Microeconomic Theory and the Kullback-Liebler Discrepancy: Some Remarkable Connections

John Fountain[a], Department of Economics, University of Canterbury, Christchurch, New Zealand
Peter E. Kennedy, Department of Economics, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

Abstract

Using a well-known utility of wealth function, the classic contingent claims model is shown to produce measures of expected wealth, certainty equivalence, gains from trade, and risk premia with relative entropy interpretations, as measured by the Kullback-Liebler discrepancy.

Author Keywords: Bayes factor; contingent claims; entropy; Pratt-Arrow risk
D8; G1

---

[a] Corresponding author. Department of Economics, University of Canterbury, Christchurch, New Zealand. John.fountain@canterbury.ac.nz

1. Introduction

Information theory has many applications in economics and econometrics. Maasoumi (1993) discusses examples in finance, industrial organization, marketing, statistical inference and model selection. These applications arise through information-based measures of concentration, inequality, aggregation, index numbers, volatility, mobility, and divergence. These measures are based on the concept of entropy, with the relative (or cross) entropy measure, known as the Kullback-Liebler (KL) discrepancy, playing a prominent role. The KL discrepancy is a logarithmic measure of discrepancy between two probability distributions, quantifying the gain in information from using the “true” distribution rather than a specific alternative distribution. For example, if John believes a random drawing is governed by q whereas Peter knows that the true probability is p, then the KL measure $D(p\|q)$ measures how much more informed Peter is relative to John about the possible outcome. Golan (2002) has a good discussion of the KL discrepancy interpretation.

One feature of this literature is that the applications to economics invariably stem from mechanical properties of information measures, or, most commonly, their statistical interpretations; a special issue of the Journal of Econometrics edited by Golan (2002) contains numerous examples. There is little or no connection to the fundamentals of economic theory, a deficiency this paper aims to remedy.

2. The classic contingent-claims model

A classic economic problem is how to allocate wealth between contingent claims, only one of which will pay off. The problem may be theoretical, as in a classic Arrow-Debreu general equilibrium analysis of complete contingent claims (Hirshleifer and Riley, 1992, Chapter 2; Gollier, 2001, chapter 13) , or practical, as in popular sports betting markets, or both, as in the Iowa Electronic markets studied by experimental economics. The purpose of this paper is to show that, remarkably, the traditional economic analysis of this problem in the context of a well-known utility of wealth function gives rise to measures of expected wealth, certainty equivalence, gains from trade, and risk premia that have interpretations in terms of relative entropy, as measured by the KL discrepancy.

Consider two contingent assets, one unit of the first of which pays one dollar if the event E occurs (and zero otherwise), and one unit of the second of which pays one dollar if the event E does not occur (and zero otherwise). The market has priced these assets at p and (1-p), revealing the market's view of the probability p that E will occur. An agent Q wishes to maximize her expected utility by purchasing x_1 and x_2 units of the first and second assets, respectively, subject to her budget constraint $px_1 + (1-p)x_2 \leq m$. We assume a negative exponential utility of wealth (w) function $u(w) = -\exp(-w/\tau)$. This well-known (Pratt, 1964; Hirschleifer and Riley, 1992; Gollier, 2001), tractable functional form exhibits constant absolute risk aversion (CARA) in the sense that her Pratt-Arrow risk tolerance index $\tau = -u'/u''$ is independent of her wealth.[1] Finally, Q believes that the probability that E will occur is q, not p.

Q's problem is to choose x_1 and x_2, subject to her budget constraint, to maximize

$$\text{expected utility} = q(-\exp(-x_1/\tau)) + (1-q)(-\exp(-x_2/\tau)).$$

Straightforward algebra yields the solutions

$$x_1^* = m + \tau\{D(p\|q) + \mathrm{Ln}(q/p)\}$$

$$x_2^* = m + \tau\{D(p\|q) + \mathrm{Ln}[(1-q)/(1-p)]\}$$

where $D(p\|q)$ is the KL discrepancy, or relative entropy, between two probabilty mass functions, in this case binomials with parameters p and q:

$$D(p\|q) = p\mathrm{Ln}(p/q) + (1-p)\mathrm{Ln}[(1-p)/(1-q)].$$

For those not familiar with this classic model from the contingent claims literature, a brief discussion of the meaning of this result is in order, before moving to an interpretation of the KL discrepancy measure.

