<table>
<thead>
<tr>
<th>Title</th>
<th>The minimum local cross-entropy criterion for inferring risk-neutral price distributions from traded options prices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors(s)</td>
<td>Edelman, David</td>
</tr>
<tr>
<td>Publication date</td>
<td>2004-04-18</td>
</tr>
<tr>
<td>Series</td>
<td>Centre for Financial Markets working paper series; WP-04-01</td>
</tr>
<tr>
<td>Publisher</td>
<td>University College Dublin. School of Business. Centre for Financial Markets</td>
</tr>
<tr>
<td>Link to online version</td>
<td><a href="http://www.ucd.ie/bankingfinance/docs/wp/EDELMAN1.pdf">http://www.ucd.ie/bankingfinance/docs/wp/EDELMAN1.pdf</a></td>
</tr>
<tr>
<td>Item record/more information</td>
<td><a href="http://hdl.handle.net/10197/1127">http://hdl.handle.net/10197/1127</a></td>
</tr>
</tbody>
</table>
The Minimum Local Cross-Entropy Criterion
For Inferring Risk-Neutral Price Distributions
From Traded Options Prices

David Edelman
University College, Dublin
April 18, 2004

Abstract
A quantity known as the Local Cross-Entropy (LCE) for a density is proposed, defined to be the local derivative of the Cross-Entropy between a density and a 'kernel-smoothed' version of itself, with respect to bandwidth of the smoothing. This criterion is argued to be of the 'smoothness' type and is also argued to be more sensible and 'natural' than the frequently used 'Maximum Entropy' criterion for many applications. When applied to price distributions in conjunction Options constraints the minimum LCE criterion is shown to produce estimates which share the best theoretical properties of the Maximum Entropy approach with the best practical properties of the estimators identified by Jackwerth and Rubinstein.

1 Introduction
One of the important developments which has followed the emergence of international derivatives markets in the past several decades has been the new types of information about market
belief which may be extracted from these new types of nonlinear instruments. Initially, following the results of Black and Scholes on Option Valuation, the attention focused on the newly available Implied Volatility deducible by backsolving for the only unobservable parameter in an Option contract from actual traded Option price. Implicit in this, of course, was the assumption of underlying Lognormality of price processes, a condition which when violated would give different answers to different options on the same underlying price. Such an inconsistency did soon emerge, giving rise to speculation as to the ‘real’ distribution of underlying asset implied by several Options prices written thereon. Of course such type of information is never complete, giving rise to the need for a modelling approach for filling in that which is unknown or ambiguous, subject to the known constraints. A number of approaches then emerged, arguably the most ‘natural’ of which was the Maximum Entropy approach to probability modelling, put forward in the Options context by Edelman and Buchen (1995) and Buchen and Kelly (1996).

In principle, this method maximises the ‘randomness’ of a distribution in some sense, subject to known constraints, or alternately minimises the Entropy distance to some ‘prior’ distribution. Authors Jackwerth and Rubinstein (1996) performed a survey analysis of this and a number of other methods for extracting distributional information from traded Options prices, and found practical drawbacks with the Maximum Entropy approach, and recommended an alternative of maximum ‘smoothness’, though they did not put forward any justification of such an approach other than the fact that it seemed to work.

In what follows, a brief heuristic motivation of Entropy and Entropy Modeling in general will be presented, followed by an introduction to a new modelling approach which shares the appeal of the ‘naturalness’ of Maximum Entropy-type methods with the performance and ‘smoothness’ properties of approaches such as that put forward by Jackwerth and Rubinstein (1996).
In probability modeling problems where certain aspects of a desired distribution are assumed to be known, it is often considered a priority to be as ‘Objective’ as possible about other aspects about which nothing is known. For instance, for a variable with $n$ distinct possible outcomes about which nothing else is known, the most ‘Objective’ distribution would arguably be the equiprobable, which happens to also be the most ‘Entropic’ distribution, also often called the least ‘Informative’ distribution.

On the other hand, suppose for example that the distinct possible outcome values were given by $x_1, \ldots, x_n$ and it was known that the distribution defined by corresponding $p_1, \ldots, p_n$ were such that
\[ \sum_{i=1}^{n} x_i p_i = 0. \]
If were to adopt the measure of Entropy from Thermodynamics, Data Compression, and elsewhere
\[ E = - \sum p_i \log(p_i), \]
then we might wish to find the most Entropic distribution satisfying the previous 'first moment' condition.

