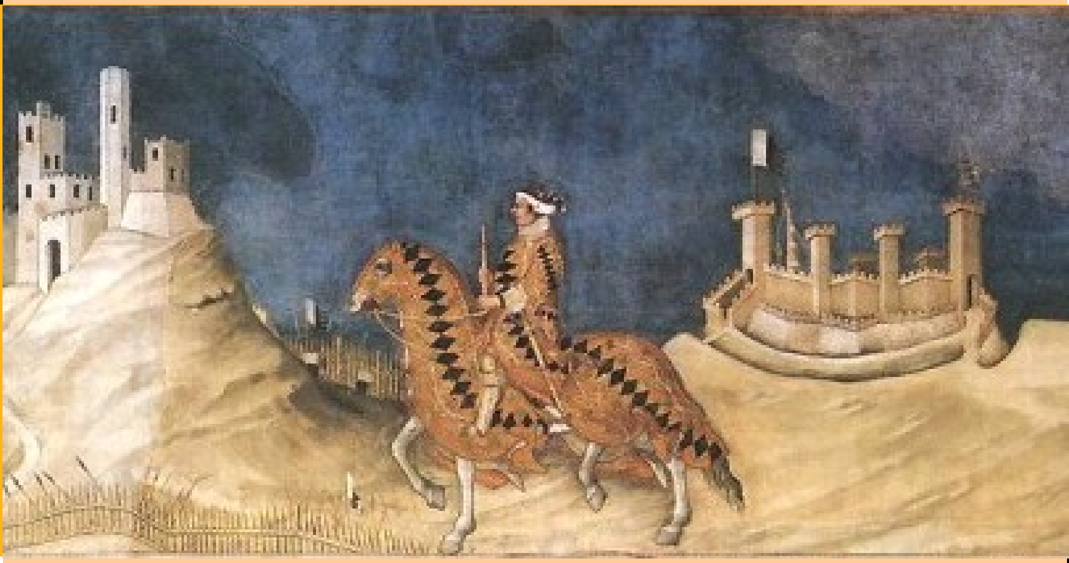
UNIVERSITÀ DI SIENA 1240

QUADERNI DEL DIPARTIMENTO
DI ECONOMIA POLITICA E STATISTICA

Marcello Basili
Luca Pratelli

Aggregation of not necessarily independent opinions

n. 677 – Giugno 2013



Abstract - We consider an aggregation scheme of opinions expressed through different probability distributions or multiple priors decision model. The decision-maker adopts entropy maximization as a measure of risk diversification and a rational form of prudence for valuing uncertain outcomes.
We show a new aggregation rule formalization based on the idea that the decision-maker has a more reliable set of outcomes called ordinary and two fat tails that include more ambiguous and extreme events.

Keywords: Ambiguity, Aggregation, Entropy, Multiple Priors, Quantiles.
JEL classification: D81.

Marcello Basili, DEPS, University of Siena; email: marcello.basili@unisi.it
Luca Pratelli, Naval Academy of Livorno; email: luca_pratelli@marina.difesa.it

1 Introduction

It is challenging to determine the probability distributions of decisive problems that affect human life, such as: climate change for future emission concentrations, prediction of weather statistics, hazards of GMO, re-emerging of infectious diseases, occurrence of pandemic flu etc.. It requires synthesis and assessment of uncertainty represented by experts' opinions about "the correctness of underlying data, models or analyses...or confidence about statistical analysis of a body of evidence (e.g. observations or model results)" (IPCC 2007).

The assessment of the likelihood of such occurrence has largely defied quantification due to insufficient data, limited ability to model the underlying processes and a vague knowledge Moreover when the experts' opinions are represented through probability distributions, such distributions could not be all independent and equally likely. In these scenarios, characterized by ambiguity and stochastically dependent experts' opinions, the Bayesian axiomatic approach to consensus distribution does not appear satisfying, not even in the sophisticated version of copula models, and scientists often resort to formal elicitation protocols. The elicitation protocols, because of interaction (sharing assessment) among experts (e.g. Delphy, Nominal Group Technique, Kaplan's approach), suffer from many problems such as polarization (Plous 1993), strategic manipulation, overconfidence, self-censorship, pressure to conformity, and more extreme probability estimates (Cooke 1991) in order to generate some kind of consensus distribution[1]. Moreover the behavioral combination approaches involve the experts' risk perception.[2] Risk perception is a specific concept different from expected return, that is concerned with evaluation rules and attractiveness, and that involves evaluation and attitude (relative and absolute) toward risk. Risk perception is a subjective multidimensional notion that encompasses interpretation (awareness, knowledge, information, familiarity, dread, etc.), likelihood of probability, and appraisement of possible consequences of feasible options.

