ESTIMATION THEORY FOR THE CUSP CATASTROPHE MODEL

Loren Cobb, Medical University of South Carolina

ABSTRACT

The 'cusp' model of catastrophe theory is very closely related to certain multiparameter exponential families of probability density functions. This relationship is exploited to create an estimation theory for the cusp model. An example is presented in which the independent variable has a 'bifurcation' effect on the dependent variable.

INTRODUCTION

The 'elementary' catastrophe models of Thom (1975) and Zeeman (1977) have attracted the attention of researchers and theorists throughout the sciences. A persistent problem with virtually all published applications, however, has been the absence of statistical procedures for detecting the presence of a catastrophe in any given body of data. This lack has resulted in some severe criticism of catastrophe models for being, among other things, speculative and unverifiable (Sussmann and Zahler, 1978). Thus catastrophe models have become associated in many minds with reckless speculation and intellectual irresponsibility. As part of an effort to overcome this problem, this paper presents an estimation theory and the beginnings of an inferential theory, in a form useful for survey research applications of catastrophe models.

Catastrophe models come in both dynamic and static forms, the static forms being simply the equilibria (stable and unstable) of the dynamic forms. The capacity for multiple stable equilibria is inherent in catastrophe models: this is the principal feature which distinguishes them from the standard models used in linear and polynomial regression. In effect, the 'control' factors of a catastrophe model correspond to the independent variables of a statistical model, and the 'behavioral' variable of a catastrophe model corresponds to the dependent variable of a statistical model. When the control factors are such that the behavioral variable is in a multistable situation, then <u>each</u> stable equilibrium value is a predicted value of the behavioral variable - thus there is more than one predicted value. In addition, the unstable equilibria which separate the stable equilibria are also predictions of a sort: they are the values that we predict the behavioral variable will <u>not</u> have. This feature of catastrophe models makes it difficult to define the size of an error of prediction.

There are two ways of overcoming this difficulty. Both of these ways have emerged from a study of various forms of dynamic stochastic catastrophe models (Cobb, 1978, 1981, and Cobb and Watson, 1981). One of these is based on the method of moments and is an estimation method only, while the other is based on maximum likelihood estimation and permits hypothesis testing with the use of the chi-square approximation to the likelihood ratio test. The former has the advantage of computational simplicity, while the latter is clearly preferable when hypotheses must be tested.

THE CUSP MODEL

The canonical cusp model can be thought of as a rather peculiar response surface model. It's shape may be seen in Figure 1 on the next page. Note that sections taken through the depicted surface parallel to the α-axis are just cubic polynomials in y, the dependent variable. The entire surface is defined by the implicit equation

$$0 = \alpha + \beta(y-\lambda)/\sigma - \{(y-\lambda)/\sigma\}^3.$$

If we let $z = (y-\lambda)/\sigma$ be the 'standardized' dependent variable, then the cubic equation is simply

$$0 = \alpha + \beta z - z^3.$$

It may be seen that λ and σ are the <u>location</u> and <u>scale</u> parameters, respectively. The roots of the cubic polynomial are the predicted values of z, given α and β. When there are three roots, the central root is an 'anti-prediction': a prediction of where the dependent variable will <u>not</u> be. This feature of the cusp model is clarified in Figure 2, which shows the sequence of conditional probability density functions for y, with α fixed as β is increased. This sequence corresponds to the trajectory and its projection that are shown in Figure 1. These probability density functions will be discussed in a later section.

The two dimensions of the 'control' space, α and β, are canonical factors which depend upon the actual measured independent variables, say $X_1, \ldots, X_v$. As a first approximation, we may suppose that the control factors depend linearly upon the independent variables:

$$\alpha = \alpha_0 + \alpha_1 X_1 + \ldots + \alpha_v X_v,$$

$$\beta = \beta_0 + \beta_1 X_1 + \ldots + \beta_v X_v.$$

Thus the statistical estimation problem is to find estimates for the 2v+4 parameters

$$\{\lambda,\sigma,\alpha_0,\ldots,\alpha_v,\beta_0,\ldots,\beta_v\},$$

from n observations of the v+1 variables

$$\{Y, X_1, \ldots, X_v\}.$$

As β changes from negative to positive, the conditional probability density function of y changes in shape from unimodal to bimodal. For this reason the β factor will be called the <u>bifurcation</u> factor (it has also been called the