This is a constrained optimisation problem in $p_1, \ldots, p_n$ which is easily solved using lagrange multipliers, and the functional form of the solution found to be
\[ p_i = p(x_i) = \exp(c_1 + c_2 x_i), \quad i = 1, \ldots, n \]
c_1 and $c_2$ chosen to satisfy the first-moment condition and $\sum p_i = 1$. Thus, given only partial information, here, the first moment condition, we have a specific solution, where the unspecified aspects of the distribution are assumed to be as 'Entropic' as possible. This is a Maximum Entropy solution.

However, it must be noted that a Maximum Entropy solution based on
\[ - \sum_{i=1}^{n} p_i \log(p_i) \]
is identical to a solution via Minimum Cross-Entropy to the equiprobable as given by

\[ \sum_{i}^{n} p_{i} \log \left( \frac{p_{i}}{1/n} \right) \]

Thus, another interpretation of such a solution is that it is as close to a certain 'reasonable' distribution (here given by the equiprobable) as possible. But on the other hand, perhaps there is another distribution which seems to be the 'most reasonable' in the absence of any other information, given by \( g_{1}, \ldots, g_{n} \). Then a reasonable modeling criterion might be the minimisation of

\[ \sum_{i}^{n} p_{i} \log \left( \frac{p_{i}}{g_{i}} \right) \]

given the other conditions. The difficulty encountered here is the standard Bayes problem of specifying a 'reasonable' prior. This might at first seem reason enough for discarding the approach altogether in favour of the canonical Maximum Entropy one, except for the fact that when variables are involved, solutions under variable transformation are inconsistent. If we wish to find the most Entropic function \( p(x) \) subject to \( \sum x p(x) = 0 \) and then solve the problem of finding the most Entropic function of \( \tilde{p}(y) \) subject to \( \sum y^3 p(y) \), where \( y_{i} = (x_{i})^{\frac{1}{3}} \), the first moment conditions are identical, but the resulting distributions will be very different in general. This suggests an element of arbitrariness in the choice of variable, which tends to undermine the intuition of "naturalness" of using the equiprobable distribution as a canonical yardstick.

This paradox may be resolved using Prior distributions, leading us (perhaps somewhat grudgingly) to the inevitable conclusion that Minimum Cross Entropy may not be any less "canonical" than Maximum Entropy.

However, the problem of how to choose the appropriate Prior distribution is not an easily solvable one, either philosophically nor technically.

Here, we propose a suggestion. For a given candidate distribution, why not use as a Prior distribution, a slightly-smoothed version of the candidate distribution itself? In engineering, the process of annealing is often applied, whereby a molten object is cooled with a series of gentle 'reheats' to achieve a much higher level of structural integrity than would have occurred had
this process not been applied. Similarly, in Numerical Optimisation problems, the technique of simulated annealing, whereby slight perturbations are made to successive candidate numerical solutions during optimisation, has been found to be very effective in many applications, especially those involving highly nonlinear systems.

So perhaps what is being proposed here can be referred to as a sort of “Probabilistic Annealing”, and perhaps when viewed in such a light might have some intuitive appeal.

3 Local Cross Entropy

As discussed above, instead of seeking a solution satisfying the constraints which is as close as possible to the Uniform (‘Maximum Entropy’) distribution, we shall investigate the criterion of minimising the cross-entropy of a distribution to a slightly smoothed version of itself. It therefore becomes necessary to make the notion of Local Cross Entropy precise.

To this end, imagine a kernel function $g(\cdot)$, having the properties of

$$g(s) \geq 0 \ \forall s$$

$$\int g(s)dx = \int s^2 g(s)ds = 1,$$

$$\int sg(s)ds = 0, \int s^3 g(s)ds = 0, \int s^4 g(s)ds = c_4$$

and for a given provisional density $f(x)$ define $f_\varepsilon(x)$ by

$$f_\varepsilon(x) = \int f(x-s)g(\frac{s}{\varepsilon})\frac{ds}{\varepsilon}$$

$$= f(x) + \frac{1}{2}\varepsilon^2 f''(x) + \frac{c_4}{24}\varepsilon^4 f'''(x) + O(\varepsilon^5)$$

as $\varepsilon \to 0$.