The entropy and cross entropy methods are often used in Decision Theory and Statitics to elicit optimal probability distributions or combine them optimally. In economics an economic system is often considered like a physic system[3] and the entropy maximization is consequently seen as a principle of optimality.[4] We assume that the decision-maker (DM, henceforth) has a prudential attitude about uncertain experts' opinions. Facing the set of all probability distributions

[1] Formal elicitation protocols and methods have been developed to assess subjective probabilities of experts in different fields, e. g. the response of the climate system to future changes in radiative forcing (Zickfeld et al 2010) or the prospect of triggering major changes in the climate system (Kriegler et al 2009).

[2] A review of classifier combination methods and major advancements in this field can be found in Tulyakov *et al* 2008.

[3] See Mirowsky (1989), Tsallis et al (2003), Grunwald and Dawid (2004).

[4] In a recent paper Foley (2002) put in evidence that applying the methods of statistical mechanics to the standard economic model of pure exchange; "whereas the maximum entropy exchange equilibrium is the most probable Pareto-efficient, individually utility improving state of an exchange economy, Walrasian competitive equilibrium turns out to have a vanishingly small probability. Thus if we observe Walrasian competitive equilibrium, there must be very powerful forces sustaining it as an allocation".

assigned by experts to possible events, the DM adopts the entropy maximization principle as a rule of inference when the information is ambiguous and scanty.[5] As a result the DM focuses on the particular probability distribution π_j that maximizes the entropy. The elicited probability π_j is the *closest to uniformity* and the two tails, derived by assuming pessimism on catastrophic events and optimism on extremely positive events, originates the consensus distribution.

The paper is organized as follows: Section 2 provides a discussion of some related literature; Section 3 presents our model and introduces a new aggregation rule; Section 4 includes some concluding remarks.

2 Related Literature

There exist some problems to the full implementation of consensus rules in an axiomatic Bayesian approach to the expert priors and the DM's prior updating by a likelihood function, indeed the arbitrariness of the pooling weights, the use of invariant combination rules, the dependence between the DM's information and the experts' information, the dependence among experts' probability distributions (e.g. stochastic dependence), and the calibration of experts' opinion.[6]

The maximum entropy approach is known as a method to aggregate a set of opinions expressed through probability distribution functions, into a single one. Crucially, a probability distribution with higher entropy has a greater multiplicity and it is more capable of realization in Nature, that is more likely to occurr (Jaynes 1982). There are some axiomatized aggregation formulas based on the maximum entropy inference such as Levy and Delic (1994) and Miung *et al.* (1996). These formulas describe the combination of two or more experts' opinions in a single calibrated distribution. The calibration or competence measure reflects the quality of an expert's prediction. The competence measure is expressed by a real valued function that is a monotonically increasing function from 0 to 1, the absolute distance or quadratic absolute distance between the predicted event and the expert prediction. More complex it does appear to include pairwise interaction among experts because they share information sources, education backgrounds, theoretical dispositions, common training and experience. Levy and Delic and Miung *et al.* express statistical pairwise dependence (covariance) introducing correlation among experts' opinions in the constraints. With the exception of trivial cases of full and null dependence, the constrained problem of entropy maximization does not admit a closed form solution but requires the application of numerical methods to find the solution

[5] The maximization of entropy implies diversification, the simplest strategy of risk management. In fact it is well known that diversification, as opposed to concentration, is an efficient way to reduce likely loss, indeed it would induce a smaller loss with respect to a concentrative choice (DeMiguel et al 2009, Huang 2010 and Boyle et al 2012).

