On Combining Statistical and Fuzzy Techniques: Detection of Business Cycles From Uncertain Data

Hung T. Nguyen
Dept. of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003
email hunguyen@nmsu.edu

Berlin Wu
Dept. of Mathematical Sciences
National Chengchi University
Taipei, Taiwan
email berlin@math.nccu.edu.tw

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
email vladik@cs.utep.edu

Abstract

Detecting the beginning and the end of the business cycle is an important and difficult economic problem. One of the reasons why this problem is difficult is that for each year, we have only expert estimates (subjective probabilities) indicating to what extent the economy was in growth or recession. In our previous papers, we used fuzzy techniques to process this uncertain information; namely, we used the operation $\min(a, b)$ to combine the subjective probabilities (expert estimates) of two events into a probability that both events happen. This function corresponds to the most optimistic estimate of the joint probability. In this paper, we use another operation which corresponds to the most cautious (pessimistic) estimate for joint probability. It turns out, unexpectedly, that as we get better extrapolations for subjective probabilities, the resulting change times become fuzzier and fuzzier until, for the best (least sensitive) extrapolation, we get the largest fuzziness. We explain this phenomenon by showing that in the presence of noise, an arbitrary continuous (e.g., fuzzy) system can be well described by its discrete analog, but as the description gets more accurate, the continuous description becomes necessary.

Introduction

Detecting a Business Cycle: an Important Economic Problem

Economy is changing in cycles: a growth period is followed by recession, and recession changes back to growth. It is extremely important to be able to predict the future economic behavior, and for this prediction, we must collect the statistics of the previous cycles. However, transitions are gradual, and it is therefore very difficult to find out when exactly growth changes into recession and vice versa. Both terms are not precisely defined, they express the expert's opinion and can be, therefore, best described *subjective probabilities* (*fuzzy values*).

Formalization of the Problem

Let us describe this problem in formal terms. Let us assume that we are analyzing a transition between growth and recession. We start with a year (let us denote it by 0) of clear growth; we know that at some following year T, we have a clear recession. We want to find the year when the change occurred, i.e., a year c which was a recession year, while the next year $c+1$ was a growth.

Let us denote the property "n was a recession year" by $R(n)$. Then, the opposite property (that n is a growth year) can be expressed as the negation $\neg R(n)$. We assume that for every year n, we know the subjective probability (fuzzy degree) $P(R(n))$ that this year was a recession year. We will denote this probability by $r(n)$. Then, the probability that n was a growth year is equal to

$$P(\neg R(n)) = 1 - P(R(n)) = 1 - r(n).$$

For each year n, the subjective probability that this year was a change year can be described as $P(R(n) \& \neg R(n+1))$. We want to produce a crisp answer; so, it is natural to choose a year n for which this probability is the largest possible, i.e., for which

$$P(R(n) \& \neg R(n+1)) \to \max_n . \quad (1)$$

The Main Difficulty of Solving This Problem

The main difficulty of solving this problem comes from the fact that although we know the subjective probabilities of individual events $R(n)$, we do not know the relation between these events and therefore, we cannot uniquely determine the joint probabilities like $P(R(n) \& \neg R(n+1))$.

Our Previous Work

Fuzzy methodology (see, e.g., [Klir et al. 1995], [Nguyen et al. 1999]) can be viewed as a technique which provides reasonable estimates of subjective probabilities of joint events in the situations in which we do not know the relation between these events.

Namely, fuzzy logic takes into consideration that in the situations when we know the subjective probabilities $P(A)$ and $P(B)$ of two events A and B, and this is the only information that we know about A and B, then we must somehow estimate our degree of belief $P(A \& B)$ in $A \& B$. Since this degree of belief must be computed based on the values $P(A)$ and $P(B)$, the desired estimate for $P(A\&B)$ must be a function of the known probabilities $a = P(A)$ and $b = P(B)$. This function is called an *&-operation* or a *t-norm*, and it is usually denoted by $f_\&(a,b)$.