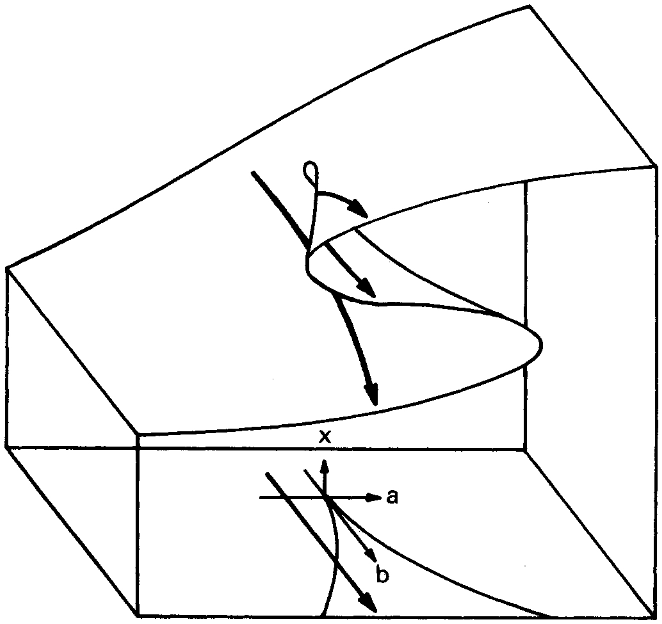

Figure 1: The cusp catastrophe model.

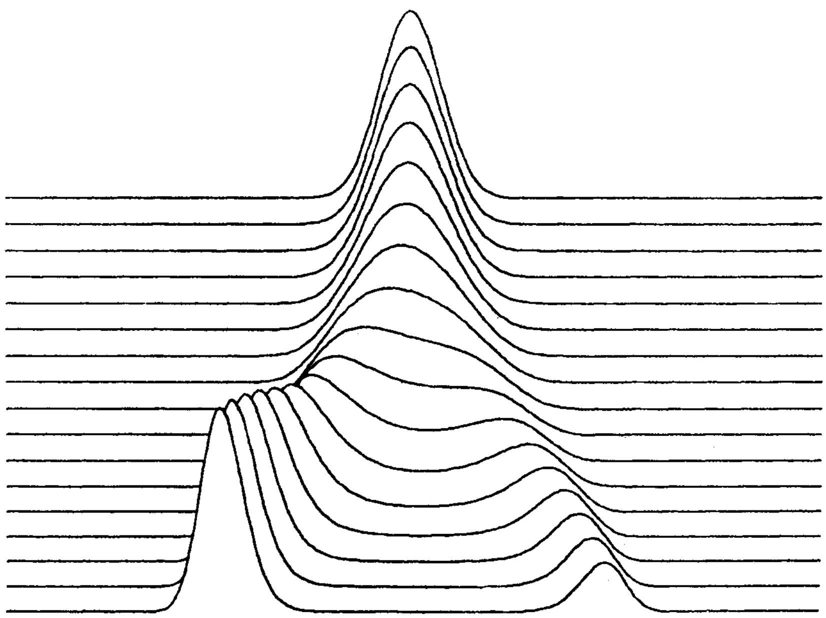

Figure 2: The cusp probability density function.

'splitting' factor, by Zeeman and others). When α is zero the pdf is symmetrical no matter what the value of β. When the pdf is unimodal, α determines its skew: α positive implies positive skew and vice versa. However, when the pdf is bimodal, then α determines the relative height of the two modes: α positive implies that the right-hand mode is higher, and vice versa. To encompass these attributes, α will be called the asymmetry factor (it has also been called the 'normal' factor, a rather misleading term).

Because the model is based on a cubic, it is possible to define a statistic which discriminates between the unimodal and bimodal cases. This is Cardan's discriminant:

$$\delta = (\alpha/2)^2 - (\beta/3)^3.$$

When δ is negative the pdf is unimodal, and when it is positive the pdf is bimodal.

STATISTICAL THEORY

The probability density function upon which all of the preceding descriptive statistics were based is the standard 4-parameter cusp pdf:

$$f(y) = \xi\exp(\alpha z + \beta z^2/2 - z^4/4),$$

in which $z = (y-\lambda)/\sigma$.

The constant ξ merely normalizes the pdf so that it has unit integral over its range, which is the whole real line. The modes and antimodes of the cusp pdf may be found by solving $df/dy = 0$. This yields the equation

$$\alpha + \beta z - z^3 = 0,$$

which is exactly the same as the implicit equation which defined the cusp surface. The modes of the cusp pdf are the predicted values of the cusp model, and the antimodes of the cusp pdf are the 'anti-predictions' of the cusp model. The derivation of the cusp pdf from stochastic catastrophe theory, using stochastic differential equations, may be found in (Cobb and Watson, 1981). The statistical theory was first presented in rudimentary form in (Cobb, 1978).

