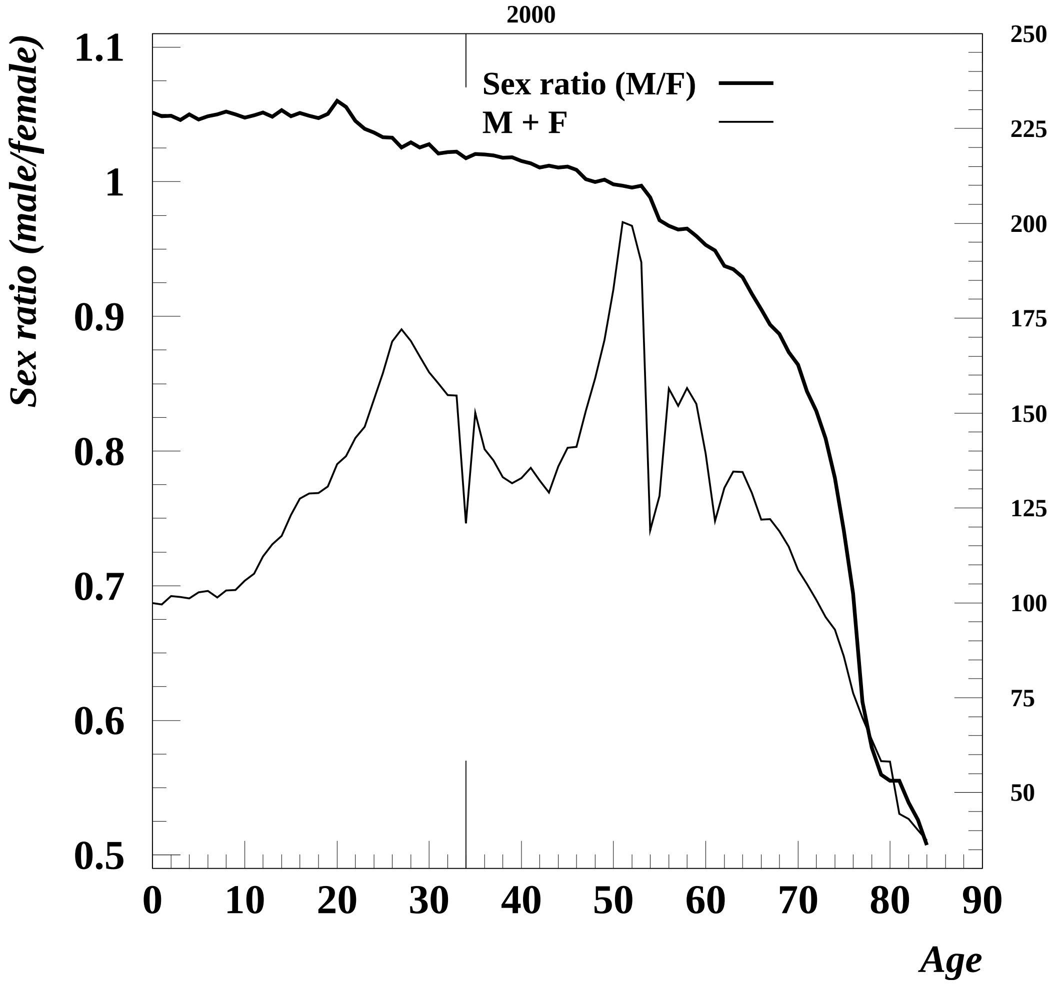

Fig. 3.2 c Sex ratio and population by age in Japan (2000). The Fire Horse year of 1966 lead to a sharp trough in births but, in contrast to 1846 and 1906, the sex ratio remained almost normal; more precise figures are given in Table 3.1. *Sources: Same as in Table 3.1.*

