Revised and extended from Chapter 2 of Cobb & Thrall (eds) **Mathematical Frontiers of the Social and Policy Sciences**. Westview Press, 1981.

Stochastic Differential Equations for the Social Sciences

by Loren Cobb

Abstract Stochastic differential equations are rapidly becoming the most popular format in which to express the mathematical models of such diverse areas as neural networks, ecosystem dynamics, population genetics, and macroeconomic systems. It seems only a question of time before the social sciences begin to rephrase their dynamic models in these terms, and indeed, in some isolated cases the process has already begun. This paper is designed to introduce social scientists to the fundamental concepts and uses of stochastic differential equations, as applied to several models of current interest: linear feedback, epidemics, and the cusp catastrophe. The approach taken here is to focus on the stationary probability density function of the model, which avoids most of the advanced mathematical techniques that are ordinarily required to describe the solutions of the equations. Further, it will be seen that this approach yields new statistical insights into the process under study, and in some cases even new descriptive statistics.

Introduction

The purpose of this paper is to acquaint social scientists with some of the ideas of the stochastic calculus, particularly in the area of stochastic differential equations. There are two messages here: one is that in many cases these techniques are surprisingly easy to apply — a year of calculus and a second of probability and statistics should be sufficient preparation. The second is that these techniques are remarkably powerful, and can be used immediately in the construction of social theories. Five detailed examples will be provided in this paper.

Although in the course of their college educations many social scientists are exposed to differential equations and their applications in physics and chemistry, they seldom use these ideas in the construction of dynamic social models. Why? Perhaps the most important reason is the strong stochastic component that is obvious in virtually all social processes. The quantitative variables used in the social sciences tend to vary over time in ways that are apparently part systematic and part random. This is true in such diverse areas as economic indicators, vital statistics, opinion polls, and individual attitudes, behaviors, and characteristics. This significant random component plays havoc with the most fundamental tool of differential calculus: the derivative. If x(t) stands for a quantitative variable at time t, we should like to explain its derivative with respect to time, dx/dt, in terms of some function of x and other variables. This is impossible if the trajectory of x(t) is so irregular that it is nowhere differentiable, and yet this is exactly the situation found in the social sciences! Several strategies have been used whenever this problem has been confronted. They are as follows:

1. Divide time into discrete steps and use difference equations.
2. Abandon the ordinary mathematical definition of the derivative and use a statistical definition.

An example of strategy 2 is the Langevin (1908) Equation for certain kinds of Brownian motion. The Langevin Equation is worthy of attention because it was the first attempt to construct a stochastic differential equation, and because it worked in spite of its relatively ad hoc nature:

$$dx/dt = -\beta x + f(t) \tag{1}$$

In this equation Langevin was trying to say that the acceleration (dx/dt) of a small particle suspended in a fluid is composed of a random force f(t) and a systematic (i.e. non-random) frictional force which is a linear function of the velocity of the particle (x). Unfortunately, even this simple equation is ill-defined as stated, since dx/dt does not exist even in the statistical (i.e. mean-square) sense for Brownian motion.

The discovery of mathematical principles adequate to the task of describing "random" systems is one of the outstanding achievement of twentieth-century mathematics. The list of major contributors to this theory includes the names of Bachelier, Einstein, Wiener, Kolmogorov, Levy, Feller and Itô. It was the latter who finally brought a measure of rigor to the theory of stochastic differential equations. As a result of this theory it is now possible to write down a wide variety of stochastic differential equations with confidence that what is written is not meaningless. Further, there is now general agreement on what constitutes a solution to these equations.

Social scientists may be interested to learn that the first mathematical treatment of stochastic differential equations (Bachelier, 1900) was concerned with the behavior of stock market prices. Bachelier's doctoral thesis, "Théorie de la Spéculation" anticipated Einstein's fundamental work in all major respects, although he did not receive credit for his work during his lifetime (Mandelbrot, 1978).

The Itô Formulation

Virtually all contemporary texts and papers on stochastic differential equations (SDEs) focus on the novel definitions and constructions that are necessary to prove that Itô's theory works. Lost amid the endless formulae are two important facts: for the scientific user the Itô SDEs are (a) easy to understand, and (b) easy to use. In order to extract these benefits from the Itô theory, however, it is essential to shift ones s attention away from the behavior of single systems whose behavior is described by a stochastic differential equation, and focus instead upon the statistical behavior of very large ensembles of such systems. Stated differently, it is more useful and informative to study the shape of a stochastic variable's probability density function (PDF) than it is to study the trajectories of individual cases of such a variable. This shift in perspective is conceptually difficult for scientists who are accustomed to ordinary differential equations, but it is perhaps relatively easier for social scientists who are accustomed to dealing with populations and statistical distributions.

At the heart of Itô's definition of a stochastic differential equation for a random variable x(t) is the specification of two functions. Intuitively described, the first, μ(x), specifies the expected rate of change in x(t), and the second, σ^2(x), specifies the "variance" of this rate of change. The word variance is in quotes here because it corresponds only loosely to the usual meaning of the term. The random input, w_t, is assumed (in the Itô formulation) to be what is technically known as a Wiener Process, a mathematical idealization of Brownian notion. This assumption is quite reasonable for simple models of a very wide class of stochastic phenomena in the physical, biological, and social sciences.

For those who would like precise definitions, the functions μ(x) and σ^2(x) can be defined in terms of the conditional expectations of a stochastic process X(t) as:

$$\mu(x) = \lim_{h \to 0} \frac{E\{X(t+h) - X(t) \mid X(t) = x\}}{h} \tag{2}$$

$$\sigma^2(x) = \lim_{h \to 0} \frac{E\{\frac{1}{2}[X(t+h) - X(t)]^2 \mid X(t) = x\}}{h} \tag{3}$$

In the field of population genetics these are referred to as the "drift" and "diffusion" functions, and this terminology will be used here.

Having described all of its parts, we can now exhibit the general form of an Itô stochastic differential equation:

$$dx_t = \mu(x_t)dt + \sigma(x_t)dw_t. \tag{4}$$

Comparing (4) with (1), we can see that the Langevin Equation is specified by $\mu(x) = -\beta x$, $\sigma^2(x) = 1$, and $dw_t = f(t)dt$. Thus in the Itô form, Langevin's equation would read:

$$dx_t = -\beta x_t dt + dw_t. \tag{5}$$

The change in appearance between (1) and (5) is slight and superficial, but there is a major difference in the mathematical machinery that can be brought to bear on (5). It is not the appearance of an equation that is important, but what can be done with it.

Social scientists whose grounding is stronger in statistics than in calculus may gain some insight into the meaning of (4) by comparing it with its approximation in discrete time:

$$\Delta x_t = \mu(x_t)\Delta t + \sigma(x_t)u_t, \tag{6}$$

where u_t is a sequence of random variables, assumed to be Normally distributed with zero mean and variance Δt, serially independent, and independent of x_t. The difference operator Δ has the usual meaning: $\Delta x_t = x_{t+\Delta t} - x_t$. Thus the discrete-time Langevin Equation would read

$$\Delta x_t = -\beta x_t \Delta t + u_t,$$

which is a first-order linear stochastic difference equation. The time series produced by this equation would be called "first-order auto-regressive" by a statistician. Dynamic models based on equations of this type are now fairly common in economics, sociology, political science, and psychology.

