Mixing Probabilities, Priors and Kernels via Entropy Pooling

Attilio Meucci[1]

published[2]: December 1 2011
this revision: December 07 2011
last revision and code at `http://symmys.com/node/353`

Abstract

We show how to mix machine learning signals such as kernel smoothing and fuzzy memberships via the Entropy Pooling approach by Meucci (2008). We illustrate a case study, where we overlay an exponentially time-decayed prior to a pseudo-Gaussian kernel that emphasizes market scenarios where volatilities and interest rates are similar to today's levels.

The code for the case study is available at `http://symmys.com/node/353`.

JEL Classification: C1, G11

Keywords: Black-Litterman, effective number of scenarios, fuzzy membership, Mahalanobis distance, pseudo-Gaussian kernel, crisp conditioning, exponential decay, machine learning

[1] The author is grateful to Garli Beibi and David Elliott

[2] This article appears as Meucci, A. (2011), *Mixing Probabilities, Priors and Kernels via Entropy Pooling* - GARP Risk Professional, December 2011, p. 32-36

1 Introduction

The Fully Flexible Probabilities framework discussed in Meucci (2010) represents the multivariate distribution f of an arbitrary set of risk drivers $\mathbf{X} \equiv (X_1, \ldots, X_N)'$ non-parametrically in terms of scenario-probability pairs

$$f \iff \{\mathbf{x}_j, p_j\}_{j=1,\ldots,J}, \tag{1}$$

where the joint scenarios $\mathbf{x}_j \equiv (x_{1,j}, \ldots, x_{N,j})'$ can be historical realizations or the outcome of simulations. The use of Fully Flexible Probabilities permits all sorts of manipulations of distributions essential for risk and portfolio management, such as pricing and aggregation, see Meucci (2011), and the estimate of portfolio risk from these distributions.

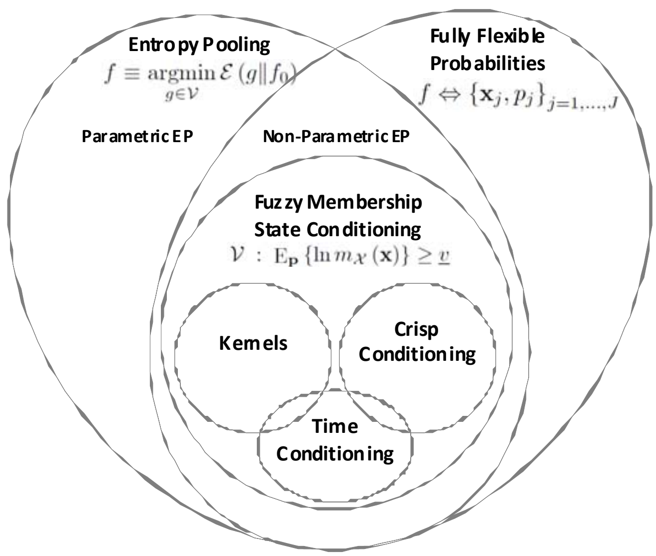

Figure 1: Fully Flexible Probabilities specification via Entropy Pooling

The probabilities in the Fully Flexible Probabilities framework (1) can be set by crisp conditioning, kernel smoothing, exponential time decay, etc., refer to Figure 1.

Another approach to set the probabilities in (1) is based on the Entropy Pooling technique by Meucci (2008). Entropy Pooling is a generalized Bayesian approach to process views on the market. Entropy Pooling starts from two inputs, a prior market distribution f_0 and a set of generalized views or stress-tests $\mathcal{V}$, and yields a posterior distribution f that is close to the prior, but incorporates the views. Entropy Pooling can be used in the non-parametric scenario-probability representation of the Fully Flexible Probabilities framework (1), in which case it provides an optimal way to specify the probabilities $\mathbf{p} \equiv$

$(p_1, \ldots, p_J)'$ of the scenarios. Alternatively, Entropy Pooling can be used with parametric distributions $f^{\boldsymbol{\theta}}$ that are fully specified by a set of parameters $\boldsymbol{\theta}$, such as the normal distribution.