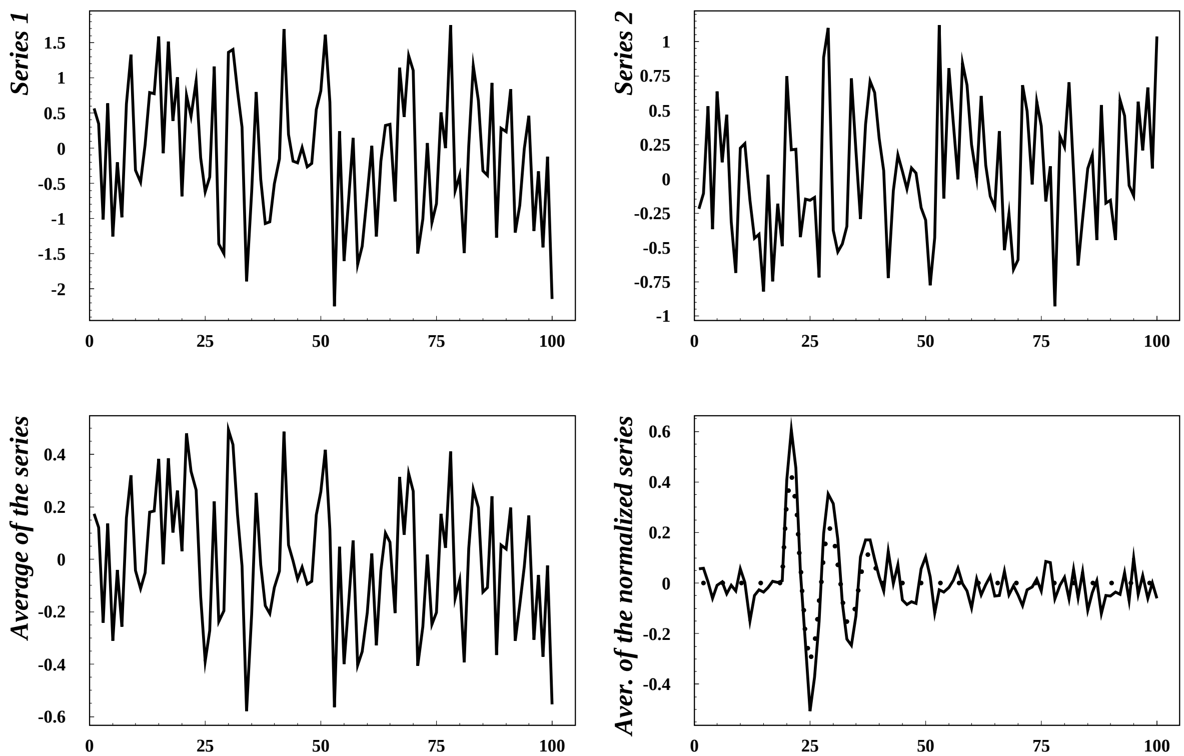

Fig. 3.3 Average of negatively correlated time series A pseudo-periodic deterministic signal is hidden in series 1 and 2; the challenge is to extract it. The two series have zero mean, correlation of -0.96 and standard deviations of 0.97 and 0.49 respectively. Taking the average of the two series (panel 3) reduces the level of noise but not enough to make the deterministic signal clearly visible. However if we take the average after dividing the series by their standard deviation, one expects the level of noise to be cut drastically. This is indeed what happens: the averaging process reveals the deterministic signal almost in its initial shape (which is represented by the dotted curve). It can be noted that the technique of computing the autocorrelation function which is often used to reveal hidden periodicities does not work when applied to the series in panel 3: the pseudo-periodic component is just too small. More generally the same procedure works for n series containing a common deterministic component provided that their average cross-correlation is close to $-1/(n-1)$.