It is not too difficult to show mathematically that if $\beta<0$ then the probability density function of x_t in the discrete-time Langevin Equation converges to the Normal density, no matter what its initial shape. Thus one possible explanation for the appearance of Normally distributed empirical data is that individual cases are executing a random walk (Brownian Motion) with a linear restoring force. Of course if $\mu(x)$ is anything other than linear, or if $\sigma^2(x)$ is anything other than constant, then the probability density to which the population converges will not be Normal. This implies that by examining the shape of an empirical frequency distribution we can deduce a substantial amount of information about the nature of the functions $\mu(x)$ and $\sigma^2(x)$, assuming that the population is in statistical equilibrium. Let us see how this is done.

The evolution of the probability density function for a variable which behaves according to a stochastic differential equation is described, necessarily, by a partial differential equation. This is because the probability density function $f(x,t)$ is a function of both x and t (time). Interestingly, the evolution of f over time is entirely determined by the functions $\mu(x)$ and $\sigma^2(x)$:

$$\frac{\partial f}{\partial t} = \frac{\partial(\mu(x)f(x,t))}{\partial x} + \frac{\partial^2(\sigma^2(x)f(x,t))}{\partial x^2} \tag{7}$$

In some cases this equation, known as the Kolmogorov forward equation, can be explicitly solved. Generally, however, an explicit solution is not available. On the other hand, it is quite easy to discover the probability density function when it has reached its equilibrium state (i.e. when $\partial f/\partial t = 0$). It is given by Wright's (1938) Formula:

$$f(x) = \frac{\psi}{\sigma^2} \exp\left[\int_{-\infty}^{x} \frac{\mu(s)}{\sigma^2(s)} ds\right], \tag{8}$$

where the constant ψ is chosen so as to make $\int_{-\infty}^{+\infty} f(x)dx = 1$.

In a sense, Wright's Formula provides the *stationary distribution* for any process whose dynamics are described by an Itô stochastic differential equation. The value $f(x)dx$ is the probability that individual trajectories will pass through the interval between x and $x+dx$, once the population ensemble has reached statistical equilibrium. It therefore yields exactly the kind of statistical information which is most valuable to the social sciences.

We will now work through a complete example of the application of Wright's Formula to a familiar problem: a linear feedback system responding to a randomly changing environment. Consider a system whose state is described by a single variable x_t, with a "goal" G towards which it moves, given no exogenous disturbance. Because it is linear, its expected rate of change is proportional to its deviation from G:

$$\mu(x_t) = r(G - x_t),$$

where r > 0 is the proportionality constant. Suppose further that the random disturbance to this system has a constant diffusion function. Thus in Itô's formulation $\sigma^2(x) = \varepsilon$, where ε is a constant. Putting all this together, we get the Itô SDE:

$$dx_t = r(G - x_t)dt + \sqrt{\varepsilon}dw_t \tag{9}$$

What is the probability density function of x when statistical equilibrium is reached? To find out, we apply Wright's Formula, as follows:

$$f(x) = \frac{\psi}{\varepsilon} \exp\left[\int_{-\infty}^{x} \frac{r(G-s)}{\varepsilon} ds\right]$$

$$\Rightarrow f(x) = \frac{\psi}{\varepsilon} \exp\left[-\frac{(x-G)^2}{2\varepsilon / r}\right]. \tag{10}$$

We see that the stationary probability density function is nothing other than the familiar Normal distribution, with mean G and variance ε/r. Thus the probability that this system is, at any given moment, a specified distance from its goal is given by a Normal probability distribution whose variance is proportional to the disturbance variance and inversely proportional to the strength of the feedback system. Wright's formula describes the stationary distribution of the Itô stochastic differential equation. Note that it was found by integrating a linear function within Wright's Formula: not exactly difficult. *One of the central points of this essay is that finding the stationary distribution of an Itô SDE involves a simple exercise in ordinary integration of the sort found in any introductory calculus texts.* Thus the formidable mathematical machinery of the stochastic calculus may be finessed by those who are willing to settle for the stationary probability density function of a system whose dynamics are described by a stochastic differential equation.

In this paper we shall examine in some detail five examples of stochastic differential equations. The first three are linear, and will cover the forms that show the most promise for use in the social sciences. The fourth, a quadratic, has an interesting application in social epidemiology. The fifth, a cubic, is the so-called "cusp catastrophe" of Thom & Zeeman. In each case we shall proceed from a specification of the stochastic differential equation to an application of Wright's Formula. Examination of the resulting probability density functions will be seen to yield valuable statistical information concerning the process itself.

Linear Stochastic Systems

A linear stochastic differential equation is one for which the drift function $\mu(x)$ is linear in x, *regardless of the diffusion function $\sigma^2(x)$*. In many cases the appropriate form for $\sigma^2(x)$ can be deduced from elementary knowledge about the range and behavior of the variable x(t). For example, if it is known that the size of the fluctuations in x(t) do not depend upon the size of x(t), then $\sigma^2(x)$ must be constant. Thus $\sigma^2(x) = \varepsilon$, where ε is a small positive constant, specifies one of the principal types of stochastic differential equations. By contrast, the fluctuations in some variables do depend on their levels. For example, the variance of day-to-day stock market prices is proportional to the stock prices: the higher the price the larger the fluctuations. For such variables the form $\sigma^2(x) = \varepsilon x$ is appropriate, and this specifies another principal type of SDE. A third distinct type is seen in public opinion polls: the month-to-month variance in the fraction x of the population that supports a country's leader is proportional to $x(1-x)$. Thus the fluctuations are largest when x is close to 50%. For such variables the form $\sigma^2(x) = \varepsilon x(1-x)$ is appropriate. These distinctions important, because the shape of the stationary probability density function depends in several interesting ways upon the diffusion function $\sigma^2(x)$.