In this article, we show that Entropy Pooling represents the most general approach to optimally specify the probabilities of the scenarios (1) and includes common approaches such kernel smoothing as special cases, when the prior distribution f_0 contains no information. Then we use Entropy Pooling to overlay different kernels/signals to an informative prior in a statistically sound way.

In the remainder of the article we proceed as follows. In Section 2 we review common approaches to assign probabilities in (1) and we generalize such approaches using fuzzy membership functions.

In Section 3 we review the non-parametric implementation of Entropy Pooling. Furthermore, we show how Entropy Pooling includes fuzzy membership probabilities. Finally, we discuss how to leverage Entropy Pooling to mix the above approaches and overlay prior information to different estimation techniques.

In Section 4 we present a risk management case study, where we model the probabilities of the historical simulations of a portfolio P&L. Using Entropy Pooling we overlay to an exponentially time-decayed prior a kernel that modulates the probabilities according to the current state of implied volatility and interest rates.

2 Time/state conditioning and fuzzy membership

Here we review a few, popular methods to exogenously specify the probabilities in the Fully Flexible Probabilities framework (1). For applications of these methods to risk management, refer to Meucci (2010).

If the scenarios are historical and we are at time T, a simple approach to exogenously specify probabilities is to discard old data and rely only on the most recent window of time τ. This entails setting the probabilities in (1) as follows

$$p_j \propto \begin{cases} 1 & \text{if } t_j > T - \tau \\ 0 & \text{otherwise.} \end{cases} . \tag{2}$$

An approach to assign weights to historical scenarios different from the rolling window (2) is exponential smoothing

$$p_j \propto e^{-\frac{\ln 2}{\tau}|t_j - T|}, \tag{3}$$

where $\tau > 0$ is a given half-life for the decay.

An approach to assigning probabilities related to the rolling window (2) that does not depend on time, but rather on state, is crisp conditioning: the probabilities are set as non-null, and all equal, as long as the scenarios of the market drivers $\mathbf{x}_j$ lie in a given domain $\mathcal{X}$. Using the indicator function $1_{\mathcal{X}}(\mathbf{x})$, which is 1 if $\mathbf{x} \in \mathcal{X}$ and 0 otherwise, we obtain

$$p_j \propto 1_{\mathcal{X}}(\mathbf{x}_j) \tag{4}$$

An enhanced version of crisp conditioning for assigning probabilities, related to the rich literature on machine learning, is kernel smoothing. First, we recall that a kernel $k_{\epsilon,\boldsymbol{\mu}}(\mathbf{x})$ is defined by a positive non-increasing generator function $k(d) \in [0,1]$, a target $\boldsymbol{\mu}$, a distance function $d(\mathbf{x},\mathbf{y})$, and a radius, or bandwidth, ϵ as follows

$$k_{\epsilon,\boldsymbol{\mu}}(\mathbf{x}) \equiv k(\frac{d(\mathbf{x},\boldsymbol{\mu})}{\epsilon}). \tag{5}$$

For example, the Gaussian kernel reads

$$k_{\epsilon,\boldsymbol{\mu}}(\mathbf{x}) \equiv e^{-\frac{1}{2\epsilon^2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \tag{6}$$

where the additional parameter $\boldsymbol{\sigma}$ is a symmetric, positive definite matrix. The Gaussian kernel (6) is in the form (5), where d is the Mahalanobis distance $d^2(\mathbf{x},\boldsymbol{\mu}) \equiv (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$.

Using a kernel we can condition the market variables $\mathbf{X}$ smoothly, by setting the probability of each scenario as proportional to the kernel evaluated on that scenario

$$p_j \propto k_{\epsilon,\boldsymbol{\mu}}(\mathbf{x}_j). \tag{7}$$

As we show in Appendix A.1, the crisp state conditioning (4) includes the rolling window (2) as a special case, and kernel smoothing (7) includes the time decayed exponential smoothing (3) as a special case, see Figure 1.