The function $f_\&(a,b)$ must satisfy several reasonable conditions: e.g., if we want to estimate the degree of belief in a formula $A\&B\&C$, we can do it either by representing this formula as $(A\&B)\&C$, which leads to the estimate $f_\&(f_\&(a,b),c)$, or as $A\&(B\&C)$, which leads to an estimate $f_\&(a, f_\&(b,c))$. It is reasonable to require that these two estimates coincide, i.e., that $f_\&(a,b)$ is *associative*.

The simplest possible &-operation is $f_\&(a,b) = \min(a,b)$; this is the first operation proposed by L. Zadeh, the father of fuzzy logic, and so far, probably still the most widely used &-operation. In our previous papers [Kreinovich et al. 1999], [Kreinovich et al. 1999a], [Kreinovich et al. 1998], we used this simplest operation to formalize the problem of detecting the business cycle.

When we use this operation, the problem (1) turns into

$$\min(r(n), 1 - r(n+1)) \to \max_n . \qquad (2)$$

In our papers, we showed how to solve this optimization problem without using exhaustive search. The application of the resulting algorithm to Taiwan business cycle led to reasonable results [Wu et al. 1998] (see also [Wu et al. 1999]).

What We Are Planning To Do

In this paper, we start answering the natural next question: what if, instead of using the simplest possible &-operation, we use other &-operations? We will show that this approach leads to unexpected complications which, however, will be shown to be in line with general ideas about fuzziness.

New Approach: From Optimistic to Pessimistic Estimates

How Does $\min(a,b)$ Fit Into the Range of Possible And-Operations

It is known (see, e.g., [Klir et al. 1995], [Nguyen et al. 1999]) that there are many different &-operations $f_\&(a,b)$; all these operations are in between two extreme ones:

$$\min(a+b-1,0) \le f_\&(a,b) \le \min(a,b). \qquad (3)$$

This inequality does not require that we only consider &-operations, i.e., associative operations. It is also true for an arbitrary *copula*, i.e., for an arbitrary way of transforming the probabilities $a = P(A)$ and $b = P(B)$ of two events A and B into a probability of $A\&B$ (see, e.g., [Nelsen 1999]).

The $\min(a,b)$ operation that we used corresponds to the upper bound in the inequality (3). This upper bound $\min(a,b) = \min(P(A), P(B))$ is the largest possible probability of $P(A\&B)$. In other words, it represents the *optimistic* approach to determining the joint probability.

From Optimistic to Pessimistic Approach: A Natural Next Step

Since $\min(a,b)$ corresponds to the optimistic approach, a natural next step is to consider the *cautious* (*pessimistic*) approach, i.e., to consider, as an &-operation, the lower bound $\min(a+b-1,0)$ in the inequality (3) – the bound which represents the smallest possible probability of the joint event $A\&B$.

Using Pessimistic &-Operation In Detecting the Business Cycle

Let us see what happens when we use this pessimistic &-operation in the problem of detecting the business cycle. For this operation, we estimate the probability $P(R(n)\&\neg R(n+1))$ as

$$\max(P(R(n)) + P(\neg R(n+1)) - 1, 0),$$

i.e., as

$$\max(r(n) + 1 - r(n+1) - 1, 0) =$$

$$y\max(r(n) - r(n+1), 0).$$

Normally, we assume that there is a gradual transition between recession and growth and therefore, that the probabilities $r(n)$ gradually decrease. Under this assumption, the difference $r(n) - r(n+1)$ is always non-negative and therefore, the desired estimate for the probability $P(R(n)\&\neg R(n+1))$ is simply equal to this difference $r(n) - r(n+1)$. Therefore, the problem (1) can be formulated as follows:

$$r(n) - r(n+1) \to \max_n . \qquad (4)$$

A General Description of The Resulting Approach: Discrete Case

In deriving the formula (4), we did not use any specific features of this particular transition – from recession to growth. The same arguments can be repeated about an arbitrary gradual transition in which we need to select a single (crisp) "change point". In each such case, if the time is discrete, and if for every n, we know the degree of belief $r(n)$ that n belongs to the previous stage, then it is reasonable to select, as a crisp change point, the value n for which the change $r(n) - r(n+1)$ is the largest possible. When formulated in these terms, the condition (4) starts sounding very natural, irrespective of our arguments about pessimistic or optimistic &-operations.