The standard cusp pdf can clearly be reparametrized so that it is a canonical exponential family, as in:

$$f(y) = \exp(-\eta+\tau_1 y+\tau_2 y^2+\tau_3 y^3+\tau_4 y^4).$$

Now the well-developed theory (e.g. Lehmann, 1959) of exponential families can be applied: we know that maximum likelihood estimators (MLE's) exist, are unique, and can be found, for example by a Newton-Raphson search. This search procedure proceeds as follows. Let τ stand for the vector of parameters

$$\tau = (\tau_1, \ldots, \tau_4),$$

let S be the vector of sample means defined by

$$S_k = (1/n)\sum_{i=1}^{n} Y_i^k, \quad \text{for } k=1,2,3,4,$$

let $M(\tau)$ be the vector of expectations of $f(y)$ given τ defined by

$$M_k(\tau) = E[Y^k], \quad \text{for } k=1,2,3,4,$$

and let $H(\tau)$ be the 4x4 matrix of covariances given τ defined by

$$H_{ij}(\tau) = \mathrm{Cov}[Y^i,Y^j],$$

for $i,j=1,2,3,4$. The Newton-Raphson search is an iterative procedure which starts from an initial guess for τ, say t. Each guess is improved by adding to it the quantity

$$[S-M(t)][H(t)]^{-1}.$$

This calculation is performed repetitively until $S = M(t)$, within the limits of computational accuracy.

It should be noted that after each iteration the vector $M(t)$ and the matrix $H(t)$ must be recalculated. These moments must be found by numerical integration, since closed form expressions for the moments of the cusp pdf are not known.

The preceding discussion applies to the estimation of the four parameters of the cusp pdf, given observations of the variable Y. If, however, the cusp pdf is to be used as the conditional density of Y, given the values of the independent variables, then the maximum likelihood procedure becomes more complicated. If we use the previously stated assumption that the factors α and β are linear combinations of the independent variables, then the extension of the Newton-Raphson technique is straightforward. There are now $2v+4$ parameters to be estimated, and the matrix $H(t)$ becomes $(2v+4)\mathrm{x}(2v+4)$ dimensional. The problem is that calculating the elements of $H(t)$ and $M(t)$ now becomes almost prohibitively tedious, especially for large data sets, since the computation time is now proportional to the sample size.

We now proceed to estimation by the method of moments. It will be seen that this method, in sharp contrast to the method of maximum likelihood, is extremely easy to implement. As noted before, it does not, unfortunately, permit hypothesis testing.

Even though closed-form expressions for the moments of the cusp pdf do not exist, moment estimators are trivial to derive with the aid of the following general theorem:

Theorem: Let $g(x,y)$ be a polynomial function of x and y such that

$$0 < \int_{-\infty}^{+\infty}\exp\{-\int g(x,y)dy\} < \infty.$$

Let ξ be the reciprocal of this quantity. Define the conditional density of a random variable Y given X as

$$f(y|x) = \xi\exp\{-\int g(x,y)dy\}.$$

Assume that the joint density of X and Y has moments of all orders, and let $h(x)$ denote the density of X. Then for any non-negative j and k

$$E[X^j Y^k g(X,Y)] = kE[X^j Y^{k-1}].$$

Proof: Note that $f(y|x)$ is asymptotically zero as y tends to either + or $-\infty$. Since $g(x,y)$ is a polynomial, we also have that $y^k f(y|x)$ tends to zero in the same way. Further, we can write $g(x,y)$ as

$$g(x,y) = -\{\partial f(y|x)/\partial y\}/f(y|x).$$

Substituting this expression into the moment formula, we obtain

$$E[X^j Y^k g(X,Y)] = \int\int x^j y^k g(x,y) f(y|x) h(x) dydx$$

$$= \int\int x^j y^k \{-\partial f(y|x)\} h(x) dx$$

$$= \int x^j h(x) \{-\int y^k \partial f(y|x)\} dx.$$

Now use integration by parts on the inner integral, and observe that one of the parts is identically zero:

$$-\int y^k \partial f(y|x) = -y^k f(y|x) \Big|_{-\infty}^{+\infty} + k\int y^{k-1} f(y|x) dy$$