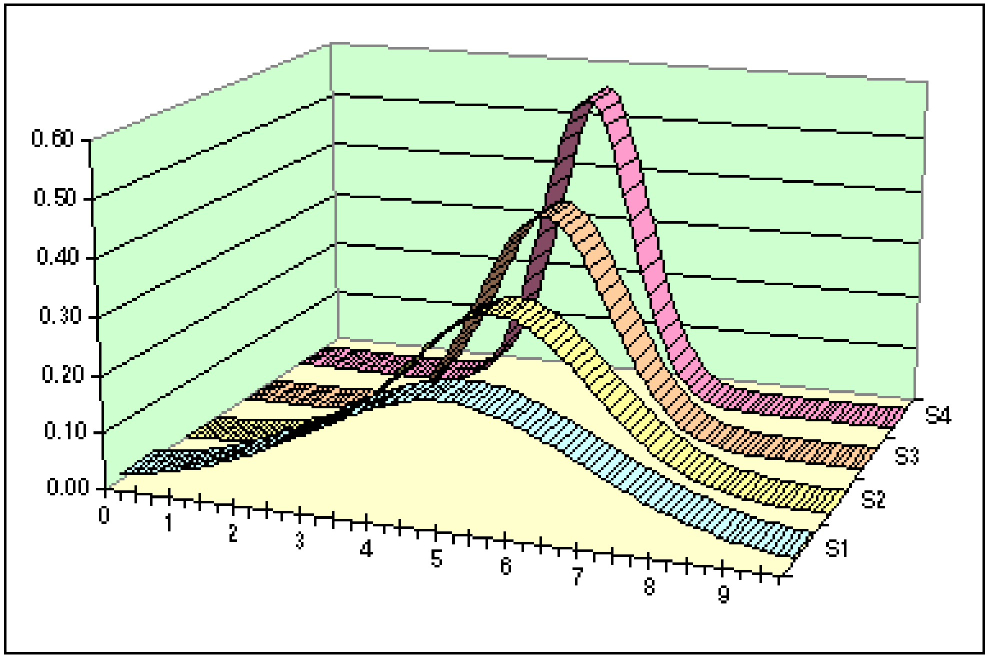

Figure 1. Stationary probability density functions for four linear feedback models of **Type N**, each with the same mean. Note that these are "**Normal**" probability densities. The characteristics of the **Type N** model are as follows:

Drift: $\mu(x) = r(G-x)$, $(r > 0)$.

Diffusion: $\sigma^2(x) = \varepsilon$, $(\varepsilon > 0)$.

SDE: $dx_t = r(G - x_t)dt + \sqrt{\varepsilon}dw_t$.

PDF: $f(x) = \frac{1}{2\pi\delta}\exp\left[-\frac{(x-G)^2}{2\delta}\right]$, where $\delta = \varepsilon/r$.

Statistics: Mean = G, mode = G, variance = δ.

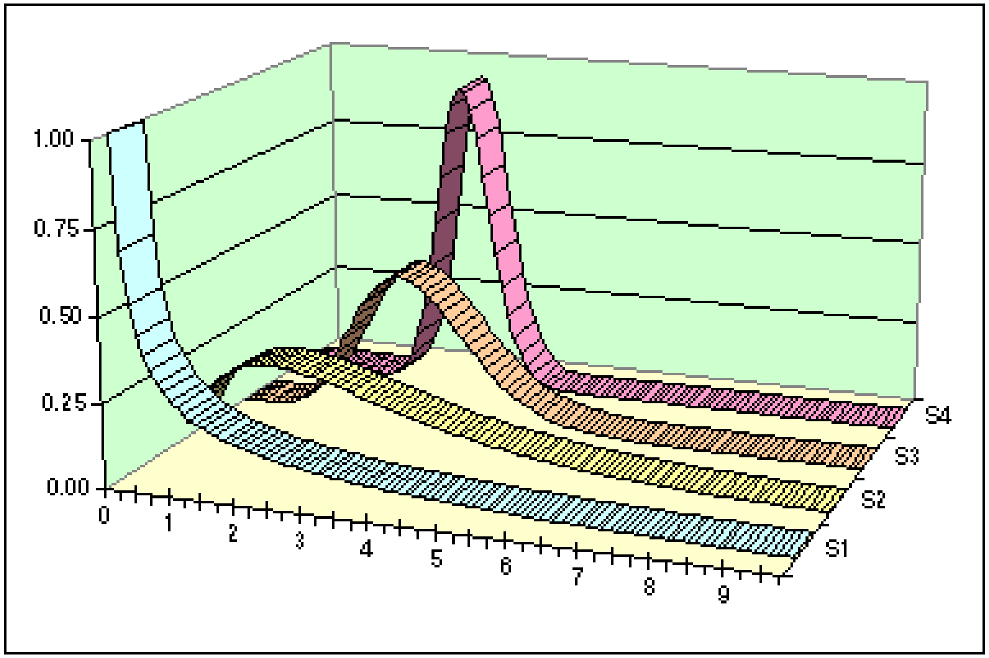

Figure 2. Stationary probability density functions for four linear feedback models of **Type G**, each with the same mean. Note that this is the "**Gamma**" probability density. The characteristics of the **Type G** model are as follows:

Drift: $\mu(x) = r(G-x)$, $(r > 0)$.

Diffusion: $\sigma^2(x) = \varepsilon x$, $(\varepsilon > 0)$.

SDE: $dx_t = r(G - x_t)dt + \sqrt{\varepsilon x_t}\, dw_t$.

PDF: $f(x) = (x/\delta)^{-1+G/\delta} e^{-x/\delta} \big/ \Gamma(G/\delta)$, where $\delta = \varepsilon/r$.

Statistics: Mean = G, mode = $G-\delta$, variance = δG.

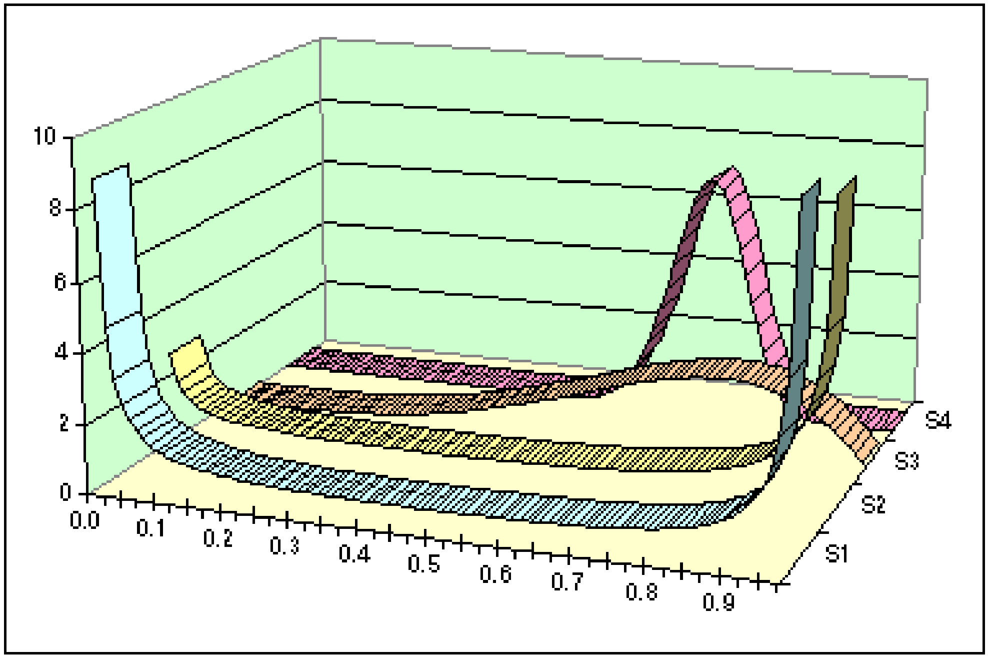

Figure 3. Stationary probability density functions for four linear feedback models of **Type B**, each with the same mean. Note that these are "**Beta**" probability densities. The characteristics of the **Type B** model are as follows:

Drift: $\mu(x) = r(G-x)$, $(r > 0)$.

Diffusion: $\sigma^2(x) = \varepsilon x(1-x)$, $(\varepsilon > 0)$.

