Mathematical Modelling, Vol. 1, pp. 311–317, 1980
Printed in the USA.

0270-0255/80/040311-07$02.00/0

STATISTICAL CATASTROPHE THEORY: AN OVERVIEW

LOREN COBB

Medical University of South Carolina
Charleston, South Carolina 29403, USA

BILL WATSON

College of Charleston
Charleston, South Carolina 29401, USA

Abstract—This paper is a summary of an address given to the Conference on Frontiers of Applied Geometry. A stochastic version of catastrophe theory is presented, using stochastic differential equations. We show that there is a nontrivial relationship between the potential functions of the deterministic models and the stationary probability density functions of the stochastic models. In the second part of the paper, we use maximum likelihood theory to derive estimators for the stationary densities, and we demonstrate how to test statistical hypotheses for these models.

1. INTRODUCTION

Statistical catastrophe theory is a term that sounds inherently paradoxical. Statistical models do not, as a rule, contain degenerate singularities, and catastrophe theory is generally perceived as a purely deterministic branch of differential topology. However, one may restate catastrophe models in stochastic form, using stochastic differential equations. The resulting stochastic processes have stationary probability density functions which are of some topological interest in themselves. It will also be seen that these densities are, as a class, amenable to statistical analysis by the classical method of maximum likelihood. This enables the construction of algorithms for estimating the parameters of given models from data, and for testing statistical hypotheses concerning the existence of degenerate singularities within the data.

2. STOCHASTIC CATASTROPHE MODELS

Let $x(t)$ be the real-valued state variable of a (deterministic) system whose dynamics are controlled by a smooth potential function $U(x)$, so that

$$dx/dt = -\partial U/\partial x. \tag{1}$$

The *singularities* of U are those points for which $\partial U/\partial x = 0$. These singularities are said to be *degenerate* if $\partial^2 U/\partial x^2 = 0$. Catastrophe theory is a family of topological theorems which is useful for classifying these degenerate singularities and for describing the behavior of systems such as (1) in the neighborhood of such singularities [see 1,2,3 for details].

One way to render (1) stochastic is to introduce a white noise driving term, viz.,

$$dx = (-\partial U/\partial x)dt + \sqrt{v(x)}\cdot dw(t). \tag{2}$$

In this stochastic differential equation (SDE), the function $w(t)$ is to be understood as a standard Wiener process (idealized Brownian motion). This construction is the usual one within the stochastic calculus of Itô and Stratanovich [4]. The function $v(x)$ in (2) modulates the intensity of the random input $dw(t)$; it is called the *infinitesimal variance function* of the SDE. Let A, an interval of the real line, be the range of x (i.e., $x: T \times \Omega \rightarrow A \subset R$, where Ω is the sample space for the random variable x). Then $v(x) > 0$ for $x \in \text{Interior}(A)$, and $v(x) \geq 0$ for $x \in \text{Boundary}(A)$. Some common examples for the infinitesimal variance function are

(1) $v(x) = \epsilon, \quad A = (-\infty, +\infty),$

(2) $v(x) = \epsilon x, \quad A = (0,\infty),$

(3) $v(x) = \epsilon x(1-x), \quad A = (0,1).$

The last occurs very commonly in the theories of population genetics, but we shall primarily concern ourselves with the first two.

Let $f_x(u,t,x_0)$ be the probability density function

$$\frac{d}{du}\text{Prob}\{x(t)<u|x(0)=x_0\},$$

for the random variable x at time t, given an initial position x_0.

THEOREM (Itô) [4, chap. 7].

$$\begin{aligned} &\text{If } dx = m(x)dt + \sqrt{v(x)}\,dw, \\ &\text{then } \quad \partial f/\partial t = -\partial(mf)/\partial u + \tfrac{1}{2}\partial^2(vf)/\partial u^2. \end{aligned} \tag{3}$$

In effect, this theorem shows that the probability density of an SDE obeys a (deterministic!) partial differential equation. We shall find it convenient to focus our attention on the evolution of the density f, rather than on the individual trajectories of x.