[6] In empirical studies Clemen and Winkler (1986) and Figlewski and Urich (1983) found that correlations among expert forecasts can be above 0.80.

such as Darroch and Ratcliff (1972), that introduced an iterative proportional fitting procedure, or Agmon *et al.* (1979) that proposed an efficient algorithm.

3 Entropy Maximization in a Quantile Utility Model

Our approach is meant to make a consensus distribution that adopts a quantile-function (Basili and Chateauneuf 2011). This is our setting: a DM has multiple priors, the set of all probability distributions of each expert on possible events, none of which is considered fully reliable. The set of probability distributions of all experts can be considered to reflect the DM's assessment of the reliability of the available information about the underlying uncertainty, that is her perception of ambiguity, and the suggested aggregation rule incorporates the DM's attitude about scanty and vague information.[7]

Facing the set of all experts' probability distributions about possible events, the DM focuses on the particular probability(ies) that maximizes the entropy, the $s-closest\ to\ uniformity$ probability distribution. Differently from other possible formulations, our new one appears to be less conservative and extreme: the DM is supposed to be pessimistic with respect to purely catastrophic losses, ambiguity neutral with respect to ordinary outcomes and optimistic with respect to purely windfall gains.[8]

So for any given act $X_{i=1,...,m} \in \mathcal{F}$ and $(\alpha_0, \beta_0) \in [0,1]^2, \alpha_0 \leq \beta_0$, such that it determines an interval of cumulative probability between which outcomes can be considered ordinary, we assume that the DM values these outcomes between these two quantiles in an ambiguity neutral way by π_j.

3.1 Set Up

Let S be the set of states of the world, $\mathcal{A}$ a σ-*algebra* of events in S and $(\pi_j)_{j=1,...,N}$ a class of probability measures on $(S, \mathcal{A})$.[9] Let $X_1, ..., X_m$ be a set of real random variables (*rrv*) with numerable support, technically $X_{i=1,...,m}$ is an act, such that $X_{i=1,...,m} : S \to \mathbb{R}$.

Definition 1 *The measure υ is a capacity on $(S, \mathcal{A})$ if $\upsilon : A \in \mathcal{A} \to \upsilon(A) \in \mathbb{R}$, where $\upsilon(\varnothing) = 0$, $\upsilon(S) = 1$ and such that, $(A, B) \in \mathcal{A}^2$, $A \subseteq B \Longrightarrow \upsilon(A) \leq \upsilon(B)$, then $\upsilon(A) = \min_{i,j} \pi_j(X_i \in A)$, $\forall A \in \mathcal{A}$.*

Definition 2 *The dual capacity $\overline{\upsilon}$ of a capacity υ is defined by $\overline{\upsilon}(A) = 1 - \upsilon(A^C) = \max_{i,j} \pi_j(X_i \in A)$, $\forall A \in \mathcal{A}$.*

[7] Different methods of aggregation based on multiple priors or capacity are in Cres *et al.* 2011 and Gajdos and Vergnaud 2012.

[8] See Basili and Chateauneuf (2011).

[9] Note that $\mathcal{A}$ is a subset of 2^S such that $\mathcal{A} = 2^S$ if S is finite.

Definition 3 *The core of a convex capacity* υ *on* $\mathcal{A}$ *is defined by* $C(\upsilon) := \{\pi_i : \pi_i(A) \geq \upsilon(A)\ \forall A \in \mathcal{A}\}$.

The core of a convex capacity υ includes the distributions $X_{i=1,\ldots,m}$ with respect to $(\pi_j)_{j=1,\ldots,N}$. It is worth to note that $\upsilon(a)$ and $\overline{\upsilon}(A)$ allows to take into account globally, i.e. on all acts, the experts evaluations and thus to evaluate the general reliability of everyone.

Consider $\alpha_0, \beta_0 \in]0,1[$ with $\alpha_0 \leq \frac{1}{2} < \beta_0$ and for any given $u \in [0,1]$, then indicate

$\alpha(u) = \alpha_o + (\frac{1}{2} - \alpha_o)u$ and $\beta(u) = \beta_o + (\frac{1}{2} - \beta_o)u$.

Let us also note that $\alpha(u) \leq \frac{1}{2} < \beta(u)$, for all $u \in [0,1]$.