If Q's assessment of the probability of E occurring matches that of the market, so that $q = p$, then x_1^* and x_2^* both equal m; Q will buy m units of the E asset at price p, and m units of the not E asset at price 1-p. This is a certain wealth of m, and so is interpreted as the

---

[1] Fountain (2004) extends the analysis reported here to n (rather than two) contingent claims, and non-CARA agents.

investor choosing not to buy either contingent claim. If Q's assessment of the probability of E occurring is greater than that of the market (q > p), we would expect her to buy more of the E asset and less of the not E asset (i.e, x_1* greater than x_2*). This is indeed the case in the formulas above, because when q > p, Ln(q/p) is positive and Ln[(1-q)/(1-p)] is negative. For sufficiently large m (which we assume throughout to make the results correspond to the real world) the expressions x_1* and x_2* indicate that positive amounts of both contingent assets will be bought. But as noted earlier, if x units of both assets are bought, the result is equivalent to a wealth of x with certainty. Consequently, the results x_1* and x_2* are interpreted to mean that the investor actually buys only the contingent asset with the greater x_i*, in an amount equal to the difference between the higher and lower of x_1* and x_2*.[2] For q > p, this implies that the investor will purchase the first asset in an amount equal to Ln(q/p) - Ln([(1-q)/(1-p)], and will purchase none of the second asset.

3. Interpretation

How might Ln(q/p) and Ln([(1-q)/(1-p)] be interpreted? Consider obtaining a new observation, which will be E or not E. If E is observed, then q/p would be the Bayes factor (Kass and Raftery, 1995; Lavine and Schervish, 1999) in favor of Q's hypothesis Hq that the binomial parameter is q, versus the market's hypothesis Hp that the binomial parameter is p. Similarly, (1-q)/(1-p) would be the Bayes factor in favor of Hq if the event not E were observed. The Bayes factor (q/p) is a measure of the change in the odds in favor of Hq resulting from observing E, and the Bayes factor (1-q)/(1-p) is the change

---

[2] An implication of the CARA utility function is that the amount of contingent claim purchase is not affected by wealth/income, as is evident from these algebraic results. The CARA utility function thus models price sensitive demands that are unresponsive to income, as many believe is the case for products such as tomatoes and peas. To many this is far superior to assuming risk neutrality or assuming that only variance of wealth matters (each of those utility functions is very restrictive). The CARA formulation suits a model of a typical investor who limits risk exposure each week to amounts that are price sensitive but not income sensitive.

in the odds in favor of Hq resulting from observing not E. Continuing with Bayesian terminology, Ln(q/p) is termed the *weight of the evidence* in favor of Hq (versus Hp) if E were to be observed, and Ln[(1-q)/(1-p)] is termed the weight of the evidence in favor of Hq if not E were to be observed.

From its algebraic definition, the KL discrepancy D(q‖p) is Q's *expected value* of the weight of the evidence in her favor (Hq over Hp) resulting from an additional observation. If the two probability mass functions (in this case the two binomials with parameters p and q) are very similar (in this case, p and q are close in value), an additional observation will not provide much evidence regarding which of the two distributions is correct. But if the two probabilities p and q differ markedly, an additional observation provides considerable information regarding which of the two distributions is correct. Consequently, a large expected weight of the evidence corresponds to distributions that differ markedly. This is the rationale behind the KL discrepancy as a measure of the difference between two distributions. In summary, Ln(q/p) in case E were to be observed, and Ln[(1-q)/(1-p)] in case not E were to be observed, provide weights of evidence for Hq against Hp; the KL discrepancy D(q‖p) is a measure of the difference between these two distributions, measuring the *expectation* for Q of the weight of the evidence in her favor.

To determine her optimal contingent wealth, Q looks at two measures. First, she assesses the difference between her and the market's probabilities, as measured by Ln (q/p), the weight of the evidence should E occur; and second, she measures the weight of the evidence should not E occur. Her decision is based on the difference between these two measures of the difference between the two probability mass functions. In this respect she could be said to follow the weight of the evidence in allocating her investment monies.