Henceforth, it will be helpful, though not necessary, to assume that

$$\int f^{(m)}(x)dx = f^{(m-1)}(\infty) - f^{(m-1)}(-\infty) = 0$$
for $m = 1, \ldots, 4$.

The Cross-Entropy from $f(\cdot)$ to $f_\varepsilon(\cdot)$, then (up to terms of order $\varepsilon^4$), is

$$\int f(x) \log \left\{ \frac{f(x)}{f_\varepsilon(x)} \right\} dx,$$

or

$$\int f(x) \log \left\{ \frac{f(x)}{f(x) + \frac{1}{2} \varepsilon^2 f''(x) + \frac{\varepsilon^4}{24} f^{(iv)}(x)} \right\}$$

$$= - \int f(x) \log \left\{ 1 + \frac{1}{2} \varepsilon^2 \frac{f''(x)}{f(x)} + \frac{c_4 \varepsilon^4}{24} \frac{f^{(iv)}(x)}{f(x)} \right\} dx$$

$$= - \frac{1}{2} \varepsilon^2 \int f''(x) dx + \frac{1}{8} \varepsilon^4 \int \frac{f''(x)^2}{f(x)} dx - \frac{c_4}{24} \varepsilon^4 \int f^{(iv)}(x) dx$$

yielding a leading term, as $\varepsilon \to 0$, proportional to

$$\int \frac{f''(x)^2}{f(x)} dx$$

which henceforth we shall refer to as the Local Cross Entropy (LCE).

### 4 The Minimum Local Cross-Entropy Criterion

In order to apply the Minimum Local Cross-Entropy (MLCE) Criterion with integral constraints

$$\int c_1(x) f(x) dx - k_1 = 0, \ldots, \int c_m(x) f(x) - k_m = 0,$$

we are faced with the problem of optimising

$$\int_{-\infty}^\infty \left[ \left\{ \frac{f''(x)}{f(x)} \right\}^2 + \sum \lambda_i c_i(x) \right] f(x) dx$$

Again, we adopt a variational approach by adding $t g(x)$ to $f(x)$ and differentiating.

$$0 = \frac{d}{dt} \int_{-\infty}^\infty \left[ \left\{ \frac{f''(x)}{f(x)} + t g''(x) \right\}^2 + \sum \lambda_i c_i(x) \right] f(x) + t g(x) ] dx$$

$$= \int_{-\infty}^\infty \left[ \frac{2 f''(x) g''(x)}{f(x)} - \frac{f''(x)^2 g(x)}{f(x)^2} \right] + \{ \sum \lambda_i c_i(x) \} g(x) dx,$$

which must equal zero for all $g(x)$. Taking

$$\lim_{x \to \pm \infty} g^{(m)}(x) = 0,$$
we may use integration by parts to transform this integral equation into

\[ \int_{-\infty}^{\infty} \left[ \frac{2f''(x)}{f(x)} - \frac{f''(x)}{f(x)} \right] dx + \sum \lambda_i c_i(x) g(x) dx = 0, \quad \forall g(x), \]

which is equivalent to

\[ 2h''(x) - \frac{1}{4} h(x)^2 + \sum \lambda_i c_i(x) = 0, \]

where \( h(x) = \frac{f''(x)}{f(x)} \).

Thus, the fourth-degree ordinary differential equation may be reduced to a system of two second-order ordinary differential equations which may be solved sequentially:

\[ 2h''(x) - \frac{1}{4} h(x)^2 + \sum \lambda_i c_i(x) = 0, \]

then

\[ \frac{f''(x)}{f(x)} = h(x). \]

In fact, the latter may be further reduced to the first-order sequence

\[ v'(x) + v(x)^2 = h(x), \]

and

\[ v(x) = \frac{f'(x)}{f(x)}. \]

While a general solution to the above system does not appear to exist as in the Maximum Entropy case, the above equations are not difficult to solve numerically, and analytically for some special cases.