SDE: $dx_t = r(G - x_t)dt + \sqrt{\varepsilon x_t(1-x_t)}\,dw_t$.

PDF: $f(x) = \dfrac{\Gamma(\frac{1}{\delta})}{\Gamma(\frac{G}{\delta})\Gamma(\frac{1-G}{\delta})} x^{-1+G/\delta}(1-x)^{-1+(1-G)/\delta}$, where $\delta = \varepsilon/r$.

Statistics: Mean = G, mode = $\frac{G-\delta}{1-2\delta}$, variance = $\frac{G(1-G)}{1+\delta}$.

Let us now examine these three major types in the case of linear feedback. In other words, we shall assume that $\mu(x) = r(G-x)$ in each of these types:

Type N: $\sigma^2(x) = \varepsilon$,

Type G: $\sigma^2(x) = \varepsilon x$,

Type B: $\sigma^2(x) = \varepsilon x(1-x)$.

We have already seen how to use Wright's Formula for Type N. In general, the stationary distributions are as shown on the preceding three pages. Note that in each case the mean of the distribution is the same, but the mode and variance depend upon the diffusion function, $\sigma^2(x)$.

Linear SDEs have potential applications throughout all the behavioral sciences. A particularly interesting application of the Type B system comes from a model of the process of political polarization. Let x_t be a person's political persuasion on the liberal-conservative dimension, where $x = 0$ is an extremely "liberal" conviction and $x = 1$ is an extremely "conservative" conviction. Suppose further that there is a general tendency to move toward the average persuasion of the whole population, but that people who hold extreme views are much less subject to random fluctuations than are those near the center. Thus $\sigma^2(x) = \varepsilon x(1-x)$, and

$$dx_t = r(G-x_t)dt + \sqrt{\varepsilon x_t(1-x_t)}dw_t \tag{11}$$

describes the motion of each person through the political spectrum. Note that in statistical equilibrium the average of x will be G, no matter what values r and ε have. The coefficient r in equation (11) expresses the strength of the general tendency to conform: we may reasonably assume that this is a constant. On the other hand, the coefficient ε measures the amount of random change occurring in political persuasion, and this is certainly large during times of unrest and small during times of political tranquility. What happens to the shape of the distribution of political persuasion as ε changes? Figure 3 shows this distribution for $G = 0.7$, $r = 1.0$, and four different values of ε (which increase from back to front). Note that as ε increases the political consensus breaks down and a polarization of opinion occurs.

Of course, the same effect would occur if ε were constant while r declined: the two coefficients have inverse effects. This is plain from the probability density function for this model:

$$f(x) = \psi x^{-1+rG/\varepsilon}(1-x)^{-1+r(1-G)/\varepsilon} \tag{12}$$

Thus it is the *ratio* ε/r which controls the extent of polarization in this model. Since the probability density function (12) is the well-known Beta density, it is easy to show that the ratio $\delta = \varepsilon/r$ can be calculated directly from the mean μ and variance σ^2, as follows:

$$\delta = \frac{\sigma^2}{\mu(1-\mu)-\sigma^2} \tag{13}$$

If we denote this polarization ratio by $\delta = \varepsilon/r$, then it can be shown that the Beta density is bipolar when $\delta > 0.5$. Perhaps this new polarization statistic will prove useful in the social sciences. Two examples are shown in Figures 4 and 5.

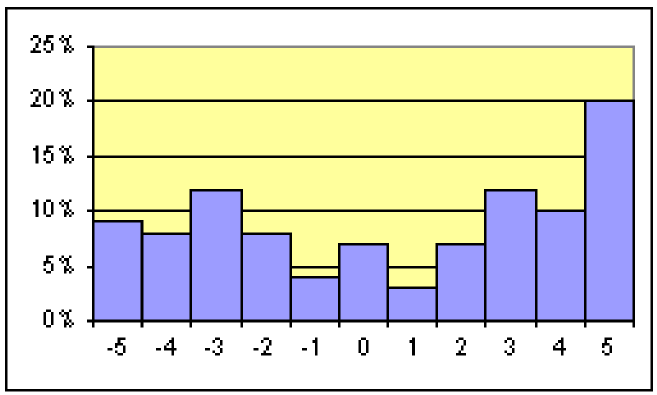

Figure 4. An example of a polarized public opinion histogram. Depicted are German attitudes towards a 1979 candidate for the presidency of West Germany (Carstens), a man who is alleged to have had connections with the Nazi party. The polarization statistic for this histogram is 0.7, indicating a moderate degree of polarization. Source: **Der Spiegel** (a German newsweekly), 21 May 1979, page 34.

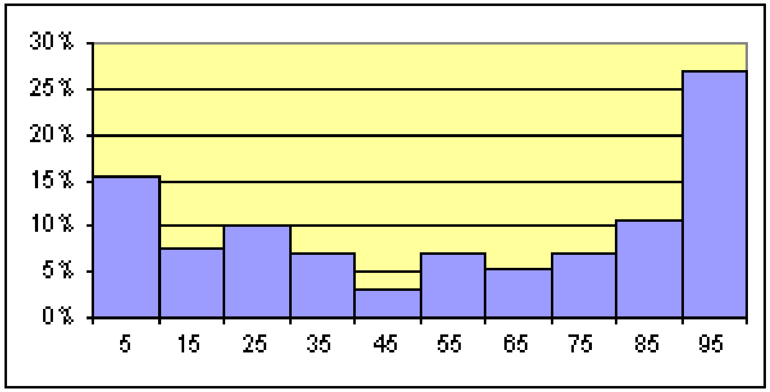

Figure 5. Strong polarization is visible in this histogram of 1970 literacy rates among 130 countries of the world. The polarization statistic is $\delta = 1.02$. Source: **The World Almanac 1978**.

It should be clear from these examples that linear SDE models have much to offer that ordinary linear models do not. In the first place, they yield explicit probability distributions which differ dramatically according to their diffusion functions. Secondly, parameters of these distributions can be estimated from empirical data by means of standard methods in each of the three cases discussed above. Thus these models are ideally suited to empirical research in the social sciences. Thirdly, a study of the Type B case produced a new statistic measuring the degree of polarization in a public opinion histogram.

Stochastic Epidemic Theory

The mathematical theory of epidemiology (Bailey, 1957) contains many types of models, both deterministic and stochastic. These have been used and expanded in the social sciences (e.g. Bartholomew, 1973) under the rubric of social epidemiology, which usually refers to the epidemiology of ideas, rumors, innovations, etc., rather than infectious diseases. In this section we shall compare a deterministic epidemic model, which is described by an ordinary differential equation, with a model described by the corresponding stochastic differential equation.

A fairly simple epidemic model can be constructed as follows: Let x(t) be the fraction of a population that has an infectious disease (or has heard a hot rumor, or whatever) at time t. If the disease does not confer immunity (e.g. gonorrhea), then the rate of change in x(t) usually follows an equation such as:

$$dx/dt = ax(1-x) - bx + c(1-x), \tag{14}$$

Where a = rate of person-to-person transmission, a > 0,
b = rate of recovery (or forgetting), b > 0,
c = rate of transmission from an external source, c > 0.