The KL discrepancy D(p‖q) = pLn(p/q) + (1-p)Ln[(1-p)/(1-q)] is the market's expectation of the weight of the evidence in its favor should another observation appear; the KL discrepancy D(q‖p) = qLn(q/p) + (1-q)Ln[(1-q)/(1-p)] is Q's expectation of the weight of the evidence in her favor. Using this result, straightforward algebra shows that

Q's expectation of her wealth is $ew = qx_1^* + (1-q)x_2^* = m + \tau[D(p\|q) + D(q\|p)]$. The sum of the two KL discrepancies, $D(p\|q) + D(q\|p)$, is Jeffrey's symmetric version (Soofi, 1994, p.1245) of the divergence between two distributions p and q. Hence, when a CARA agent invests optimally, subjectively expected wealth increases in proportion to Jeffrey's measure of divergence between the agent's beliefs and the market's beliefs, with the factor of proportionality being the agent's risk tolerance index. Since this measure is positive, when trading optimally Q always expects her wealth to be greater than what the market expects. This makes good sense: Q would not invest otherwise.

Wealth risk is also of concern to any CARA agent. The *certainty equivalent* (ce) for Q's optimal choice, the amount of certain money she regards as equivalent to her optimal choice of x_1^* and x_2^* (which embodies wealth risks), is found by setting her expected utility (eu) from her optimal strategy equal to her utility evaluated at the ce. This implies that $ce = -\tau Ln(-eu)$; substituting $eu = -q\exp(-x_1^*/\tau) - (1-q)\exp(-x_2^*/\tau)$, straightforward algebra yields $ce = m + \tau D(p\|q)$. As makes good sense, Q prefers market prices p that are farther from her own beliefs q, as measured by the KL discrepancy $D(p\|q)$. This discrepancy is the market's expectation of the log Bayes factor in its favor relative to Q's beliefs; when multiplied by Q's risk tolerance index, it is an exact measure of Hicksian compensating and equivalent variations in income, or gains from trade, for Q. This is the maximum amount she would pay to trade at prices p rather than not at all, and the mimimum amount she would have to be paid in compensation were the opportunity to trade be taken from her.

Earlier we saw that Q's expected wealth $ew = m + \tau[D(p\|q) + D(q\|p)]$. From this we find that an alternative expression for certainty equivalence is $ce = ew - \tau D(q\|p)$. This shows that Q values her optimum portfolio of contingent claims as her expected money wealth minus a *risk premium*, $\tau D(q\|p)$. This Pratt-Arrow risk premium is Q's expectation of the log Bayes factor in favor of her beliefs relative to the market's beliefs, times her risk tolerance index. As makes sense, the larger is the difference between Q's beliefs and the market's beliefs, as measured by the KL discrepancy D(q||p), the greater is the risk that Q faces.

4. Summary

In summary, we have shown that the Kullback Liebler discrepancy, remarkably, plays an important measurement role in microeconomic analysis of decisions regarding contingent wealth. In particular, it is an ingredient in an investor's measurement of expected wealth, certainty equivalent, gains from trade, and risk premium.

References

Fountain, J., 2004. Measuring Risk Relatively. University of Canterbury working paper.

Golan, A., 2002. Information and Entropy Econometrics – Editor's View. Journal of Econometrics 107, 1-15.

Gollier, C., 2001. The Economics of Risk and Time (MIT Press, Cambridge, Mass.)

Hirshleifer, J. and J. G. Riley, 1992. The Analytics of Uncertainty and Information (Cambridge University Press, Cambridge, UK)

Kass, R. and A. Raftery, 1995. Bayes Factors. Journal of the American Statistical Association 90, 773-795.

Lavine, M. and M. J. Schervish, 1999. The American Statistician 53, 119-122.

Maasoumi, E., 1993. A Compendium to Information Theory in Economics and Econometrics. Econometric Reviews 12, 137-181.

Pratt, J. W., 1964. Risk Aversion in the Small and in the Large. Econometrica 32, 122-136.

Soofi, E. S., 1994. Capturing the Intangible Concept of Information. Journal of the American Statistical Association 89, 1243-1254.