$$= 0 + k\int y^{k-1} f(y|x) dy.$$

Thus we now have

$$E[X^j Y^k g(X,Y)] = \int x^j h(x) k \int y^{k-1} f(y|x) dydx$$

$$= k\int\int x^j y^{k-1} f(y|x) h(x) dydx$$

$$= kE[X^j Y^{k-1}].$$

This theorem enables the method of moments to be applied to models that, like the elementary catastrophes, are expressed as implicit equations. Before examining the cusp model, it may be worthwhile to show how it can be applied to ordinary linear regression. The linear regression model

$$y = a + bx + \varepsilon$$

can be written in implicit equation form as

$$g(x,y) = (y - a - bx)/\sigma^2 = 0,$$

where σ^2 will turn out to be the variance of ε. The conditional pdf of y given x is

$$f(y|x) = \xi \exp[-(y^2/2 - ay - bxy)/\sigma^2]$$

$$= \xi \exp[-\{y-(a+bx)\}^2/2\sigma^2].$$

This is clearly a normal density, $N[a+bx,\sigma^2]$. (To obtain this formula, complete the square and absorb the terms in x into ξ, the normalizing constant).

To find estimation equations for a and b, use the theorem twice, first with j=k=0 and second with j=1 and k=0:

1) $E[g(X,Y)] = 0$

$$\Longrightarrow a + bE[X] = E[Y],$$

2) $E[Xg(X,Y)] = 0$

$$\Longrightarrow aE[X] + bE[X^2] = E[XY].$$

Notice that when sample moments are substituted for these expectations, we obtain the usual Gauss-Markov normal equations for linear regression. To estimate σ^2, use the theorem again, this time with j=0 and k=1:

3) $E[Yg(x,Y)] = E[Y^0] = 1$

$$\Longrightarrow E[Y^2] - aE[Y] - bE[XY] = \sigma^2,$$

which is the correct formula for the residual variance of Y after the linear effect of X has been removed by linear regression.

Turning now to the cusp model, let us consider first the model with no independent variables:

$$g(y) = a + by + cy^2 + dy^3, \quad (d>0).$$

This model has a pdf given by

$$f(y) = \xi \exp\{ay + by^2/2 + cy^3/3 + dy^4/4\},$$

which has modes and antimodes at the roots of $g(y) = 0$. The transformation from (a,b,c,d) to the standard coefficients is accomplished by

$$\lambda = -c/3d, \quad \sigma = d^{-1/4},$$
$$\alpha = -\sigma(a+b\lambda+c\lambda^2+d\lambda^3), \quad \text{and}$$
$$\beta = -\sigma^2(b+c\lambda).$$

Estimation of (a,b,c,d) from the moments of Y proceeds from an application of the theorem. Let

$$\mu_k = E[Y^k].$$

From a single application of the theorem we can derive a linear difference equation (with one varying coefficient) for the moments of Y:

$$E[Y^k g(Y)] = kE[Y^{k-1}]$$

$$\Longrightarrow k\mu_{k-1} = a\mu_k + b\mu_{k+1} + c\mu_{k+2} + d\mu_{k+3}.$$

Simply apply this result with k=0,1,2,3 to obtain a system of four linear equations in the four unknowns (a,b,c,d). Substitute sample moments for the expectations and solve the system. Transform the resulting estimates as indicated to obtain $(\lambda,\sigma,\alpha,\beta)$.

It is trivial to expand this technique for models with independent variables. For example, suppose there is one independent variable, say X. Then the model is

$$g(x,y) = b_1 + b_2 x + b_3 y + b_4 xy + b_5 y^2 + b_6 y^3.$$

The standard coefficients are obtained from

$$\lambda = -b_5/3b_6, \quad \sigma = b_6^{-1/4},$$

$$\alpha_0 = -\sigma(b_1 + b_3\lambda + b_5\lambda^2 + b_6\lambda^3),$$

$$\alpha_1 = -\sigma(b_2+b_4\lambda),$$

$$\beta_0 = -\sigma^2(b_3+b_5\lambda),$$

$$\beta_1 = -\sigma^2 b_4.$$

Estimation of the coefficients from moments proceeds as before. Apply the theorem six times, with j=0 and k=0,1,2,3, and then with j=1 and k=0,1. Solve the resulting system and transform to get the standard coefficients.