Alternatively, instead of considering the degree of belief $r(n)$ that n belongs to the previous stage, we can consider the degrees of belief $\bar{r}(n) = 1 - r(n)$ that n belongs to the following stage. In this case, n is the value for which the change $\bar{r}(n+1) - \bar{r}(n)$ is the largest possible.

We can combine these two cases by saying that we choose n for which the change $|r(n) - r(n+1)|$ is the largest possible

$$|r(n) - r(n+1)| \to \max_n . \tag{5}$$

A General Description of The Resulting Approach: Continuous Case

In some real-life problems, the time is continuous. How can we then select the reasonable moment of change? A natural way to do it is to consider a sequence of closer and closer discrete approximations to continuous time. Namely, for each approximation, we select a small period Δt, and consider moments of time 0, Δt, $2\Delta t$, etc. As $\Delta t \to 0$, we get a better and better approximation of continuous time.

For each Δt, we select, as the moment of change, the moment of time t for which

$$|r(t) - r(t + \Delta t)| \to \max_t . \tag{6}$$

For small $\Delta t \approx 0$, the maximized expression in (6) is equal to

$$|r'(t) \cdot \Delta t + o(\Delta t)| = \Delta t \cdot (|r'(t)| + o(1)).$$

Since multiplication of an objective function by a positive constant Δt does not change the value t where the maximum is attained, optimizing (6) is equivalent to maximizing $|r'(t)| + o(1)$. For $\Delta t \to 0$, we therefore get the following result: the change time is the time t for which

$$|r'(t)| \to \max_t . \tag{7}$$

At first glance, this seems to be a reasonable approach, but, as we will see, the attempts to apply this approach lead to an unexpected problem.

A Problem with the New Approach to Detecting the Moment of Change

Extrapolation is Needed

In the discrete case, we can, in principle, elicit, from the expert, the value of subjective probability $r(n)$ for all n. However, in the continuous case, when we have infinitely many possible values of t, we cannot elicit the value $r(t)$ for all these t; all we can do is elicit some values of this membership function $r(t)$ and then extrapolate and/or interpolate.

In Continuous Case, the Value t Can Only Be Obtained by Using Measurements, Which Are Never 100% Accurate

In discrete case, we know the exact value of n. In the continuous case, when the value t can be an arbitrary real number, we need *measurement* to give us the value of t, and measurement is never 100% accurate. When we measure, then, due to inevitable noise, we get a value $\widetilde{t}$ which is, in general, slightly different from the actual value t.

To Make Extrapolated Values More Accurate, We Must Choose the Extrapolation Procedure Which is the Least Sensitive to the Measurement Uncertainty

Since the result $\widetilde{t}$ of measuring t comes with an uncertainty $\Delta t = \widetilde{t} - t$, it is reasonable to select the extrapolation procedure for which this uncertainty leads to the least possible influence on the resulting extrapolated values. Let us assume that we know the values $r_i = r(t_i)$ and $r_{t+1} = r(t_{i+1})$ for two values t_i and t_{i+1}. How can we interpolate from these values to the whole interval $[t_i, t_{i+1}]$?

The error in $r(t)$ caused by the measurement inaccuracy Δt is equal to

$$\Delta r = r\left(\widetilde{t}\right) - r(t) = r(t + \Delta t) - r(t).$$

For small measurement errors, we can linearize this expression and get $\Delta r = r'(t) \cdot \Delta t + o(\Delta t)$. Thus, the absolute value of this error is equal to $|r'(t)| \cdot \Delta t + o(\Delta t)$. We do not know which values t will be used, so, to estimate the quality of a given extrapolation $r(t)$, we can use the worst-case error

$$\max_t |r(t)| = \max_t (|r'(t)| \cdot \Delta t + o(\Delta t)) =$$

$$\Delta t \cdot \max_t |r'(t)| + o(\Delta t).$$

Thus, if we want to find the extrapolation procedure which is the least sensitive with respect to small measurement errors, we must look for a function $r(t)$ for which

$$\max_t |r'(t)| \to \min \tag{8}$$

among all functions $r(t)$ for which $r(t_i)$ and $r(t_{i+1})$ get the known values $r(t_i) = r_i$ and $r(t_{i+1}) = r_{i+1}$.