Definition 4 *Call* $\overline{\upsilon} -$ *quantile with respect to* $\alpha(u)$ *and* $\overline{\upsilon} -$ *quantile with respect to* $\beta(u)$ *the values*

$$q'_u = \inf\{t : \overline{\upsilon}(]-\infty, t]) \geq \alpha(u)\} \quad \textit{and} \quad q''_u = \inf\{t : \overline{\upsilon}(]-\infty, t]) \geq \beta(u)\}.$$

Clearly $u \to q'_u$ (resp. $u \to q''_u$) is increasing (resp. decreasing) and $q'_u \leq q''_u$. We consider the $\overline{\upsilon} - quantile$ because it makes the interval between the quantiles wider, insofar we are coherent with a precautionary attitude (defective approach) of the DM.

It is possible to define the relative entropy restricted to the interval $[q'_u, q''_u]$:

Definition 5 *For any* $u \in [0,1]$ *and probability measure* π_j*, the relative entropy of* $X_i I_{\{X_{i \in [q'_u, q''_u]}\}}$ *with respect to* π_j *is defined by*

$$\tau_{i,j}(u) = -n_{i,j}(u) \sum_{x \in [q'_u, q''_u]} \pi_j(X_i = x) \log(\pi_j(X_i = x))$$

where $n_{i,j}(u) = \frac{1}{\log(\max(2, card\{x \in [q'_u, q''_u] : \pi_j(X_i = x) > 0\}))}$

In this setup the relative entropy $\tau_{i,j}(u)$ represents the degree of ignorance in term of closeness to uniform distribution, i.e. complete ignorance.

Definition 6 *For any* $s \in]0,1[$ *, the* $s -$ *level of uniformity of the relative entropy* $\tau_{i,j}$ *is defined by the real number* $u_{i,j}(s)$ *such that*

$u_{i,j}(s) = \inf\{u \in [0,1] : \tau_{i,j}(u) \geq s\}$, *where* $\inf \varnothing = 1$.

Interesting enough $u(s)$, i.e. the $s-level$ of uniformity, denotes the $\min u_{i,j}(s)$, that is the maximum interval of relative ignorance. The value $u(s)$ calibrates quantiles among all acts respect each expert, by assuming a $s - level$ of the relative entropy. It is worth to note that the choice of the parameter $s \in]0,1[$ expresses the relaibility of the experts evaluation and is related to the uniform distribution, i.e. the distribution corresponding to full relaibility or complete ignorance.

In other words, when the DM considers the experts predictions of very high quality, s is choosen close to 1 and the interval $[q'_u, q''_u]$ is very narrow, on the contrary when experts predictions are sloppy, s is choosen close to 0 and the interval $[q'_u, q''_u]$ is very large. Crucially, determination of the interval $[q'_u, q''_u]$ is endogenous and depends on reliability attached to experts opinions. Moreover according to the assumption on the $s - level\ of\ uniformity$, the DM considers the probability measure π^*_j hence there exists i^* such that the distribution of X_i according to π_j on the interval $[q'_u, q''_u]$. If there are more than one probability distributions that maximize entropy, then π^*_j is equal to their arithmetic mean.[10]

From the previous developments, it turns out that the DM determines what she conceives as the lower tail and the upper tail, through her choice of α_0, β_0, s and therefore computes the value of a *rrv* X_i, through the Choquet integral or

$J(X_i) = J_1(X_i) + J_2(X_i) + J_3(X_i)$ where

$J_1(X_i) = \displaystyle\int X_i I_{\left\{X_i < r'_i(s)\right\}} d\overline{\lambda}$;

$J_3(X_i) = \displaystyle\int X_i I_{\left\{X_i > r''_i(s)\right\}} d\lambda$;

$J_2(X_i) = E^{\pi^*_i}\left[X_i I_{\left\{r'_i(s) \leq X_i \leq r''_i(s)\right\}}\right]$ and

$\lambda = \min_j \pi_j$, $\overline{\lambda}$ is the dual capacity of λ; $r'_i(s) = \sup\{t : 1 - \alpha(u(s)) \leq \lambda(X_i > t)\}$; $r''_i(s) = \inf\left\{t : \overline{\lambda}(X_i \leq t) \geq \beta(u(s))\right\}$.

