# Estimation Theory for the Cusp Catastrophe Model

***REVISED EDITION*[1]**

Loren Cobb

## Abstract

The cusp model of catastrophe theory is very closely related to certain multi-parameter 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 an 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 catastrophes in any given body of data. This lack has resulted in some severe criticism of catastrophe models 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 multi-stable situation, then *each* 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 that the behavioral variable will *not* have. This feature of catastrophe models makes it difficult to define the size of an error of prediction.

*Written June 1980*
*Published Spring 1981*
*Revised June 2010*

*1980 Proceedings of the American Statistical Association*
*Section on Survey Research Methods*
*pp. 772–776*

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, 1980; Cobb & Watson, 1981). One of these is based on the method of moments, while the other is based on maximum likelihood estimation. The latter permits hypothesis testing through 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.

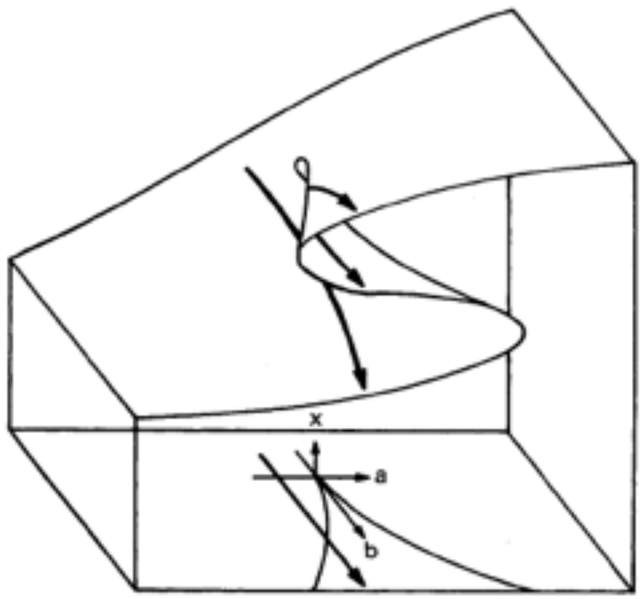


**Figure 1: The cusp catastrophe model.**

## 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 below. Note that sections taken through the depicted surface parallel to the $\alpha$-axis are just cubic polynomial in $y$, the dependent variable. The entire surface is defined by the implicit equation

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

If we let $z = \dfrac{y-\lambda}{\sigma}$ be the "standardized" dependent variable, then the cubic equation is simply

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

It may be seen that $\lambda$ and $\sigma$ are *location* and *scale* parameters, respectively, for $y$. The roots of the cubic polynomial are the predicted values of $z$, given $\alpha$ and $\beta$. When there are three roots, the central root is an "*anti-prediction*": a prediction of where the dependent variable will *not* be. This feature of the cusp model is clarified in Figure 2, which shows the sequence of conditional probability density functions for $y$, with $\alpha$ fixed as $\beta$ 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.

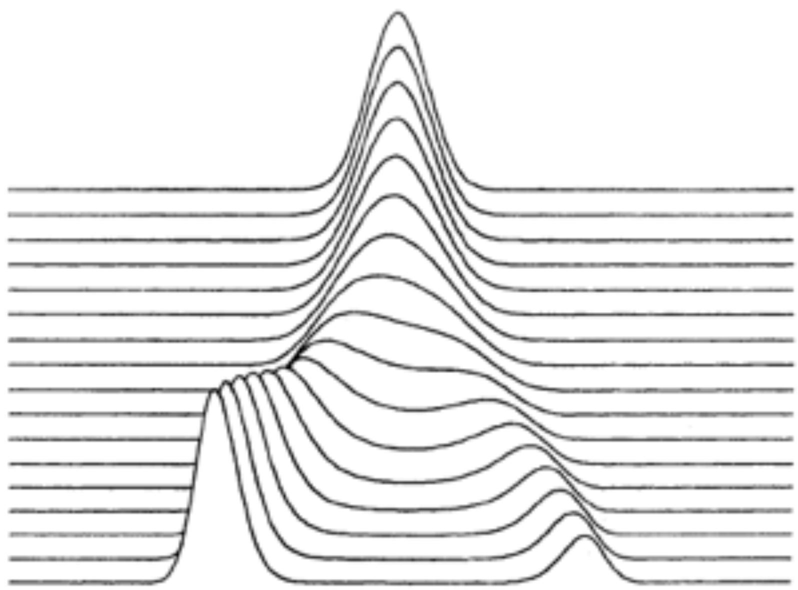

**Figure 2: The cusp family of probability density functions.**

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

$$\alpha = \alpha_0 + \alpha_1 X_1 + ... + \alpha_v X_v,$$
$$\beta = \beta_0 + \beta_1 X_1 + ... + \beta_v X_v.$$