The problem of finding such *least sensitive* extrapolation procedure was solved in [Nguyen et al. 1995], [Nguyen et al. 1995a], [Nguyen et al. 1993], where we showed that the solution to this problem is to use *linear* extrapolation, for which the function $r(t)$ is linear:

$$r(t) = r_i + \frac{t - t_i}{t_{i+1} - t_i} \cdot (r_{i+1} - r_i). \tag{9}$$

The Unexpected Phenomenon

And here comes the problem: for this membership function (9), the derivative $r'(t)$ is constant, and so we cannot pick up a single point (7) for which this derivative attains the largest possible value.

In other words, when we try the best possible extrapolation procedure, the seemingly reasonable method of determining the crisp change point stops working.

How can we explain this phenomenon?

The Better Our Description, the More Fuzziness We Observe: An Explanation

Phenomenon Reformulated in General Terms

In general terms, we can reformulate the above phenomenon as follows: We want to get the best crisp results. For that, we use the best possible extrapolation, which is supposedly the most adequate for representing the original uncertainty. However, for this supposedly perfect representation of uncertainty, we get the fuzziest possible results instead of the desired crisp value.

How can we explain this phenomenon?

Basis for Our Explanation: Tsirelson's Theorem

B. S. Tsirelson noticed [Tsirelson 1982] that in many cases, when we reconstruct the signal from the noisy data, and we assume that the resulting signal belongs to a certain class, the reconstructed signal is often an *extreme* point from this class. For example, when we assume that the reconstructed signal is monotonic, the reconstructed function is often (piece-wise) constant; if we additional assume that the signal is smooth (one time differentiable, from the class C^1), the result is usually one time differentiable but rarely twice differentiable, etc.

Tsirelson provides an elegant *geometric* explanation to this fact: namely, when we reconstruct a signal from a mixture of a signal and a Gaussian noise, then the *maximum likelihood* estimation (a traditional statistical techniques) means that we look for a signal that belongs to the priori class, and that is the closest (in the L^2−metric) to the observed "signal+noise". In particular, if the signal is determined by finitely many (say, d) parameters, we must look for a signal $\vec{s}=(s_1,\ldots,s_d)$ from the a priori set $A\subseteq R^d$ that is the closest (in the usual Euclidean sense) to the observed values

$$\vec{o}=(o_1,\ldots,o_d)=(s_1+n_1,\ldots,s_d+n_d),$$

where n_i denotes the (unknown) values of the noise.

Since the noise is Gaussian, we can usually apply the *central limit theorem* and conclude that the average value of $(n_i)^2$ is close to σ^2, where σ is the standard deviation of the noise. In other words, we can conclude that

$$(n_1)^2+\ldots+(n_d)^2\approx d\cdot\sigma^2.$$

In geometric terms, this means that the distance $\sqrt{\sum(o_i-s_i)^2}=\sqrt{\sum n_i^2}$ between $\vec{s}$ and $\vec{o}$ is $\approx\sigma\cdot\sqrt{d}$. Let us denote this distance $\sigma\cdot\sqrt{d}$ by ε.

Let us (for simplicity) consider the case when $d=2$, and when A is a convex polygon. Then, we can divide all points p from the exterior of A that are ε−close to A into several zones depending on what part of A is the closest to p: one of the *sides*, or one of the *edges*. Geometrically, the set of all points for which the closest point $a\in A$ belongs to the *side* e is bounded by the straight lines orthogonal (perpendicular) to e. The total length of this set is is therefore equal to the length of this particular side; hence, the total length of all the points that are the closest to all the sides is equal to the *perimeter* of the polygon. This total length thus does not depend on ε at all. However, the set of all the points at the distance ε from A grows with the increase in ε; its length grows approximately as the growth of a circle, i.e., as const$\cdot\varepsilon$. When ε increases, the (constant) perimeter is a vanishing part of the total length. Hence, for large ε, the fraction of the points that are the closest to one of the sides tends to 0, while the fraction of the points p for which the *closest* is one of the *edges* goes to 1.