Equation (3) is the *Kolmogorov forward equation* for the stochastic process. As $t \rightarrow \infty$, f converges to a stationary form, f^*, such that $\partial f^*/\partial t = 0$. This stationary density is either a generalized function (e.g., the Dirac δ-function, see [5]), or a proper probability density function, depending on the functions m and v. Upon solving (3) for f^* with appropriate boundary conditions, we obtain

$$f^*(x) = C\cdot\exp\,[2\int^x((m(s)-\tfrac{1}{2}v'(s))/v(s))ds]. \tag{4}$$

(See [4, p. 197] for details.) The function f^* given by (4) will be the theoretical probability density function for use in statistical catastrophe theory, in which case we shall identify $m(x)$ with $-\partial U/\partial x$. For example, suppose that the infinitesimal variance function is constant, e.g., $v(x) = \epsilon$. Then, since $m(x) = -\partial U/\partial x$, we have

$$\begin{aligned} f^*(x) &= C\cdot\exp[2\int^x(-\partial U/\partial s)ds/\epsilon] \\ &= C\cdot\exp[-2U(x)/\epsilon], \end{aligned}$$

which implies

$$\log f^* = \log C - 2U/\epsilon. \tag{5}$$

Thus, we see that $\log f^*$ is just an affine transformation of the potential function. Therefore, the entire apparatus of catastrophe theory, and, in particular, the classification of the degenerate singularities of U, now applies without change to $\log f^*$, the logarithm of the stationary probability density function of the stochastic version of (1). Of course, if the infinitesimal variance function is not constant, then the situation is not quite so straightforward (but perhaps is more interesting).

2. THE STATIONARY DENSITIES

If the infinitesimal variance function $v(x)$ is zero at one of the boundaries of the interval A, then the stationary density f^* generically has either a zero or a pole at the boundary. For example, if the potential function U is quadratic and v is proportional to x, then f^* has the form $x^a \exp(-bx)$ for $0 < x < \infty$. Thus, f^* has a pole or zero at zero depending on the sign of a, and is positive but finite at zero only if $a = 0$. We suggest that there is a need for a local classification of stationary densities near boundaries. One such classification exists [6], but it is not a topological classification.

In the interior of its domain, f^* has differentiable relative maxima (*modes*) and minima (*antimodes*). The modes and antimodes of f^* are related nontrivially to the relative minima and maxima of the potential function U. Only when the infinitesimal variance function is constant is the relationship simple: In this special case the modes and antimodes of f^* coincide with the relative minima and maxima of U, respectively. In all other cases it is necessary to refer to Eq. (4), which can be rewritten as

$$d(\log f^*)/dx = -2(\partial U/\partial x + \tfrac{1}{2}\partial v/\partial x)/v(x). \tag{6}$$

Define the *shape function* of f^* to be

$$g(x) = \partial U/\partial x + \tfrac{1}{2}\partial v/\partial x. \tag{7}$$

From (6) and (7) it can be seen that the modes and antimodes of f^* occur at the zeroes of $g(x)$, which do not necessarily coincide with the zeroes of $\partial U/\partial x$.

The *canonical catastrophe* potentials are universal unfoldings of Taylor polynomial approximations to potentials that have degenerate singularities [3]. For example, the potential $\frac{1}{4}x^4$ has a degenerate singularity at $x=0$. Its universal unfolding is $\frac{1}{4}x^4-\frac{1}{2}bx^2-ax$, the canonical cusp potential. The cusp potential has relative minima and maxima at $\{x: x^3-bx-a=0\}$. Now suppose that $v(x) = 2\epsilon(x-x_0)$, $\epsilon>0$. For this choice of infinitesimal variance function, the shape function for f^* is

$$g(x) = x^3 - bx - a + \epsilon, \ (x>x_0). \tag{8}$$

From this, the modes and antimodes of f^* are seen to be displaced away from the minima and maxima of U. Indeed, f^* may be unimodal even if U has two minima.

The parameters of an unfolding are called its *control variables*, and their number is its *codimension*. The canonical cusp potential is

$$U(x,a,b) = \tfrac{1}{4}x^4 - \tfrac{1}{2}bx^2 - ax, \tag{9}$$

where a and b are the control variables and $\operatorname{codim}(U) = 2$. The zeroes of $\partial U/\partial x$ are frequently depicted as a function of the control variables, as shown in Fig. 1, where $\partial U/\partial x = x^3-bx-a = 0$.