For instance, it may be shown that if the constraints corresponding to the first four moments of the Gaussian are imposed, then the MLCE solution yields the Gaussian, while surprisingly (and unlike the Maximum Entropy case) the solution when the constraints are reduced to the first two moments is not the Gaussian but rather a slightly heavier-tailed density.

For instance, for \( c_1(x) = 1 \), \( c_2(x) = x^2 \), and \( c_3(x) = x^4 \), if the 3 constraints of

\[ \int f(x) dx - 1 = 0 \]
\[ \int x^2 f(x) dx - 1 = 0 \]
and
\[ \int x^4 f(x) dx - 3 = 0, \]
corresponding to the same moments of the Gaussian, are applied, and \( f(\cdot) \) is taken to be the standard Gaussian, it is easily seen that the differential equations are satisfied:

\[ f''(x)/f(x) = -1 + x^2 = h(x) \]

\[ 2h''(x) - h(x)^2 = 2 - (1 - 2x^2 + x^4) = -\lambda_0 - \lambda_1 x^2 - \lambda_2 x^4, \]
where \( \lambda_0 = -1, \lambda_1 = -2, \) and \( \lambda_2 = 1. \)

However, if the fourth moment constraint is removed, the system of differential equations does not have a straightforward solution, and numerical solution suggests that the expected fourth moment for the distribution optimised subject to unit mass and variance yields a fourth moment of approximately 3.39, somewhat larger than the value of 3 for the Gaussian.

For this case (i.e., \( c_1(x) = 1, c_2(x) = x^2, \) and
\[ \int f(x) dx - 1 = 0 \]
\[ \int x^2 f(x) dx - 1 = 0 \]
with no other constraints), inspection of the numerical solution suggests that as \( |x| \) increases, \( h''(x) \to 0. \)

If this is indeed the case, then the requirement that

\[ 2h''(x) - h(x)^2 + \lambda_1 + \lambda_2 x^2 = 0 \]
implies that for \( |x| \) large,

\[ h(x) \sim \sqrt{\lambda_2} |x|, \]

while for \( x \) near zero, inspection suggests that \( h''(x) = 0 \) also, implying that

\[ h(x) \sim h(0) + xh'(0) + \frac{1}{2} x^2 h''(0) = h(0) + \frac{1}{2} x^2 h''(0) \]

8
along with

\[ 2h''(x) - h(x)^2 + \lambda_1 + \lambda_2 x^2 = 0 \]

implies that

\[ 2h''(0) - \{h(0) + \frac{1}{2} x^2 h''(0)\}^2 + \lambda_1 + \lambda_2 x^2 = 0 \]

\[ h(x) \sim k_0 + k_1 x^2 \]

The result of numerical analysis of the above is that the MLCE distribution with second moment constrained is similar to the Gaussian, but with heavier tails. It is perhaps surprising that the distribution which is closest to its 'de-focused' version subject to fixed second moment is \textit{not} the Gaussian.

If an additional 4th moment constraint is added, the solution to

\[ h''(x) - \frac{1}{4} h(x)^2 + \lambda_0 + \lambda_1 x^2 + \lambda_2 x^4 = 0 \]

is solved for some special cases by

\[ h(x) = a_0 + a_1 x^2, \]

where

\[ a_1 = 2 \sqrt{\lambda_2} \]

\[ a_0 = \frac{\lambda_1}{\sqrt{\lambda_2}} \]

and it must be that

\[ \lambda_0 = 2 \frac{\lambda_1}{\sqrt{\lambda_2}} - \frac{\lambda_1}{\lambda_0^2} \]

This case (of \( f''(x) / f(x) = a_0 + a_1 x^2 \)) with \( a_0 = -\frac{1}{\sigma^2} \) and \( a_1 = \frac{1}{\sigma^4} \) is easily seen, by substitution and differentiation, to correspond to the Gaussian with mean zero and variance \( \sigma^2 \).

Next, we proceed to the case in which the aforementioned integral constraints are standard European Call options.
5 MLCE with Option constraints

Next, we follow ideas in Edelman and Buchen (1995) and the subsequent Buchen and Kelly (1996), then Jackwerth and Rubinstein (1996) and begin by applying Options constraints, along with the standard constraints of Risk-Neutrality. For the present discussion, we include only Equality constraints, which do suffer from not allowing for nonzero Bid-Ask spreads which occur in actual markets. The extension to inequality constraints is straightforward, and will be included in a later paper in the context of actual Market data.