Similar arguments can be repeated for any dimension. For the same noise level σ, when d increases, the distance $\varepsilon=\sigma\cdot\sqrt{d}$ also increases, and therefore, for large d, for "almost all" observed points $\vec{o}$, the reconstructed signal is one of the extreme points of the a priori set A.

Much less probable is that the reconstructed signal belongs to the 1-dimensional face of the set A, even much less probable that s belongs to a 2-D face, etc.

The main *methodological consequence* of this result is that even when the actual state space is *continuous*, when we determine the state from measurements result, we inevitably obtain (most often) one of the *discretely many* states. On the large-scale level, we get one of the few *clusters*. When we add new measurements and thus, get to the next level, each original cluster subdivides into new clusters, etc., so that we get a *hierarchical* structure.

Explanation Itself

This result explains why in spite of the clearly fuzzy character of most human reasoning, binary logic describes most of this reasoning pretty well (see, e.g., [Starks at al. 1997], [Starks at al. 1998]). States with unusual "truth values" (different from 0 and 1) are not an exception, but rather a *general* rule. However, if we do the observations in the presence of some noise (e.g., if we use a not-prefect procedure for describing the values of the membership function), then we will mainly notice the *extreme* points of the set $[0,1]$ of the truth values, i.e., the values 0 and 1.

As observations become more accurate, we will observe the actual intermediate fuzzy values as well, and crisp description will become more and more difficult.

Comment. Similarly, we can explain Schrödinger's paradox in quantum mechanics (see Appendix).

Conclusion

In this paper, we consider the problem of characterizing a gradual change by a single moment of change.

In our previous papers, we have shown that the use of an "optimistic" method of combining subjective probabilities (fuzzy values) leads to a reasonable method for solving this problem. In this paper, we show that the use of a "cautious" method for combining subjective probabilities leads to a seemingly reasonable alternative method which, however, does not work if we use the "best possible" (least sensitive) techniques for describing the original uncertainty.

We show that this phenomenon is not simply a mathematical artifact: it seems to be related to the reasons why in some cases, in spite of a fuzzy nature of human reasoning, fuzzy models can be summarized by crisp models.

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF grants No. DUE-9750858 and CDA-9522207, by United Space Alliance, grant No. NAS 9-20000 (PWO C0C67713A6), by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518, and by the National Security Agency under Grant No. MDA904-98-1-0561 and MDA904-98-1-0564.

References

[Klir et al. 1995] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, Upper Saddle River, NJ, 1995.

[Kreinovich et al. 1999] V. Kreinovich, H. T. Nguyen, and B. Wu, "Justification of Heuristic Methods in Data Processing Using Fuzzy Theory, with Applications to Detection of Business Cycles From Fuzzy Data", *Proceedings of the 8th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'99)*, Seoul, Korea, August 22–25, 1999, Vol. 2, pp. 1131–1136.

[Kreinovich et al. 1999a] V. Kreinovich, H. T. Nguyen, and B. Wu, "Justification of Heuristic Methods in Data Processing Using Fuzzy Theory, with Applications to Detection of Business Cycles From Fuzzy Data", *East-West Journal of Mathematics*, 1999, Vol. 1, No. 2 (to appear).

[Kreinovich et al. 1998] V. Kreinovich, H. T. Nguyen, B. Wu, and K. T. Atanassov, "Fuzzy Justification of Heuristic Methods in Inverse Problems and in Numerical Computations, with Applications to Detection of Business Cycles From Fuzzy and Intuitionistic Fuzzy Data", *Notes on Intuitionistic Fuzzy Sets (NIFS)*, 1998, Vol. 4, No. 2, pp. 47–56.

[Nelsen 1999] R. B. Nelsen, *An Introduction to Copulas*, Springer-Verlag, N.Y., 1999 (Lecture Notes in Statistics, Vol. 139).