We can generalize further the concepts of crisp conditioning and kernel smoothing by means of fuzzy membership functions. Fuzzy membership to a given set $\mathcal{X}$ is defined in terms of a "membership" function $m_{\mathcal{X}}(\mathbf{x})$ with values in the range $[0,1]$, which describes to what extent $\mathbf{x}$ belongs to a given set $\mathcal{X}$. Using a fuzzy membership $m_{\mathcal{X}}$ for a given set $\mathcal{X}$ of potential market outcomes for the market $\mathbf{X}$, we can set the probability of each scenario as proportional to the degree of membership of that scenario to the set of outcomes

$$p_j \propto m_{\mathcal{X}}(\mathbf{x}_j). \tag{8}$$

A trivial example of membership function is the indicator function that defines crisp conditioning (4)

$$m_{\mathcal{X}}(\mathbf{x}) \equiv 1_{\mathcal{X}}(\mathbf{x}). \tag{9}$$

With the indicator function, membership is either maximal, and equal to 1, if $\mathbf{x}$ belongs to $\mathcal{X}$, or minimal, and equal to 0, if $\mathbf{x}$ does not belong to $\mathcal{X}$.

A second example of membership function is the kernel (5), which is a membership function for the singleton $\mathcal{X} \equiv \boldsymbol{\mu}$

$$m_{\boldsymbol{\mu}}(\mathbf{x}) \equiv k_{\epsilon,\boldsymbol{\mu}}(\mathbf{x}). \tag{10}$$

The membership of $\mathbf{x}$ to $\boldsymbol{\mu}$ is maximal when $\mathbf{x}$ is $\boldsymbol{\mu}$. The larger the distance d of $\mathbf{x}$ from $\boldsymbol{\mu}$, the less $\mathbf{x}$ "belongs" to $\boldsymbol{\mu}$ and thus the closer to 0 the membership of $\mathbf{x}$.

3 Entropy Pooling

The probability specification (8) assumes no prior knowledge of the distribution of the market risk drivers $\mathbf{X}$. In this case, fuzzy membership is the most general approach to assign probabilities to the scenarios in (1). Here we show how non-parametric Entropy Pooling further generalizes fuzzy membership specifications to the case when prior information is available, or when we must blend together more than one membership specification.

First, we review the non-parametric Entropy Pooling implementation. Please refer to the original article Meucci (2008) for more details, more generality, and for the code.

The starting point for non-parametric Entropy Pooling is a prior distribution $f^{(0)}$ for the risk drivers $\mathbf{X}$, represented as in (1) by a set of scenarios and associated probabilities

$$f_0 \iff \{\mathbf{x}_j, p_j^{(0)}\}_{j=1,\ldots,J}. \tag{11}$$

The second input is a view on the market $\mathbf{X}$, or a stress-test. Thus a generalized view on $\mathbf{X}$ is a statement on the yet-to-be defined distribution defined on the same scenarios $f \iff \{\mathbf{x}_j, p_j\}_{j=1,\ldots,J}$. A large class of such views can be characterized as expressions on the expectations of arbitrary functions of the market $v(\mathbf{X})$.

$$\mathcal{V} : \mathrm{E}_{\mathbf{p}} \{v(\mathbf{X})\} \geq v_*, \tag{12}$$

where v^* is a threshold value that determines the intensity of the view.

To illustrate a typical view, consider the standard views a-la Black and Litterman (1990) on the expected value $\boldsymbol{\mu}_{\mathbf{aX}}$ of select portfolios returns $\mathbf{aX}$, where $\mathbf{X}$ represents the returns of N securities, and $\mathbf{a}$ is a $K \times N$ matrix, whose each row are the weights of a different portfolio. Such view can be written as in (12), where $v(\mathbf{X}) \equiv (\mathbf{a}', -\mathbf{a}')'$ and $v_* \equiv (\boldsymbol{\mu}_{\mathbf{aX}}, -\boldsymbol{\mu}'_{\mathbf{aX}})'$.

Our ultimate goal is to compute a posterior distribution f which departs from the prior to incorporate the views. The posterior distribution f is specified by new probabilities $\mathbf{p}$ on the same scenarios (11). To this purpose, we measure the "distance" between two sets of probabilities $\mathbf{p}$ and $\mathbf{p}_0$ by the relative entropy

$$\mathcal{E}(\mathbf{p}, \mathbf{p}_0) \equiv \mathbf{p}' (\ln \mathbf{p} - \ln \mathbf{p}_0). \tag{13}$$

The relative entropy is a "distance" in that (13) is zero only if $\mathbf{p} = \mathbf{p}_0$ and it becomes larger as $\mathbf{p}$ diverges away from $\mathbf{p}_0$.