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

$$(\lambda, \sigma, \alpha_0, ..., \alpha_v, \beta_0, ..., \beta_v),$$

from $n$ observations of the $v$+1 variables

$$(Y, X_1, ..., X_v).$$

As $\beta$ changes from negative to positive, the conditional probability density function of $y$ changes in shape from unimodal to bimodal. For this reason the $\beta$ factor will be called the *bifurcation* factor (it has also been called the *splitting* factor, by Zeeman and others). When $\alpha$ is zero the PDF is symmetrical no matter what the value of $\beta$. When the PDF is unimodal, $\alpha$ determines its skew: $\alpha$ positive implies positive skew, and vice versa. However, when the PDF is bimodal, then $\alpha$ determines the relative height of the two modes: $\alpha$ positive implies that the right-hand mode is higher, and vice versa. To encompass these attributes with a single term, $\alpha$ will be called the *asymmetry* factor (it has also been called the *normal* factor, a rather misleading term in the statistical context).

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

$$\delta = \left(\tfrac{1}{2}\alpha\right)^2 - \left(\tfrac{1}{3}\beta\right)^3 .$$

When $\delta$ 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:

$$f(y;\alpha,\beta,\lambda,\sigma) = \xi \exp\left(\alpha z + \tfrac{1}{2}\beta z^2 - \tfrac{1}{4}z^4\right), \quad \text{with} \quad z = \frac{y-\lambda}{\sigma}.$$

The constant $\xi$ 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 $\frac{\partial f}{\partial y} = 0$ . This yields the equation

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

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, while 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 & Watson, 1981). The statistical theory was first presented in rudimentary form by (Cobb, 1978).

The standard cusp PDF can clearly be reparametrized so that it is an exponential family, as in:

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

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

$$\tau = \left(\tau_1, \tau_2, \tau_3, \tau_4\right),$$

let $S$ be the vector of sample means defined by

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

let $M_k(\tau)$ be the vector of expectations of $Y$ given $\tau$ defined by

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

and let $H(\tau)$ be the 4x4 covariance matrix defined by

$$H_{ij}(\tau) = \mathrm{Cov}\left[Y^i, Y^j\right], \quad \text{for } i, j = 1,2,3,4.$$

The Newton-Raphson search is an iterative procedure which starts from an initial guess for $\tau$, say $t_0$. Each iteration is given by the vector expression

$$t_{k+1} = t_k + \left(S - M(t_k)\right) H(t_k)^{-1}.$$

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.[2]

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 $\alpha$ and $\beta$ are linear combinations of the independent variables, then the extension of the Newton-Raphson technique for finding the maximum likelihood estimates is straightforward. There are now $2v+4$ parameters to be estimated, and the matrix $H(t)$ becomes $(2v+4)\times(2v+4)$ dimensional.

We now proceed to estimation by the method of moments. It will be seen that this method, in contrast to the method of maximum likelihood, is extremely easy to implement.

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

$$0 < \int_{-\infty}^{+\infty} \exp\left[-\int g(x,y)\,dy\right] < \infty, \quad \forall x.$$

Let $\psi(x)$ be the reciprocal of this quantity. Suppose that a random variable $Y$ depends upon x in such a way that its conditional density is given by

$$f(y|x) = \psi(x)\exp\left[-\int g(x,y)\,dy\right].$$

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

$$\mathrm{E}\left[X^j Y^k g(X,Y)\right] = k\,\mathrm{E}\left[X^j Y^{k-1}\right].$$