[Nguyen et al. 1995] H. T. Nguyen and V. Kreinovich, "Towards theoretical foundations of soft computing applications", *Int'l J. on Uncertainty, Fuzziness, and Knowledge-Based Systems,* 1995, Vol. 3, No. 3, pp. 341–373.

[Nguyen et al. 1995a] H. T. Nguyen, V. Kreinovich, R. N. Lea, and D. Tolbert. "Interpolation that leads to the narrowest intervals, and its application to expert systems and intelligent control", *Reliable Computing*, 1995, Vol. 1 No. 3, pp. 299–316.

[Nguyen et al. 1993] H. T. Nguyen, V. Kreinovich, D. Tolbert, "On robustness of fuzzy logics". *Proceedings of IEEE-FUZZ International Conference*, San Francisco, CA, March 1993, Vol. 1, pp. 543–547.

[Nguyen et al. 1999] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic,* CRC Press, Boca Raton, Florida, 1999.

[Starks et al. 1997] S. A. Starks, V. Kreinovich, and A. Meystel, "Multi-Resolution Data Processing: It is Necessary, It is Possible, It is Fundamental", *Proceedings of the International Conference on Intelligent Systems and Semiotics (ISAS'97)*, National Institute of Standards and Technology Publ., Gaithersburg, MD, 1997, pp. 145–150.

[Starks et al. 1998] S. A. Starks, V. Kreinovich, and A. Meystel, "Multi-Resolution Data Processing: It is Necessary, It is Possible, It is Fundamental", *Heuristics: The Journal of Intelligent Technologies*, 1997/98, Vol. 10, No. 2, pp. 51–58.

[Tsirelson 1982] B. S. Tsirel'son, "A geometrical approach to maximum likelihood estimation for infinite-dimensional Gaussian location. I", *Theory Probab. Appl.*, 1982, Vol. 27, pp. 411–418.

[Wu et al. 1999] B. Wu and S.-L. Hung, "Pattern recognition for nonlinear time series: with examples for distinguishing ARCH and bilinear models", *Fuzzy Sets and Systems*, 1999 (to appear).

[Wu et al. 1998] B. Wu and M.-C. Liaw, *Application of fuzzy time series analysis to change period detection*, Technical Report, Department of Mathematical Sciences and Statistics, National Chengchi University, Taipei, Taiwan, China, 1998.

Appendix: Explanation of Schrödinger's Paradox

In *classical physics*, it is assumed that for each state of a physical system, every property is either true or false. For example, a particle is either located in a certain interval of space coordinates $[x-\Delta, x+\Delta]$, or it is not located inside this interval.

In *quantum mechanics*, in addition to the states in which a particle is located within this interval, and to the states in which the particle is definitely outside it, there are states in which *some* measurements of the coordinate will lead to results *within* the interval, and *some* to the results *outside* this interval.

In such states, we cannot say that a statement "the particle is located in the given interval" is true or that this statement is false; at best, we can determine the *probability* of the "yes" answer. (To describe such unusual "truth value", quantum logic has been introduced.)

States with unusual "truth values" are not an exception, but rather a *general* rule in quantum mechanics: e.g., for every two states ψ and ψ' with certain values $\lambda \neq \lambda'$ of a measured quantity, there exists a state called their *superposition* in which the value of this quantity is no longer certain. (In the standard formalism of quantum mechanics, where states are described by vectors in a Hilbert space, superposition is simply linear combination.)

Such superposition state is easy to generate.

Schroedinger has shown that this *superposition principle* seemingly contradicts our intuition.

Indeed, suppose that we have a cat in a box, and a light-controlled rifle is aimed at the cat in such a way that a left-polarized photon would trigger the rifle and kill the cat, while the right-polarized photon would keep the cat alive.

If we send a photon with a circular polarization (that is, according to quantum mechanics, a superposition of left- and right-polarized states), we would get (due to the *linear* character of the equations of quantum mechanics), the *superposition* of the states resulting from using left- and right-polarized photons. In other words, we will get a *superposition of* a *dead and alive* cat states. This is, however, something that no one has ever observed: for macroscopic objects (cats included), an object is either dead or alive. Tsirelson's result explains why such non-extremal states are indeed difficult to observe.