When started at the initial value x(0) = 0, this model yields a logistic trajectory which asymptotes at an equilibrium point, found by solving (14) for dx/dt = 0. A little calculation shows that this point is:

$$\hat{x} = d + \sqrt{d^2 + c/a}, \quad \text{where } d = \frac{a-b-c}{2a}.$$

To convert this model into a stochastic differential equation model, we have to introduce an assumption concerning the diffusion function $\sigma^2(x)$. The most reasonable assumption, as before, is that the random variation is greatest when x = 1/2, and least when x = 0 or 1. Thus we write

$$\mu(x) = ax(1-x) - bx + c(1-x),$$

$$\sigma^2(x) = \varepsilon x(1-x),$$

$$dx_t = \mu(x_t)dt + \sigma(x_t)dw_t. \tag{15}$$

The stationary probability density function f(x) for this model is found, as usual, by an application of Wright's Formula:

$$f(x) = \frac{\psi}{x(1-x)} \exp \int \left(\frac{a}{\varepsilon} - \frac{b}{\varepsilon(1-x)} + \frac{c}{\varepsilon x} \right) dx$$

$$= \psi x^{-1+c/\varepsilon}(1-x)^{-1+b/\varepsilon}e^{ax/\varepsilon} \tag{16}$$

Examples of this density are depicted in Figure 6. The striking features of the density labeled S2 in Figure 6 are the reflecting boundary at 0, the antimode at x_1, and the mode at x_2. Intuitively, these features are due to a stochastic threshold effect: the epidemic is unlikely to "take off" unless more than x_1 of the people have the disease. To find the location of the mode (x_2) and antimode (x_1) in terms of the parameters (a,b,c,ε) we need only solve

$$\mu(x) - \frac{d}{dx}\sigma^2(x) = 0\,.$$

Letting d = (a–b–c+2ε)/2a, we find that the mode and antimode are given by this formula:

$$\hat{x} = d \pm \sqrt{d^2 - (\varepsilon - c)/a}$$

If there are two positive real solutions, the smaller is the epidemic threshold, while the larger is the most likely size of the epidemic. Note that if c > ε then the boundary at x = 0 is no longer "reflecting," and therefore an epidemic is guaranteed. Thus the stochastic threshold phenomenon is observed only when the transmission of infection from an exogenous source occurs at a very slow but still positive rate, (i.e. if 0 < c < ε). For example, it can be calculated that the epidemic threshold for model S2 in Figure 6 is at about 16% (i.e. if more than 16% of the population is infected then an epidemic is extremely likely), and the most likely size for the epidemic is 63%.

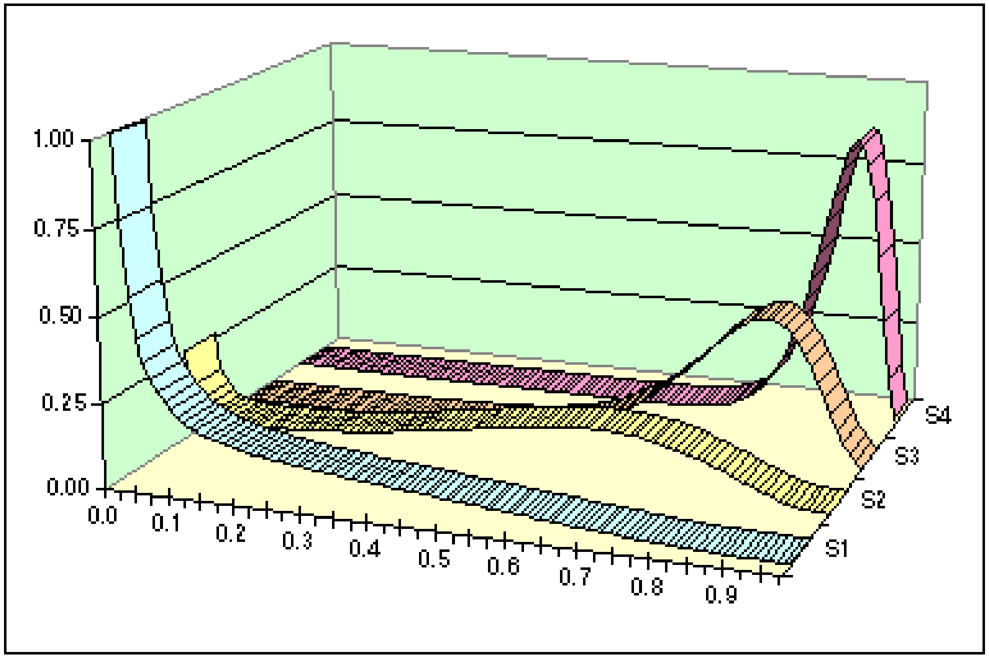

Figure 6. Stationary probability density functions for four stochastic epidemics, showing the effect of decreasing the person-to-person transmission rate from 4 (back) to 0.5 (front). The model used these parameter values:

a = 4.0, 2.0, 1.0, 0.5	(person-to-person transmission rate),
b = 0.40	(recovery rate),
c = 0.01	(external transmission rate),
ε = 0.10	(strength of random fluctuations).

The Stochastic Cusp Catastrophe

Contrary to appearances and even some published reports, the catastrophe theory has little or nothing to do with disasters. The field of catastrophe theory was so named by René Thom, its brilliant progenitor, merely to remind us of one of the qualitative features of its models: a capacity for sudden changes in state in response to gradual changes in exogenous controls. In reality, catastrophe theory is a cluster of topological concepts and theorems with which it is possible to classify functions based on their shape near their critical points. (Critical points, or singularities, occur when the first derivative of a function is zero.) The classification of nondegenerate singularities (minima, maxima, and saddle points) was accomplished by Morse, while the much more difficult task of classifying degenerate singularities (roughly, those points where the second derivative vanishes) was only recently achieved by a group of mathematicians including Thom, Mather, Malgrange, Smale, and Arnold. The catastrophe models which are developed in this classification exercise are not new, indeed some have been used throughout the sciences for centuries. However, they have attained new importance by virtue of their new role as principal members of their respective equivalence classes of functions with degenerate singularities.

A fierce controversy has arisen around catastrophe theory. The issue is not the mathematical theory, which is well accepted, but it is the applicability of the theory which has been seriously questioned. Precisely because of its topological nature, it is difficult to use catastrophe theory directly as a source for empirically testable models for the social sciences. Nevertheless, catastrophe theory has inspired a great number of attempted applications, some of dubious value. Catastrophe theory, the controversy, and many selected applications are reviewed at length for behavioral scientists in Cobb & Ragade (1978). More general treatments are also available: Haken (1978) and Poston & Stewart (1978) are particularly good. The collected papers of Zeeman (1977) may also be of interest to social scientists.

The difficulties that arise in attempts to construct empirically testable catastrophe models come from two sources, which can be heuristically stated as follows:

3. ***Plasticity:*** Two catastrophe models are topologically equivalent at a point x if there is a smooth (infinitely differentiable) transformation from one to the other, and a smooth inverse transformation. For example, at the point $x = 0$ the linear model $y = x$ is topologically equivalent to $y = \tan(x)$. The transformation is $x \rightarrow \tan(x)$, and the inverse transformation is $x \rightarrow \arctan(x)$. Both are smooth at 0.