Then we define the posterior as the closest distribution to the prior, as measured by (13), which satisfies the views (12)

$$\tilde{\mathbf{p}} \equiv \underset{\mathbf{p} \in \mathcal{V}}{\operatorname{argmin}} \, \mathcal{E}(\mathbf{p}, \mathbf{p}_0), \tag{14}$$

where the notation $\mathbf{p} \in \mathcal{V}$ means that $\mathbf{p}$ satisfies the view (12).

Applications of Entropy Pooling to the probabilities in the Fully Flexible Probabilities framework are manifold. For instance, with Entropy Pooling, we can compute exponentially decayed covariances where the correlations and the variances are decayed at different rates. Other applications include conditioning the posterior according to expectations on a market panic indicator. For more details see Meucci (2010).

As highlighted in Figure 1, the Entropy Pooling posterior (14) also includes as special cases the probabilities defined in terms of fuzzy membership functions (8). Indeed, let us assume that in the Entropy Pooling optimization (14) the prior is non-informative, i.e.

$$\mathbf{p}_0 \propto \mathbf{1}. \tag{15}$$

Furthermore, let us assume that in the Entropy Pooling optimization (14) we express the view (12) on the logarithm of a fuzzy membership function $m_{\mathcal{X}}(\mathbf{x}) \in [0,1]$

$$v(\mathbf{x}) \equiv \ln\left(m_{\mathcal{X}}(\mathbf{x})\right). \tag{16}$$

Finally, let us set the view intensity in (12) as

$$v^* \equiv \frac{\sum_{j=1}^{J} m_{\mathcal{X}}(\mathbf{x}_j) \ln m_{\mathcal{X}}(\mathbf{x}_j)}{\sum_{j=1}^{J} m_{\mathcal{X}}(\mathbf{x}_j)}. \tag{17}$$

Then, as we show in Appendix A.2, the Entropy Pooling posterior (14) reads

$$\tilde{p}_j \propto m_{\mathcal{X}}(\mathbf{x}_j). \tag{18}$$

This means that, without a prior, the Entropy Pooling posterior is the same as the fuzzy membership function, which in turn is a general case of kernel smoothing and crisp conditioning, see also the examples in Appendix A.3.

Not only does Entropy Pooling generalize fuzzy membership, it also allows us to blend multiple views with a prior. Indeed, let us suppose that, unlike in (15), we have an informative prior $\mathbf{p}^{(0)}$ on the market $\mathbf{X}$, such as for instance the exponential time decay predicate (3) that recent information is more reliable than old information. Suppose also that we would like to condition our distribution of the market based on the state of the market using a membership function as in (18), such as for instance a Gaussian kernel. How do we mix these two conflicting pieces of information?

Distributions can be blended in a variety of ad-hoc ways. Entropy Pooling provides a statistically sound answer. Indeed, we simply replace the non-informative prior (15) with our informative prior $\mathbf{p}^{(0)}$ in the Entropy Pooling optimization (14) driven by the view (12) on the log-membership function (16) with intensity (17). In summary, the optimal blend reads

$$\begin{aligned} \mathbf{p} \quad &\equiv \quad \underset{\mathbf{q}}{\operatorname{argmin}}\, \mathcal{E}(\mathbf{q}, \mathbf{p}^{(0)}). \\ &\quad\ \text{where } \ \mathrm{E}_{\mathbf{q}}\left\{\ln\left(m_{\mathcal{X}}(\mathbf{X})\right)\right\} \geq v_*. \end{aligned} \tag{19}$$

More in general, we can add to the view in (19) other views on different features of the market distribution, as in Meucci (2008).

It is worth emphasizing that all the steps of the above process are computationally trivial, from setting the prior $\mathbf{p}^{(0)}$ to setting the constraint on the expectation, to computing the posterior (19). Thus the Entropy Pooling mixture is also practical.