Crucially the definition of the interval $[\alpha_0, \beta_0]$ of the cumulative probability distribution could be related to the security integral and potential integral ensured by each act. Security level and potential level in preferences were introduced to accomodate generalized versions of expected utility theory and Allais paradoxes (Lopes 1987, Gilboa 1988, Yaffray 1988, Cohen 1992). The interval $[\alpha_0, \beta_0]$ of the cumulative probability distribution resembles the confidence interval, it contains all the outcomes that the DM considers more reliable (ordinary outcomes) and does not rely on the assumption of symmetry or constant shape. Confidence interval measures also the degree of variability of what the DM considers an ordinary outcome.[11]

It is possible to show that

Theorem 1 *For any s and X_i, X_j, such that $X_i \geq X_j$ then $J(X_i) \geq J(X_j)$*

Proof. Let the parameter $s \in]0, 1[$ be. It is a non-restrictive assumption to consider the random variables X_i and X_j with positive values on the finite set, the interval $[q'_u, q''_u]$. In other words, there exists a natural number $n \geq 2$ such that $X_i = \sum_{h=1}^n x_h I_{A_h}$ and $X_j = \sum_{h=1}^n y_h I_{A_h}$, and so $x_h \geq y_h$, $(x)_h$ and $(y)_h$ are non-decreasing sequences and $(A_h)_h$ is a partition of S. By induction on n, it is sufficient to prove the case for $n = 2$ and the case $x_h \neq y_h$, for a unique h_o. When $n = 2$, $x_h \neq y_h$ for a unique h_o ,it is straightforward to check that $J(X_i)$ and $J(X_j)$ are the integrals of X_i and X_j with respect

[10] There are some algorithms to determine the maximum entropy distribution, i.e. Agmon et al (1979) and Jaffray (1995).

[11] Details are in Basili and Chateauneuf 2011.

to a probability measure π^*. Note that $r_i^{'}(s), r_j^{'}(s), r_i^{''}(s), r_j^{''}(s)$ are elements of $\{x_1, ..., x_n, y_{h_0}\}$. If quantiles are different from x_{h_0} and y_{h_0}, the isotonicity of the Choquet Integral implies that $J(X_i)$ and $J(X_j)$ are the integrals of X_i and X_j. On the contrary, if $x_{h_0} = r_i^{'}(s)$ one obtains $y_{h_0} = r_j^{'}(s)$ and $J_1(X_i)$ $=J_1(X_j)$; $J_2(X_i) \geq J_2(X_j)$; $J_3(X_i) = J_3(X_j)$, and consequently $J(X_i)$ and $J(X_j)$ are the integrals of X_i and X_j. If $y_{h_0} = r_j^{''}(s)$ the proof is similar to the previous one ■

4 Concluding Remarks

In this paper we solve the problem of aggregating probability distributions of experts for a given set of events, by a new rule based on entropy maximization. Differently from previous models, we consider a prudent DM that facing a set of probability distributions elicits the consensus distribution by explicitly considering coherence and reliability of experts' opinions. In fact the interval of quantiles in which it is defined the maximum entropy probability distribution depends on the DM judgement on experts competence: the more they are considered skilled the narrower the interval is. Outside that variable interval, the DM is supposed to be pessimistic with respect to purely catastrophic losses and optimistic with respect to very positive events. In so far, our rule is less conservative and extreme, even if it explicitly considers events with very low probability but very large consequences.

5 References

Agmon N., Y. Alhassid, and R. Levine (1979), An algorithm for finding the distribution of maximal entropy, *Journal of Computational Physics*, 30, 250-258.

Basili M., and A. Chateauneuf (2011), Extreme events and entropy: a multiple quantile utility model, *International Journal of Approximate Reasoning*, 52, 1095–1102.

Boyle P., L. Garlappi, R. Uppal, and T. Wang (2012), Keynes meets Markowitz: The trade-off between familiarity and diversification, *Management Science*, 58, 253-272.

Clemen R., and R. Winkler (1986), Combining Economic Forecasts, *Journal of Business and Economic Statistics*, 4, 39-46.

Cohen, M. (1992). Security level, potential level, expected utility: A three criteria decision model under risk, *Theory and Decision*, 33, 101-104.