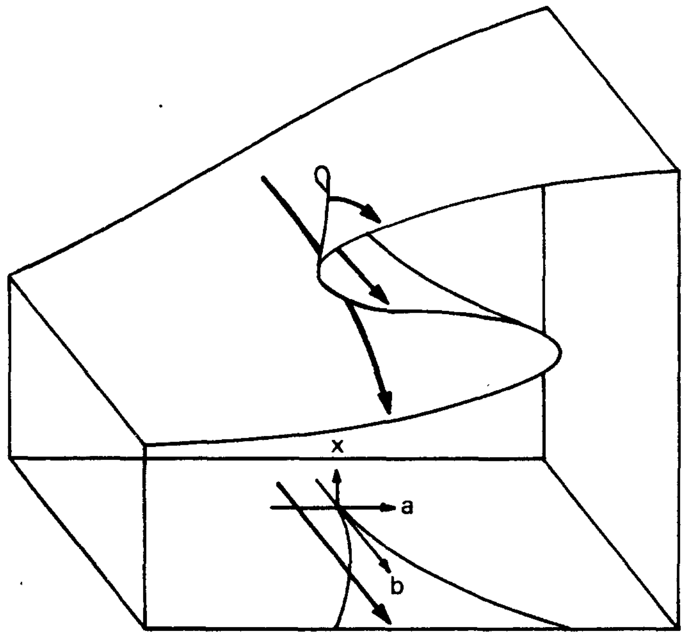

Fig. 1.

In the stochastic theory presented here there is a probability density function, f_c^* , associated with each point c in the control space of U. The precise form of f^* depends on the choice of the infinitesimal variance function, $v(x)$. Each choice of v determines a family of densities f_c^* , parametrized by c. Figure 2 shows a representative sequence from such a family, constructed from the cusp potential of Eq. (9) with $v(x)$=constant. The sequence of parameters in Fig. 2 corresponds to the pathway through the control space marked with an arrow in Fig. 1.

The nontrivial relationship between the number of minima of U and the number of modes of f^* is illustrated by the following theorem.

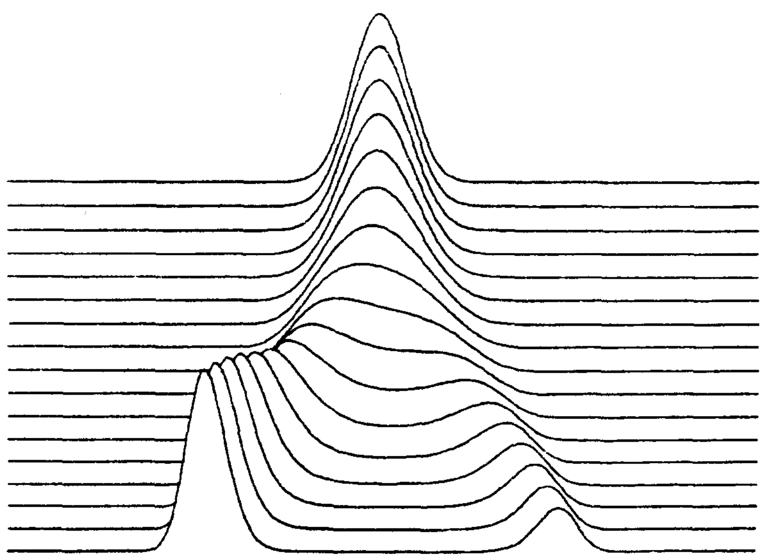

Fig. 2. Bifurcation of an exponential family of probability densities.

THEOREM: *Let $U:R^{k+1} \rightarrow R$ be the universal unfolding of a given catastrophe potential U of codimension k. Let $H \in R^k$ be the subset of the control space of U within which U has more than one relative minimum. Let $v(x)=2\epsilon(x-x_0)$, with x_0 arbitrary and $\epsilon>0$, and let f^* be the stationary probability density function associated to U and v. If $c \in h$, then there exists a value of ϵ such that f_c^* is not multimodal.*

Proof. By (7), the shape function for f_c^* is $g_c(x) = \partial U(x,c)/\partial x + \epsilon$. $U(x,c)$ can be taken to be a polynomial, from which g_c is a polynomial. Now choose ϵ to lift the relative minima of g_c above the $y=0$ axis. Then $g_c(x)$ has at most one real root.