4. ***Determinism:*** Catastrophe models are inherently deterministic, and therefore are not adequate for statistical use. Statistical models must include assumptions concerning the nature of stochastic effects.

Stochastic differential equations offer a way to render catastrophe models stochastic while at the same time preserving their topological character. Even more importantly, they point the way to statistical techniques for both parameter estimation and hypothesis testing: things that were not possible with ordinary applications of catastrophe theory.

The dynamics of a catastrophe model are usually written in terms of a ‘potential’ function: the system behaves as though it moves towards the point of lowest potential. Denoting the potential function by V(x), the dynamics are then

$$\frac{dx}{dt} = -\frac{\partial V}{\partial x}. \tag{17}$$

The singularities of V(x) are the points x for which $\partial V/\partial x = 0$, and by Equation (17) we see that these points are exactly the equilibrium points of the system. These equilibria are stable or unstable according to whether $\partial^2 V/\partial x^2$ is positive or negative, while the catastrophe points (degenerate singularities) are those values of x for which $\partial^2 V/\partial x^2 = 0$. (For the moment we are ignoring the possible influence of exogenous variables).

The simplest way to convert the differential equation (17) to stochastic form is to assume that there is a small additive stochastic driving term:

$$dx_t = -\frac{\partial V}{\partial x} dt + \sqrt{\varepsilon} dw_t. \tag{18}$$

Now applying Wright's Formula, we readily deduce the associated stationary probability density function:

$$f(x) = \psi \exp\left(-\frac{1}{\varepsilon}\int \frac{\partial V}{\partial x} dx\right)$$

$$= \psi \exp\left({}^{-V(x)}\!\!\diagup_{\varepsilon}\right). \tag{19}$$

As before, the constant ψ is chosen so that $\int f(x)dx = 1$. It is therefore a function of ε and of the coefficients of V.

Passing to logarithms in Equation (19), we find that

$$\log f = \log \psi - V/\varepsilon. \tag{20}$$

From this it is immediately evident that the singularities of log *f* correspond exactly to the singularities of V. Thus the entire panoply of topological theory which characterizes the singularities of V applies without change to log *f*.

Catastrophe theory achieves its classification by examining polynomial approximations of V in the neighborhood of degenerate singularities. In effect, a catastrophe model is just such a polynomial. For example, the 'cusp' catastrophe model specifies a quartic (fourth-degree) polynomial for V as the appropriate approximation. Cusp dynamics are described by cubic polynomials:

$$dx/dt = -(a_1 + a_2x + a_3x^2 + a_4x^3). \tag{21}$$

These polynomials can be reparametrized in an equivalent "standard" form as follows:

$$dx/dt = r(\alpha + \beta(x-\lambda) - (x-\lambda)^3), \tag{22}$$

where α and β are the so-called normal and splitting factors, respectively, of the standard cusp model. These terms were originated by Zeeman (1977). In Zeeman's models these factors are dependent upon exogenous variables. The equilibria of x are the values of x at which $dx/dt = 0$, i.e. the solutions of

$$0 = \alpha + \beta(x-\lambda) - (x-\lambda)^3.$$

The ***Type N cusp probability density function*** is derived from (22) by the usual conversion to Itô form and use of Wright's Formula:

$$f(x) = \psi \exp\left(\frac{\alpha(x-\lambda) + \frac{1}{2}\beta(x-\lambda)^2 - \frac{1}{4}(x-\lambda)^4}{\delta}\right) \tag{23}$$

where $\delta = \varepsilon/r$ as before. The variety of shapes of the cusp density are depicted in Figures 7 and 8. The four parameters of the cusp density, $(\alpha, \beta, \lambda, \delta)$, can be interpreted with the aid of ***Cardan's discriminant,*** $(C = 27\alpha^2 - 4\beta^3)$, as follows:

- **Asymmetry** (α) If $C < 0$ then the cusp density is bimodal and α determines the relative height of the two modes. If $C > 0$ then the cusp density is unimodal and α measures skewness.
- **Bifurcation** (β) If $C < 0$ then β determines the separation of the two modes, while if $C > 0$ then β measures kurtosis.

- **Location** (λ) The cusp catastrophe point is located at x = λ, with α = 0 and β = 0. Changing the value of λ simply moves the cusp density horizontally on the x-axis without changing its shape.
- **Dispersion** (δ) This parameter determines the amount of variation about the two modes of a bimodal cusp density in the same way that the variance determines the variation about the mode of a Normal density. It is *not* a scale parameter.

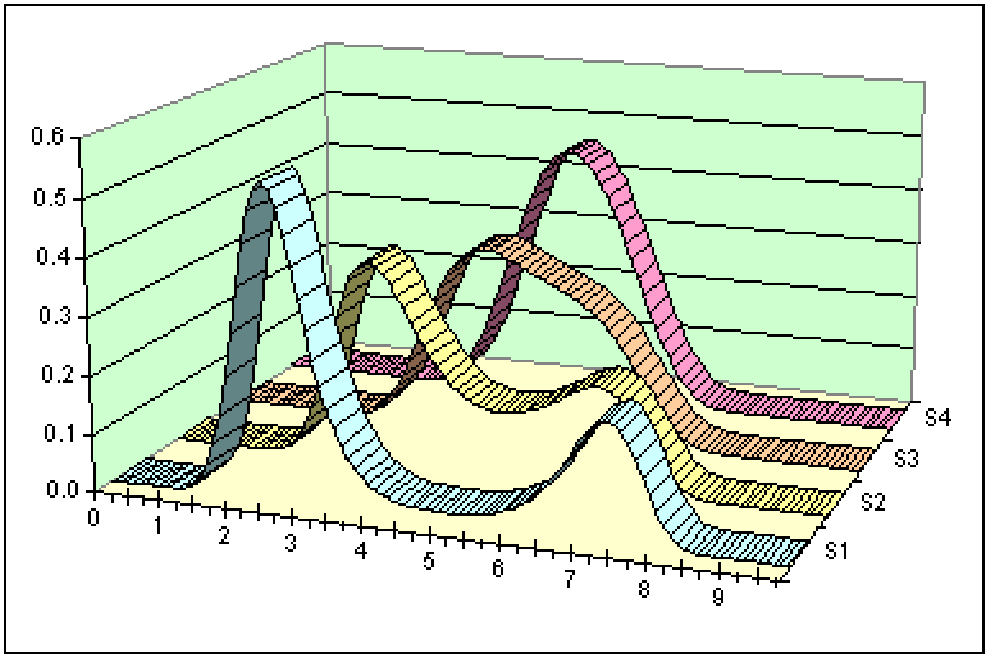

Figure 7. As β is varied from 2 (back) to –1 (front), the Type N cusp probability density function (Equation 23) changes from unimodal to bimodal. Notice that within the unimodal region β determines the kurtosis (flatness) of the pdf, while in the bimodal region it determines the degree of separation between the two modes. In this series the asymmetry parameter (α) and dispersion parameter (δ) have been held constant at –0.1 and 3.0, respectively. The location parameter (λ) is fixed at 5.0.

