x = "the data"

f = natural process that produced the data (REALITY)

f(x) = pdf of the natural process f
(ideal abstraction; unknowable "straw man")

g(x|θ) = MODEL used to approximate reality
(θ estimated from data using MLE techniques)

Kullback-Leibler Information Theorem:

All models are WRONG.

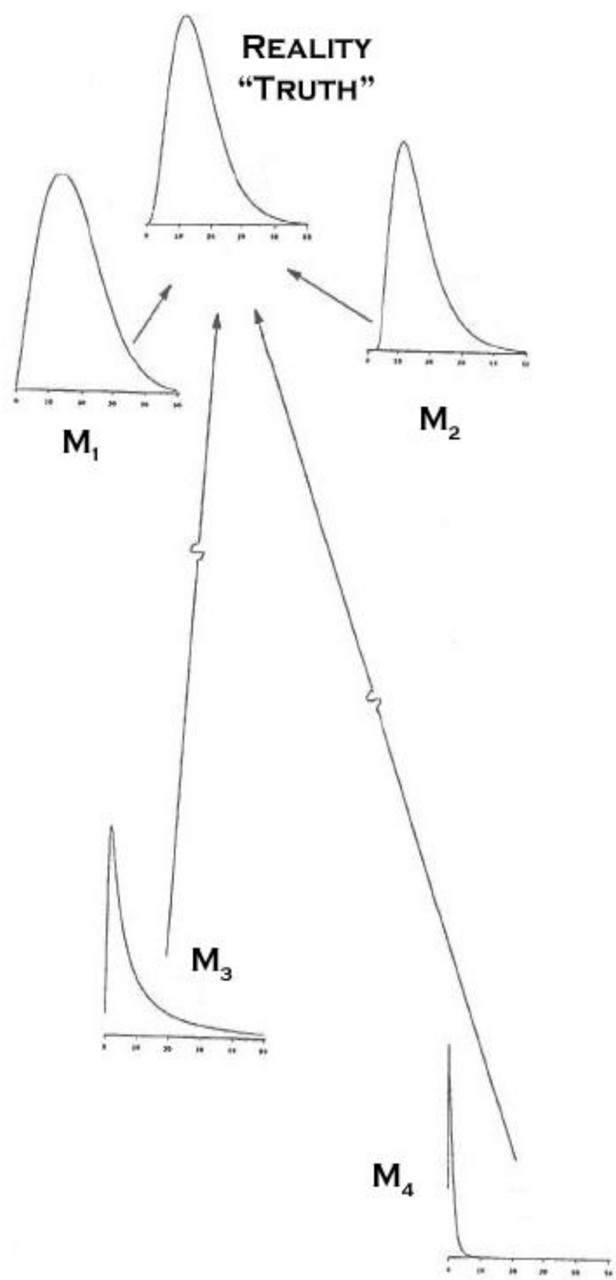
Best model demonstrates the <u>minimum loss of information</u> relative to reality

K-L Information equation can be used to quantify this loss of information:

$$I(f, g) = \int f(x) \log \left(\frac{f(x)}{g(x|\theta)} \right) dx$$

K-L Information Loss (or "distance"):

$$I(f, g) = \int f(x) \log \left(\frac{f(x)}{g(x|\theta)} \right) dx$$



<u>Akaike's Extentions to the K-L theorem</u>: (ah-kah-ee-kay)

The general equation

$$I(f, g) = \int f(x) \log \left(\frac{f(x)}{g(x|\theta)} \right) dx$$

Could be rewritten as:

$$I(f, g) = \int f(x) \log (f(x))\, dx \;-\; \int f(x) \log (g(x \mid \theta)) dx$$

Akaike noted that each of the terms on the right contained a statistical expectation with respect to "truth" (i.e., f(x)). Hence, the K-L information equation could be expressed as a difference of expected values:

$$I(f, g) = \mathrm{E}_f[\log(f(x))] - E_f[\log(g(x \mid \theta))]$$

The first expected value above depends solely on REALITY and reduces to a CONSTANT <u>for a given sample of data</u>. In practice, this value is unknowable. Nonetheless, this realization yields the following insight:

$$I(f, g) = \text{Constant} - \mathrm{E}_f[\log(g(x \mid \theta))]$$

or

$$I(f, g) - \text{Constant} = -\, \mathrm{E}_f[\log(g(x \mid \theta))]$$

That is: ln maxL(g) is PROPORTIONAL to K-L distance !!!

Hence,

For a given sample of data, the RELATIVE INFORMATION LOSS of two or more models could be derived based solely on the log maximum likelihood of the models given that data.

Akaike (ah-kah-ee-kay) “did the math” to show that such estimates were biased and provided his now famous correction for that bias:

$$\mathbf{AIC = -2 \ln (L(\theta|x)) + 2K}$$

Where K = number of estimated parameters

Lewandowsky & Farrell (2011) explore the application of the Akaike Information Criterion for model selections in Chapter 5 (pages 179-194).