Estimation by the method of moments does not yield estimators with known sampling distributions, and cannot therefore be used for hypothesis testing. The maximum likelihood method yields estimators that are efficient and that have known (asymptotic) sampling distributions, but that are, from a computational point of view, highly inefficient. It is possible, however, to use the moment estimates as the initial guess for the Newton-Raphson iterations, thus cutting down somewhat the time required to calculate the MLE's. Once MLE's have been obtained, it is possible to test hypotheses (for example, comparing a linear regression model to the equivalent catastrophe model) using the chi-square approximation for the likelihood ratio test.

AN EXAMPLE

An excellent example of published empirical data which seems to exhibit a bifurcation in the dependent variable has been quoted by Zeeman (1977, pp. 373-385). The data come from a study of driving performance before and after the ingestion of alcohol (Drew, Colquhoun, and Long, 1959). Essentially, the authors found that the change in time per lap (i.e. driving speed) was strongly affected by the position of the subjects on the Bernreuter scale of introversion. However, the effect was not linear: the correlation between change in lap time and introversion was .005 or effectively zero. However, as is visible from Figure 3, it is clear that whereas extroverts continuted to drive at about the same speed after drinking, the introverts either drove faster or slower, and did not stay at the same speed.

These data were reproduced as a figure in (Zeeman, 1977, Fig. 1), from which approximate data were recovered by digitization. Following Zeeman, three cases were eliminated as extreme outliers, leaving the 37 cases seen in Figure 3. The six parameter cusp model with one independent variable was fitted to the data using to method of moments as given above, and the resulting relationship between change in driving speed after alcohol (the dependent variable) and introversion (the independent variable) is shown. The dashed line indicates values that are predicted not to occur.

Zeeman also used a cusp model in his article, although it differs substantially from the one in Figure 3. Poston and Stewart (1978, pp. 420-423) criticize Zeeman's model on psychological grounds, and suggest the bifurcation model that appears here.

REFERENCES

Cobb, L., (1978). Stochastic catastrophe models and multimodal distributions. Behavioral Science, 23, 360-374.

Cobb, L., (1980). Stochastic differential equations for the social sciences. Chapter 2 in L. Cobb and R.M. Thrall (eds.), Mathematical Frontiers of the Social Sciences. Boulder, CO: Westview Press.

Cobb, L. and Watson, W.B., (1981). Statistical catastrophe theory: An overview. Mathematical Modelling, in press.

Drew, G.C., Colquhoun, W.P., and Long, H.A., (1959). Effect of small doses of alcohol on a skill resembling driving. Memo 38. London: Medical Research Council.

Lehmann, E.L., (1959). Testing Statistical Hypotheses. New York: Wiley.

Poston, T., and Stewart, I., (1978). Catastrophe Theory and Its Applications. London: Pitman.

Sussmann, H.J., and Zahler, R.S., (1978). A critique of applied catastrophe theory in the behavioral sciences. Behavioral Science, 23, 383-389.

Thom, R., (1975). Structural Stability and Morphogenesis. Reading, MA: Benjamin Press.

Zeeman, E.C., (1977). Catastrophe Theory: Selected Papers. Reading: MA: Addison-Wesley.

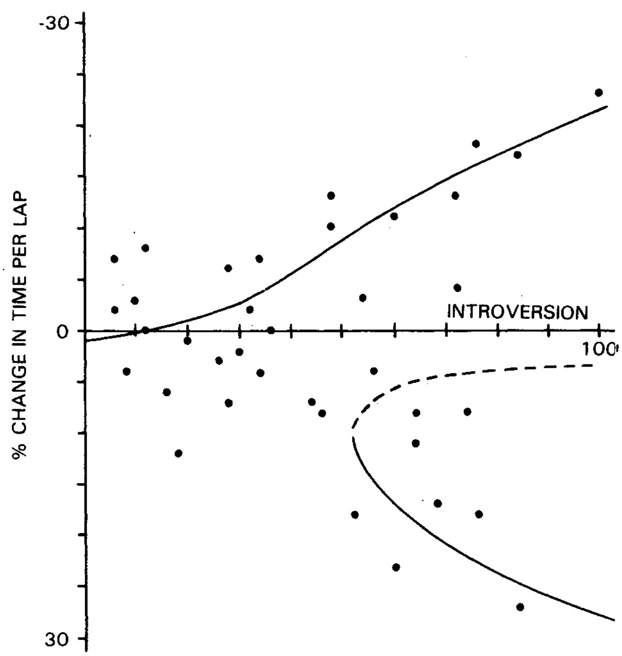

Figure 3: A bifurcation model fitted to data.