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

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

Substituting this expression into the moment formula, we obtain

$$\begin{aligned}
\mathrm{E}\left[X^j Y^k g(X,Y)\right] &= \iint x^j y^k g(x,y) f(y|x) h(x)\,dy\,dx \\
&= \iint x^j y^k \left(-\partial f(y|x)\right) h(x)\,dx \\
&= \int x^j h(x)\left(-\int y^k \partial f(y|x)\right) dx.
\end{aligned}$$

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

$$\begin{aligned}
-\int y^k \partial f(y|x) &= -y^k f(y|x)\Big|_{-\infty}^{+\infty} + k\int y^{k-1} f(y|x)\,dy \\
&= 0 + k\int y^{k-1} f(y|x)\,dy.
\end{aligned}$$

Thus we now have

$$\begin{aligned} \mathrm{E}\left[X^j Y^k g(X,Y)\right] &= \int x^j h(x)\left(k\int y^{k-1} f(y \mid x)\,dy\right)dx \\ &= k\iint x^j y^{k-1} h(x)\,dy\,dx \\ &= k\,\mathrm{E}\left[X^j Y^{k-1}\right], \end{aligned}$$

which was to be shown. ☒

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

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

where $\sigma^2$ will turn out to be the variance of $\varepsilon$. The conditional PDF of $y$ given $x$ is

$$\begin{aligned} f(y|x) &= \psi(x)\exp\left[-\frac{\frac{1}{2}y^2 - ay - bxy}{\sigma^2}\right] \\ &= \psi_1(x)\exp\left[-\tfrac{1}{2}\left(\frac{y-(a+bx)}{\sigma}\right)^2\right]. \end{aligned}$$

This is clearly a parametrized family of normal densities, $N\left(a+bx,\sigma^2\right)$. (To obtain this formula, simply complete the square and absorb the terms in $x$ into the function $\psi$, 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. $\mathrm{E}\left[g(X,Y)\right] = 0 \quad \Rightarrow \quad a + b\,\mathrm{E}[X] = \mathrm{E}[Y]$,
2. $\mathrm{E}\left[Xg(X,Y)\right] = 0 \quad \Rightarrow \quad a\,\mathrm{E}[X] + b\,\mathrm{E}\left[X^2\right] = \mathrm{E}[XY]$,

Notice that when sample moments are substituted for these expectations, we obtain the usual Gauss-Markov normal equations for linear regression.

To estimate $\sigma^2$, use the theorem again, this time with $j = 0$ and $k = 1$:

3. $\mathrm{E}\left[Yg(X,Y)\right]=1 \quad \Rightarrow \quad \mathrm{E}\left[Y^2\right]-a\mathrm{E}[Y]+b\mathrm{E}[XY]=\sigma^2,$

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

$$f(y)=\xi\exp\left[ay+\tfrac{1}{2}by^2+\tfrac{1}{3}cy^3+\tfrac{1}{4}dy^4\right],$$

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

$$\begin{aligned}
&\sigma=\sqrt[4]{d},\\
&\lambda=-\frac{c}{3d},\\
&\alpha=-\sigma\left(a+b\lambda+c\lambda^2+d\lambda^3\right), \quad \text{and}\\
&\beta=-\sigma^2(b+c\lambda).
\end{aligned}$$

Estimation of the parameter vector $(a,b,c,d)$ proceeds from an application of the theorem. Let

$$\mu_k=\mathrm{E}\left[Y^k\right].$$

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

$$\begin{aligned}
&\mathrm{E}\left[Y^k g(Y)\right]=k\mathrm{E}\left[Y^{k-1}\right]\\
&\Rightarrow k\mu_{k-1}=a\mu_k+b\mu_{k+1}+c\mu_{k+2}+d\mu_{k+3}.
\end{aligned}$$

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 above to obtain $(\alpha,\beta,\lambda,\sigma)$.