Cooke R. (1991), Experts in Uncertainty: Opinion and Subjective Probability in Science, Oxford University Press, Oxford.

Crès H., Gilboa I., and N. Vieille (2011), Aggregation of Multiple Prior Opinions, *Journal of Economic Theory*, 146, 2563-2582.

Darroch J., and D. Ratcliff (1972), Generalized Iterative Scaling for Log-Linear Models, *Ann. Math. Statistics*, 43, 1470-1480.

DeGroot M., and J. Montera (1991), Optimal Linear Opinion Pools, *Management Science*, 37, 546-558.

DeMiguel V., L. Garlappi, and R. Uppal (2009), Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies*, 22, 1915-1953.

Foley D. (2002), Maximum Entropy Exchange Equilibrium, mimeo,
http://cepa.newschool.edu/~foleyd/maxentexeq.pdf

Gajdos T. and J. Vergnaud (2012), Decisions with conflicting and imprecise information, *Social Choice and Welfare,* DOI 10.1007/s00355-012-0691-1

Genest C., McConway K, and M. Schervish (1986), Characterization of Externally Bayesian Pooling Operators, *Annals of Statistics*, 14, 487-501.

Genest C., and B. Remillard (2006), Comments on T. Mikosh's paper "Copulas: Tales and facts", *Extremes*, 9, 27-36.

Genest C., and J. Zidek (1986), Combining Probability Distributions: A Critique and an Annotated Bibliography, *Statistical Science*, 1, 114-148.

Gilboa, I. (1988). A combination of expected utility and maxmin decision criteria, *Journal of Mathematical Psychology*, 32, 405-420.

Grunwald P., and P. Dawid (2004), Game theory, maximum generalized entropy, minimum discrepancy and robust Bayesian decision theory, *The Annals of Statistics*, 32, 1367-1433.

IPCC (2007), Climate Change 2007: Synthesis Report.

Jaffray, J.Y. (1988), Choice under risk and the security factor: An axiomatic model, *Theory and Decision*, 24, 169-200.

Jaffray, J.Y. (1995), On the maximum entropy probability which is consistent with a convex capacity, I*nternational Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 3, 27-33.

Kriegler E., Hall J., Held H., Dawson R., and H. Schellnhuber (2009), Imprecise probability assessment of tipping points in the climate system, *Proceedings in the National Academy of Sciences of the United States of America*, 106, 5041-5046.

Mirowski P., (1989), More heat than light: Economics as social physics: Physics as nature's economics, New York: Cambridge University Press.

Lopes, L. L. (1987). Between hope and fear: The psychology of risk. In L. Berkowitz (ed.), Advances in experimental social psychology. New York: Academic Press.

Plous S. (1993), The Psychology of Judgment and Decision Making. New York: McGraw-Hill.

Stephanou H., and S. Lu (1988), Measuring consensus effectiveness by a generalized entropy criterion, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10, 544–554.

Stone M. (1961), The Opinion Pool, *Annals of Mathematical Statistics*, 32, 1339-1342.

Tsallis C., C. Anteneodo, L. Borland, and R. Osorio (2003), Nonextensive statistical mechanics and economics, *Physica A,* 324, 89-100.

Tulyakov S., Z. Zhang, and V. Govindaraju (2008), Comparison of Combination Methods Utilizing T-normalization and Second Best Score Model, CVPR

2008 Workshop on Biometrics.

Winkler R. (1968) The Consensus of Subjective Probability Distributions, *Management Science*, 15, 61-75 .

Winkler R. (1986). Expert Resolution, *Management Science*, 32, 298-303.

Xiaoxia Huang (2010), Portfolio Analysis: from probabilistic to credibilistic and uncertain approaches. Springer-Verlag Berlin Heidelberg.

Yu D. (1997), Expert opinion elicitation process using a fuzzy probability, *Journal of the Korean Nuclear Society,* 29, 25-34.

Zickfeld K., Morganb M., Frame D., and D. Keithd (2010), Expert judgments about transient climate response to alternative future trajectories of radiative forcing, *Proceedings in the National Academy of Sciences of the United States of America*, 107, 12451-12456.