This theorem is of great import to statistical catastrophe theory, because it states that stochastic systems with multiple stable equilibria may nevertheless exhibit unimodal stationary probability densities. It would appear possible that multiple stable equilibria could even be inferred from certain unimodal densities.

4. STATISTICAL INFERENCE

For the purposes of statistical catastrophe theory, it is fortunate that the stationary densities should have the exponential form (4). The *exponential families* of probability density functions take the form

$$f(x) = \exp[-\psi + \theta_1 \phi_1(x) + \cdots + \theta_k \phi_k(x)], \tag{10}$$

where the $\phi_i: R \rightarrow R$ are linearly independent, and ψ is chosen to normalize the density.

THEOREM [7]. *If $\{X_1, X_2, \ldots, X_N\}$ is a random sample of observations of a random variable whose density is an exponential family, then,*

(1) Maximum likelihood estimators (MLE's) for $\boldsymbol{\theta} = (\theta_1, \ldots \theta_k)$ exist.

(2) The MLE's are completely determined by the k sample statistics $\sum_{j=1}^{N} \phi_i(X_j)$, $i=1,\ldots,k$.

(3) The MLE's are asymptotically normally distributed as $N \rightarrow \infty$, and have minimum sampling variance in the class of all such estimators.

If both the potential function and the infinitesimal variance function of a stochastic catastrophe model are polynomials, then it is clearly possible to express its stationary density as an exponential family. For example, consider the canonical cusp potential of Eq. (9). If we assume that $v(x) = 2\epsilon x$, then

$$g(x) = x^3 - bx - a + \epsilon \tag{11}$$

is the shape polynomial. Clearly g has up to 3 roots, and so f^* is possibly bimodal. The density f^* is

$$f^*(x) = \exp[-\psi + ((a-\epsilon)\log(x) + bx - x^3/3)/\epsilon]. \tag{12}$$

So,

$$\begin{aligned} \theta_1 &= -1 + a/\epsilon, & \phi_1(x) &= \log(x), \\ \theta_2 &= b/\epsilon, & \phi_2(x) &= x, \\ \theta_3 &= -1/3\epsilon, & \phi_3(x) &= x^3. \end{aligned} \tag{13}$$

The theorem just stated now applies. The statistics

$$\left\{\sum_{i=1}^{N}\log X_i,\ \sum_{i=1}^{N}X_i,\ \sum_{i=1}^{N}X_i^3\right\}$$

are sufficient to calculate $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$, the maximum likelihood estimators for $(\theta_1, \theta_2, \theta_3)$. Estimates for the three original coefficients (a, b, ϵ) are easily obtained from $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$.

The only problematic part of this procedure is the calculation of the maximum likelihood estimators from the sufficient statistics. The following algorithm, for general densities of the form (10), will usually converge to the MLE's. It is based on a Newton–Raphson search for the maximum of the likelihood function. In the algorithm, $\boldsymbol{\theta}^{(n)}$ refers to the coefficient vector $(\theta_1, \ldots, \theta_k)$ on the n iteration. The initial value should be $\boldsymbol{\theta}^{(0)} = (0, 0, \ldots, 0, -1)$.

(1) Do (2–3) until $\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)}$.

(2) Using numerical integration, calculate

$$v_i = E\{\phi_i(x)|\boldsymbol{\theta}^{(n)}\} - \sum_{j=1}^{N}\phi_i(X_j)/N, \ (i=1, \ldots, k);$$

$$M_{ij} = \mathrm{Cov}\{\phi_i(X), \phi_j(X)|\boldsymbol{\theta}^{(n)}\}, \ (i,j = 1, \ldots, k).$$

(3) Let $\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} - \boldsymbol{M}^{-1}\boldsymbol{v}$.

This algorithm can easily be extended to cover cases in which control variables have also been measured. A penalty, of course, is paid in the numerical integration stage.