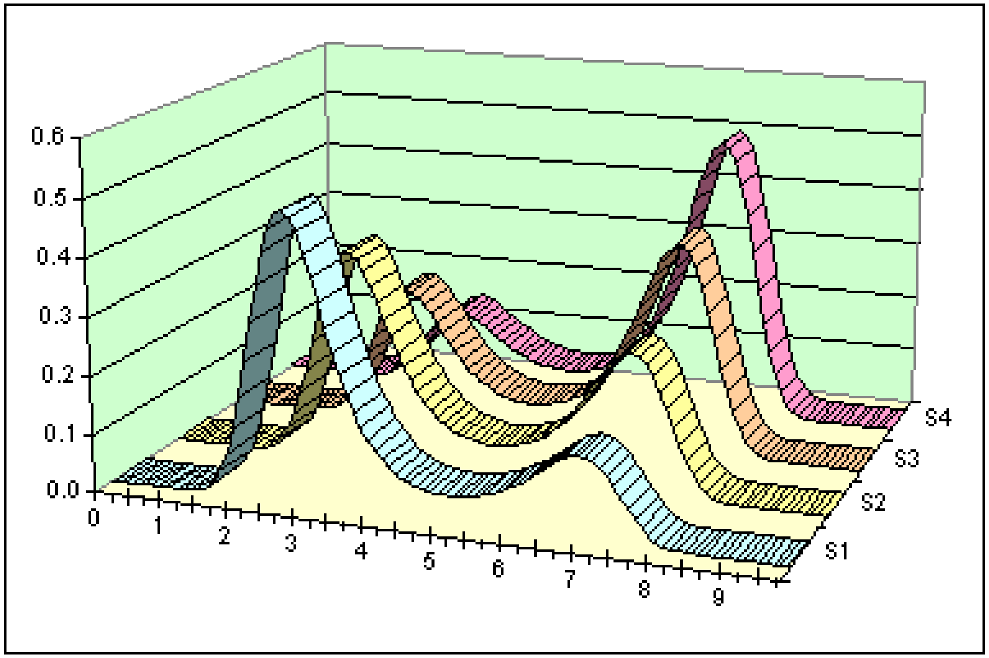

Figure 8. As α is varied from 0.3 (back) to –0.3 (front), the modes of the Type N cusp probability density function change in relative height. Thus α determines the skewness (asymmetry) of the pdf. In this series the bifurcation parameter (β) and dispersion parameter (δ) have been held constant at 1.5 and 3.0, respectively. The location parameter (λ) is fixed at 5.0.

In almost all published applications of the cusp catastrophe model the parameters α and β (and sometimes λ) are themselves dependent upon some exogenous variables. Suppose, for example, that there are m such exogenous variables, denoted z_1, z_2, …, z_m. Then a reasonable assumption would be that

$$\boldsymbol{\alpha} = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \ldots + \alpha_m z_m,$$

$$\boldsymbol{\beta} = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_m z_m, \tag{24}$$

$$\boldsymbol{\lambda} = \lambda_0 + \lambda_1 z_1 + \lambda_2 z_2 + \ldots + \lambda_m z_m,$$

where the $\{\alpha_0 \ldots \lambda_m\}$ are constant coefficients. When written this way, the parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\lambda}$ are called "factors."

The nature of the effects of the three factors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\lambda}$ on the modes of x can be seen in their generic forms in Figures 9, 10, and 11. The joint effects of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ on x are shown in Figure 12.

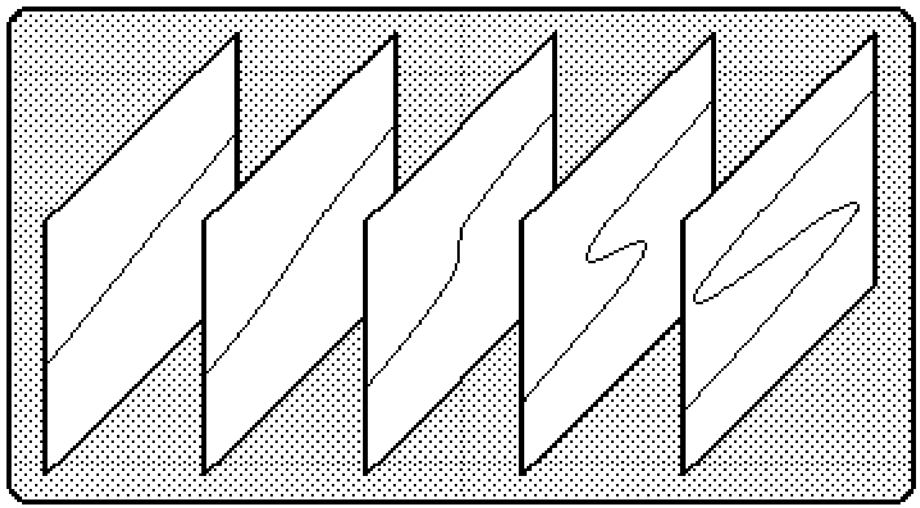

Figure 9. When the asymmetry parameter $\boldsymbol{\alpha}$ depends on an exogenous variable z, say $\boldsymbol{\alpha}(z) = \alpha_0 + \alpha_1 z$, then the equilibria of x depend on z as graphed above. The five panels show this relationship for five values of the bifurcation parameter β, ranging from negative on the left to positive on the right. Note the degenerate singularity in the center panel, where the graph rises vertically.

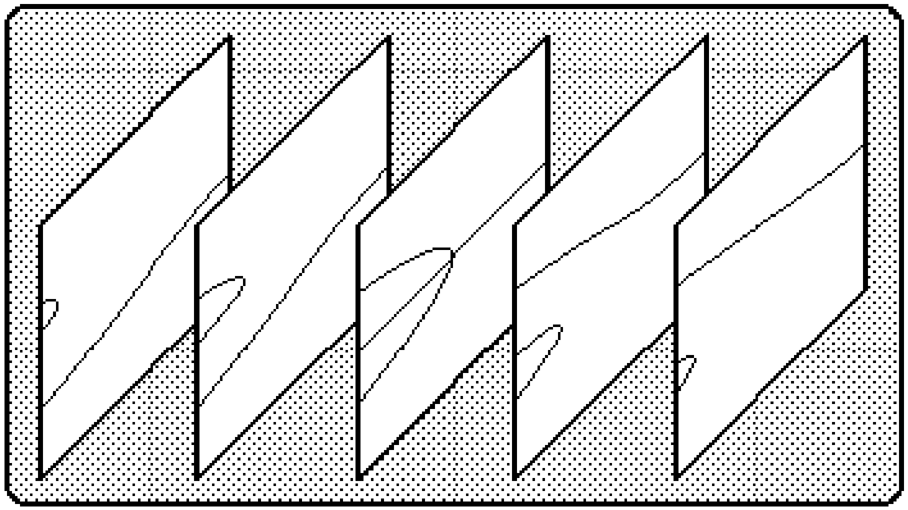

Figure 10. When the bifurcation parameter $\boldsymbol{\beta}$ depends on an exogenous variable z, say $\boldsymbol{\beta}(z) = \beta_0 + \beta_1 z$, then the modes of x depend on z as graphed above. The five panels show this relationship for five values of the asymmetry parameter α, ranging from negative on the left to positive on the right. Note the degenerate singularity in the center panel.