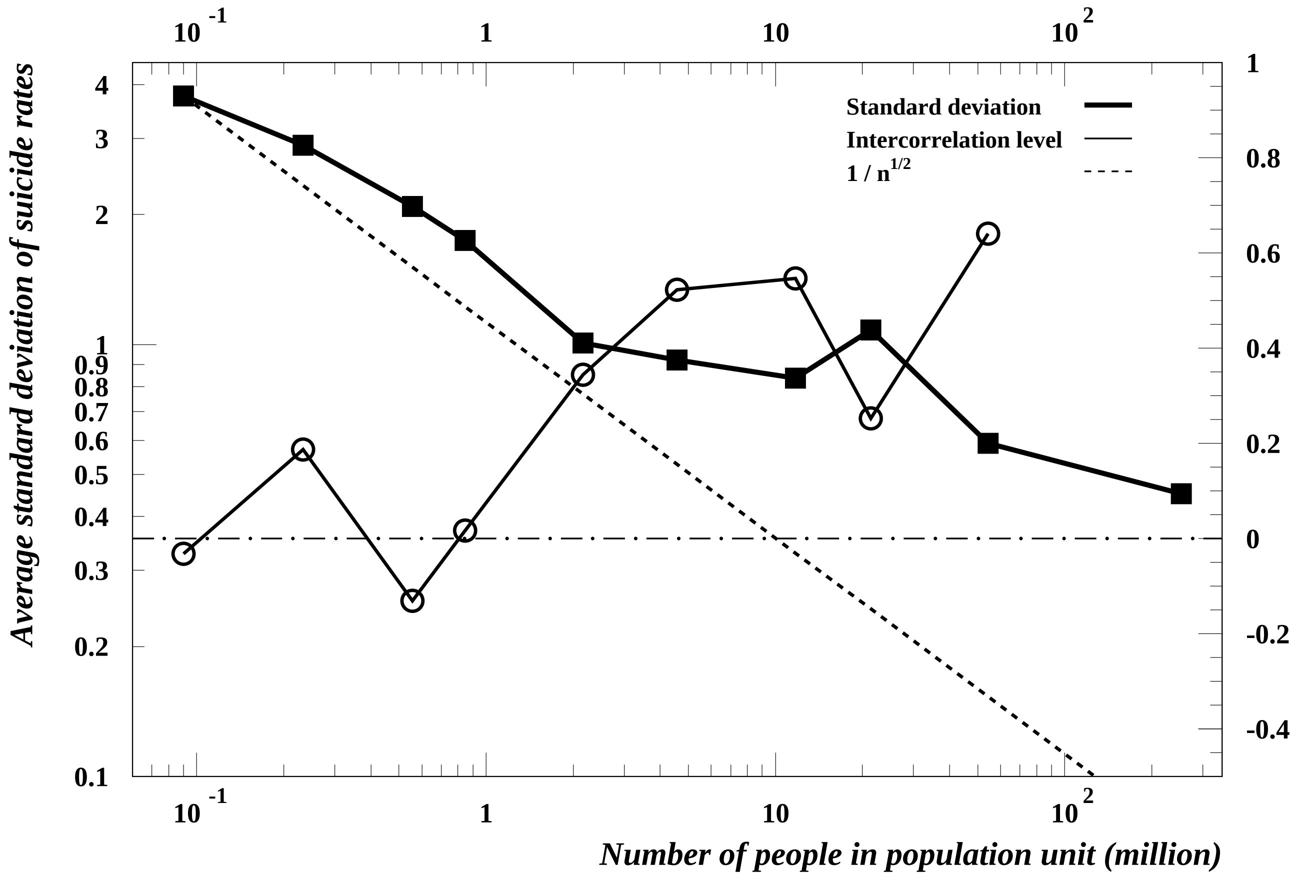

Fig. 3.4 Standard deviation of suicide rates in the U.S. The graph represents the standard deviation of suicide rates as a function of the population n of population units of increasing size. The intercorrelation curve represents the average inter-correlation of different units in the same size group. The broken curve represents the function $y = 1/\sqrt{n}$. A possible reason explaining why the standard deviation breaks away from the curve $1/\sqrt{n}$ is explained in the text. *Source: Wonder database on the website of the Center for Diseases Control.*