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

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

Estimation of the coefficients $(b_1, b_2, b_3, b_4, b_5, b_6)$ proceeds as before. Apply the theorem six times, with $j = 0$ and $k = 0, 1, 2, 3$, and then with $j = 1$ and $k = 0, 1$. Substitute sample moments for expectations, and solve the resulting system. Finally, the standard coefficients are obtained from

$$\sigma = \sqrt[4]{b_6},$$
$$\lambda = -\frac{b_5}{3b_6},$$
$$\alpha_0 = -\sigma\left(b_1 + b_3\lambda + b_5\lambda^2 + b_6\lambda^3\right),$$
$$\alpha_1 = -\sigma\left(b_2 + b_4\lambda\right),$$
$$\beta_0 = -\sigma^2\left(b_3 + b_5\lambda\right),$$
$$\beta_1 = -\sigma^2 b_4.$$

Estimation by the method of moments does not yield estimators with known sampling distributions, and cannot easily be used for hypothesis testing *[Note: The approximate sampling distribution of these estimators was later derived by Ferdon in her PhD dissertation (Ferdon, 1983).]* The maximum likelihood method yields estimators that are efficient and that have known asymptotic sampling distributions, but they are, from a computational point of view, 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 MLEs. Once MLEs have been obtained, it is possible to test hypotheses — for example, to compare a linear regression model to the equivalent cusp catastrophe model — using the chi-square approximation to 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. 373–85). 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. the driving speed) was strongly affected by the position of the subjects on the Bernreuter scale of introversion. However, as is visible from Figure 3, it is clear that whereas extroverts continued to drive at about the same speed after drinking, the introverts either drove faster or slower, and few stayed 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 depicted in Figure 3. The six-parameter cusp model with one independent variable was fitted to the data using the 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 (the anti-predictions).

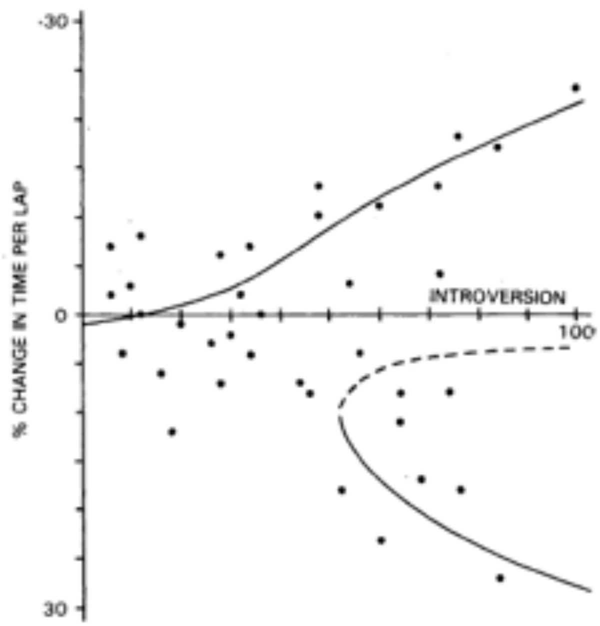


**Figure 3: A bifurcation model fitted to data.**

Zeeman also used a cusp model in his article, although it differs substantially from the one in Figure 3. Poston and Stewart (1978, pp. 420–23) criticized Zeeman's model on psychological grounds, and suggested the bifurcation model that appears here.

## References


Cobb, Loren (1978) “Stochastic catastrophe models and multimodal distributions,” *Behavioral Science, v. 23*, 360–374.

Cobb, Loren (1980) “Stochastic differential equations for the social sciences,” Chapter 2 in *Mathematical Frontiers of the Social and Policy Sciences*, edited by Loren Cobb and R. M. Thrall. Boulder, CO: Westview Press.

Cobb, Loren and Watson, William B. (1981) “Statistical catastrophe theory: An overview,” *Mathematical Modeling, v. 1*, 311–317.

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.

Ferdon, Marybeth E. (1983) *Inference for Quadratic and Catastrophe Response Surface Models*. Unpublished doctoral dissertation. Charleston, SC: Department of Biostatistics and Epidemiology, Medical University of South Carolina.

Grasman, Raoul P., van der Maas, Han L., and Wagenmakers, Eric-Jan (2009) "Fitting the cusp catastrophe in R: A cusp package primer," *Journal of Statistical Software, vol. 32*, #8, 1–27.

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

Poston, Timothy and Stewart, Ian (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, vol. 23*, 383–389.

Thom, René (1975) *Structural Stability and Morphogenesis*. Reading, MA: Benjamin.

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

## Endnotes

[1] This non-refereed proceedings paper is one of my most-cited articles. For this revision I have fixed several grammatical and mathematical errors, and brought some references up to date.
[2] There are efficient alternatives to numerical integration, as employed for example in the R package for cusp estimation by (Grasman et al, 2009).