In their paper, Jackwerth and Rubinstein identify the 'smoothness' criterion

\[ \int f''(x)^2 \, dx \]

as having the most desirable practical properties.

Below, we consider, primarily for pedagogical purposes an illuminating artificial example involving a Stock having three Call Options prices available, and compare the numerical solutions for the MLCE with those obtained via Maximum Entropy and J-R Smoothness.

Imagine an example in which the interest rate is 0, a stock is currently selling for $1.00 and three Options Prices at $.95, $1.00, and $1.05 exercise are priced at their respective Black-Scholes values of $0.1052, $0.0797, and $0.0593, corresponding to a constant implied volatility of 20%.

A numerical solution of the MLCE optimisation problem of maximising

\[ \int_0^\infty \frac{f''(x)^2}{f(x)} \, dx \]

subject to \( f(x) \geq 0, \) all \( x > 0, \) as well as the density constraint

\[ \int_0^\infty f(x) \, dx = 1, \]

the risk-neutral constraint

\[ \int_0^\infty xf(x) \, dx = 1, \]
and the three options constraints

\[ \int_{0.95}^{\infty} (x - .95) f(x) dx = 0.1052, \]
\[ \int_1^{\infty} (x - 1) f(x) dx = 0.0797, \]

and

\[ \int_{1.05}^{\infty} (x - 1.05) f(x) dx = 0.0593 \]

is straightforward and is graphed in Figure 1., with the theoretical Lognormal density consistent with these prices included for comparison [Figure 1. here]. The corresponding Maximum Entropy Solution (which turns out to be bimodal) is graphed in Figure 2. [Figure 2. here] The latter graph demonstrates the inherent instability (for which Buchen and Kelly have proposed some safeguard methods) of the Maximum Entropy method when the number of constraints supplied is small. The corresponding solution for the method recommended by Jackwerth and Rubinstein (1996), which involves the minimisation of

\[ \int_0^{\infty} f''(x)^2 dx \]

(proposed by them as an ad-hoc method with desirable practical properties) with the same constraints as above is not graphed here, but is similar in nature, with the exception that both tails are lighter than the MLCE (and theoretical) density.

Of interest as well is the ability to interpolate and extrapolate implied volatilities. In the above example, the (extrapolated) implied volatilities for the MLCE solution at exercises .85 and 1.15 reproduce very closely the constant level of the original at 20%, with only a slight increase for exercises further away from the money as shown in Figure 3. When applied to a volatility skew model with volatilities 25%, 20%, and 15% at respective exercises .95, 1, and 1.05, the (extrapolated) implied volatilities of the resulting MLCE solution at .85 and 1.15 exercise are approximately 34% and 9%, respectively. The Implied Volatility plot corresponding to this solution is shown in Figure 4.
Also, while computation algorithms were not optimised, the above optimisations each took less than a second in Linux-SCILAB on a 2 GHz processor for 100 grid points, or alternatively (with virtually no programming), approximately 2 minutes in Excel Solver on the same machine, though undoubtedly macros could reduce this to a matter of seconds or less.

6 Conclusion

A new criterion for density modelling has been introduced which appears to arise ’naturally’ in a manner reminiscent of Maximum Entropy, but which focuses more local distributional properties than global ones. As a result, the solutions appear, at least in the case of Options-type constraints, to have better stability, similar to that of the (ad-hoc) ’smoothness’ criterion identified by Jackwerth and Rubinstein, also sharing with the latter a lack of dependence on (subjective) prior information.

The MLCE criterion may also be used in a manner analogous to that in which Minimum Cross Entropy for (temporal) modeling of price processes was applied in Allevenada et. al (1997).

Subsequent to the submission of this paper, Hawkins and Friedan (2003) have recently generalised this approach and have found a Fisher Information interpretation of such methods, though as these authors note, the practical performance of their approach appears to be similar to MLCE.

References


**Acknowledgments**

The Author is indebted to Dr. Nigel O’Brian for comments which helped inspire the LCE, and to Jens Jackwerth and Mark Rubinstein for their helpful suggestions and data which was used to compare performance.