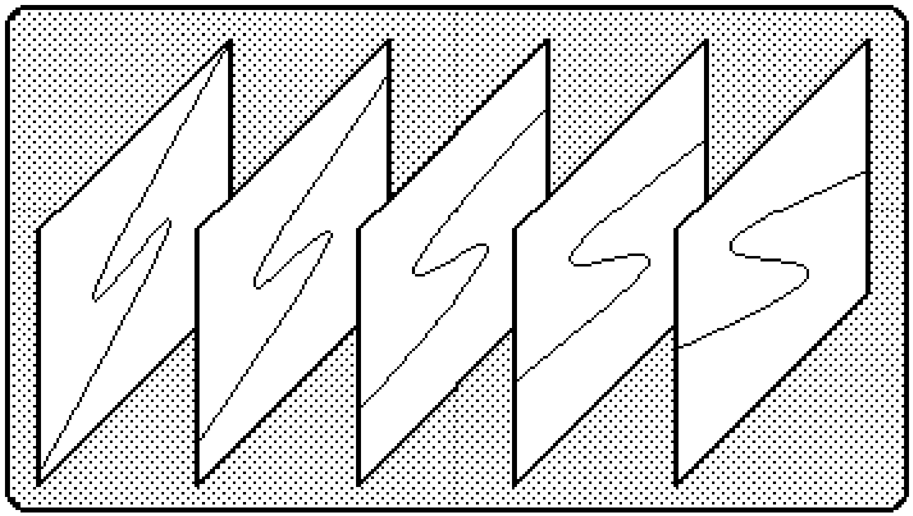

Figure 11. When the location parameter $\boldsymbol{\lambda}$ depends on an exogenous variable z, say $\boldsymbol{\lambda}(z) = \lambda_0 + \lambda_1 z$, then the modes of x depend on z as graphed above. The five panels show this relationship for five values of the asymmetry parameter α, ranging from negative on the left to positive on the right. The bifurcation parameter β is fixed at a postive value.

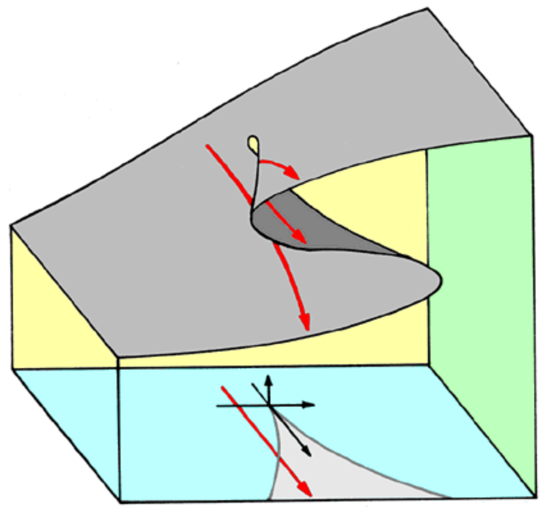

Figure 12. The 'standard' cusp model. This representation shows the joint effects of the asymmetry and bifurcation factors (**α** and **β**, respectively) on the dependent variable (x). The cusp point is the degenerate singularity that marks the beginning of the wrinkle in the graphed surface. The location factor (**λ**) is assumed to be a *constant* in this model. When **λ** is a *factor* (i.e. dependent upon one or more exogenous variables) then cusp surface will appear to be tilted with respect to the orientation shown above.

Cusp Statistics

The cusp probability density function derived in the previous section would be of little use to social scientists without some means for estimating its parameters. Fortunately, the cusp probability density (23) lends itself readily to statistical analysis. The key to a statistical approach to catastrophe theory is the observation that the cusp density is a multiparameter exponential family. Several well-known results from statistical theory apply to all such families:

1. Unique maximum likelihood estimators (MLE's) for the parameters exist, and may be found by the Newton-Raphson iteration technique.
2. The sampling distribution of the MLE's is asymptotically multivariate Normal.
3. Among all asymptotically Normal estimators, the MLE's have the minimum asymptotic variance.

4. Formal hypotheses concerning the MLE's may be tested through the use of the likelihood ratio criterion.

These results are very encouraging for the future of applied catastrophe theory, because they indicate that the full power of the maximum likelihood theory can be brought to bear on the cusp probability density function.

Conclusion

Stochastic differential equations appear to hold considerable promise for the social sciences. First, they provide a powerful way to express the stochastic and deterministic components of a model on an equal basis. This ability overcomes the major drawback of ordinary differential equations for use in the social sciences. Second, the stationary probability density function can be found, using Wright's Formula, which requires mathematical techniques found in any first course in integral calculus. In other words, it is relatively easy to derive the probability density function of the dependent variable when it has reached statistical equilibrium. Third, these stationary probability densities have many qualitatively interesting characteristics: bipolarity, nultimodality, etc. These characteristics are of theoretical interest in themselves, and have been generally overlooked by social scientists. Fourth and last, it may be seen from the discussion in this paper that a careful study of these densities yields new statistics with which to describe and analyse empirical data. The polarization statistic is one such, and the asymmetry and bifurcation factors of the cusp model are others.

[*First delivered as a presentation to the AAAS Annual Meetings in 1979.*]

[*Published as Chapter 2 of* ***Mathematical Frontiers of the Social and Policy Sciences****. Westview Press, 1981.]*

[*Last revised 26 August, 1999.*]

References

Bachelier, Louis. Théorie de la Spéculation. **Annales Scientifique de l'Ecole Normale Supérieure**, **3e Serie**, **Tome 17**, 1900, 21–86.

Bailey, Norman T. J. **The Mathematical Theory of Epidemics**. London: Griffin, 1957.

Bartholomew, David J. **Stochastic Models for Social Processes**, **2nd Edition**. New York: Wiley, 1973.

Cobb, Loren. Stochastic Catastrophe Models and Multimodal Distributions. **Behavioral Science**, 1978, 23, 360–374.

Cobb, Loren & Rammohan Ragade (eds). Applications of Catastrophe Theory in the Social and Biological Sciences: A Special Issue. **Behavioral Science**, 1978, 23, #5.

Gardner. Crispin W. **Handbook of Stochastic Methods**. Springer-Verlag, 1983.

Haken, Herrmann. **Synergetics**, **2nd Edition**. New York: Springer-Verlag, 1978.

Langevin, Paul. Sur la Théorie du Mouvement Brownien. **C. R. Acad. Sci.**, **146**, 1908, 530–533.

Mandelbrot, Benoit B. Fractals: **Form, Chance & Dimension**. San Francisco: Freeman, 1977.

Poston, Tim & Ian N. Stewart. **Catastrophe Theory and its Applications**. London: Pittman, 1978.

Wright, Sewall. The Distribution of Gene Frequencies under Irreversible Mutation. **Proc. Nat'l. Acad. Sci.**, **24**, 1938, 253–259.

Zeeman, E. C. **Catastrophe Theory: Selected Papers**. Reading, Massachusetts: Addison-Wesley, 1977.