5. HYPOTHESIS TESTING

The procedure described in the previous section yields estimates for the elements of the parameter $\boldsymbol{\theta}$ of a specified hypothetical model, given a random sample of observations. The particular estimates so obtained maximize the *likelihood* of the observed data, where the maximum is taken over all the parameter values permitted by the hypothesis. The likelihood of a parameter $\boldsymbol{\theta}$ is defined as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} f(X_i|\boldsymbol{\theta}), \tag{14}$$

where $f(x|\boldsymbol{\theta})$ is the probability density function determined by $\boldsymbol{\theta}$. This likelihood function also permits the comparison of different models, using the straightforward notion that the model with the larger maximum likelihood is the better model. When one model is contained within another, in the sense that the parameter space of the first is a proper subset of the parameter space of the second, then this comparison can be formalized into a statistical *test*, with known asymptotic sampling properties. This is particularly convenient for statistical catastrophe theory, because the canonical catastrophe potentials in one variable form a heirarchy of polynomials of increasing degree. It is therefore a relatively simple matter to construct a corresponding heirarchy of statistical tests, with which each catastrophe model can be tested against models of higher codimension.

Let $L_{k,v}$ be the maximum of the likelihood function for a stochastic catastrophe model with codimension k and infinitesimal variance function v. Let $\lambda = L_{k,v}/L_{k+j,v}$ be the *likelihood ratio* for the test of the hypothesis H_k: codim$(U)=k$ against the alternative hypothesis H_{k+j}: codim$(U)=k+j$. Because H_k is contained within H_{k+j}, we will have $0 < \lambda \leq 1$. In essence, we want to reject H_k if λ is too small. We must determine what "too small" means.

Under certain mild regularity conditions it can be shown [7, p. 311] that $-2 \log\lambda$ converges in distribution to the χ^2 random variable with j degrees of freedom. This provides a criterion for the decision concerning the value of λ: H_k is rejected at the *significance level* α if $-2 \log\lambda$ exceeds the value c such that $\text{Prob}\{\chi_j^2 > c\} = \alpha$. This means that if H_k is true there is a (known) probability α that H_k will mistakenly be rejected.

6. SUMMARY

The broad outlines of a statistical catastrophe theory have been given. The four steps are (1) construction of a general class of stochastic catastrophe models, (2) determination of their stationary probability density functions, (3) specification of an algorithm for estimating their parameters, (4) identification of a statistical test by which competing models may be evaluated.

Since functions of maximum likelihood estimators are themselves maximum likelihood estimators, the estimation and testing paradigm described here is invariant under invertible changes of parametrization. However, MLE's are *not* invariant under general diffeomorphisms of the measured variables. Therefore, much of the topological generality of catastrophe theory may have been lost in the statistical portion of our theory.

There are a number of directions for future work. Catastrophes in two dependent variables (the umbilics) should be considered. Nonparametric statistical theory should be integrated with topological models—perhaps this combination is more natural than the pairing of maximum likelihood estimation (a parametric theory) with catastrophe theory. Lastly, it would be desirable to find a statistical procedure for parameter estimation and hypothesis testing that does not require numerical integration.

Acknowledgements—Thanks to Clint Miller, Chan Lam, Alan Cantor, Tze Rhe Chang, David Cozart, and Jean-Marie Pages for helpful discussions.

REFERENCES

1. R. Thom, *Structural Stability and Morphogenesis*, Benjamin, New York (1975).
2. E. C. Zeeman, *Catastrophe Theory: Selected Papers*, Addison-Wesley, Reading, Mass. (1977).
3. T. Poston and I. N. Stewart, *Catastrophe Theory and its Applications*, Pitman, San Francisco (1978).
4. T. T. Soong, *Random Differential Equations in Science and Engineering*, Academic, New York (1973).
5. L. Cobb, Stochastic Catastrophe Models and Multimodal Distributions. *Behav. Sci.* **23**, 360–374 (1978).
6. W. Feller, Diffusion Processes in One Dimension. *Trans. Am. Math. Soc.* **77**, 1–31 (1954).
7. E. L. Lehman, *Testing Statistical Hypotheses*, Wiley, New York (1966).