4 Case study: conditional risk estimates

Here we use Entropy Pooling to mix information on the distribution of a portfolio's historical simulations. A standard approach to risk management relies on so-called historical simulations for the portfolio P&L: current positions are evaluated under past realizations $\{\mathbf{x}_t\}_{t=1,\ldots,T}$ of the risk drivers $\mathbf{X}$, giving rise to a history of P&L's $\{\pi_t\}_{t=1,\ldots,T}$. To estimate the risk in the portfolio, one can assign equal weight to all the realizations. The more versatile Fully Flexible Probability approach (1) allows for arbitrary probability weights

$$f \iff \{(\pi_t, \mathbf{x}_t), p_t\}_{t=1,\ldots,T}, \tag{20}$$

refer to Meucci (2010) for more details.

In our case study we consider a portfolio of options, whose historically simulated daily P&L distribution over a period of ten years is highly skewed and kurtotic, and definitely non-normal. Using Fully Flexible Probabilities, we model the exponential decay prior that recent observations are more relevant for risk estimation purposes

$$p_t^{(0)} \propto e^{-\frac{\ln 2}{\tau}(T-t)}, \tag{21}$$

where, τ is a half-life of three years. We plot these probabilities in the top portion of Figure 2.

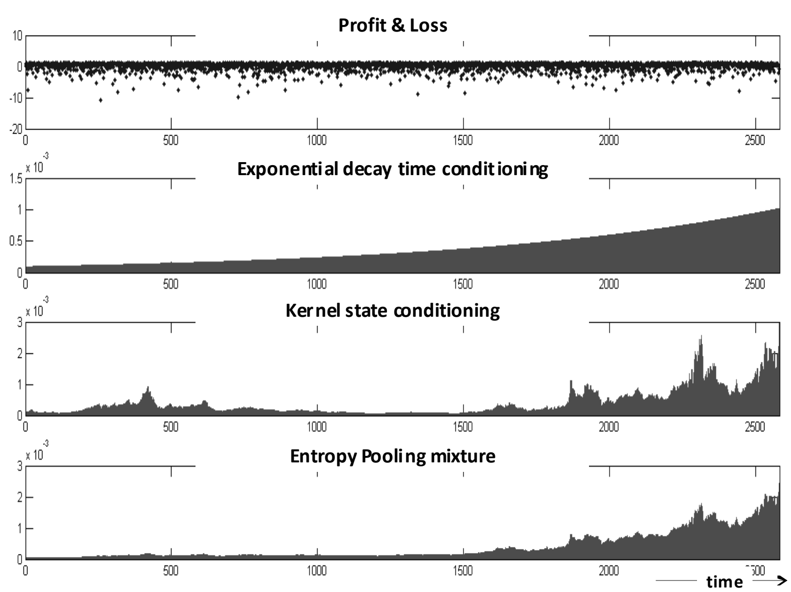

Figure 2: Mixing distributions via Entropy Pooling

Then we condition the market on two variables: the five-year swap rate, which we denote by X_1, and the one-into-five swaption implied at-the-money volatility, which we denote by X_2. We estimate the 2×2 covariance matrix $\boldsymbol{\sigma}^2$ of these variables, and we construct a quasi-Gaussian kernel, similar to (6), setting as target the current values $\mathbf{x}_T$ of the conditioning variables

$$m_t \equiv \exp(-\tfrac{1}{2\epsilon^2}[(\mathbf{x}_t - \mathbf{x}_T)' (\boldsymbol{\sigma}^2)^{-1} (\mathbf{x}_t - \mathbf{x}_T)]^\eta). \tag{22}$$

In this expression the bandwidth is $\epsilon \approx 10$ and $\eta \approx 0.4$ is a power for the Mahalanobis distance, which allows for a smoother conditioning than $\eta = 2$.

If we used directly the membership levels (22) as probabilities $p_t \propto m_t$, we would disregard the prior information (21) that more recent data is more valuable for our analysis. If we used only the exponentially decayed prior (21), we would disregard all the information conditional on the market state (22). To overlay the conditional information to the prior, we compute the Entropy Pooling posterior (19), which we write here

$$\mathbf{p} \equiv \underset{\mathbf{q}' \ln \mathbf{m} \geq v_*}{\operatorname{argmin}} \ \mathcal{E}(\mathbf{q}, \mathbf{p}^{(0)}). \tag{23}$$

Notice that for each specification of the kernel bandwidth ϵ and radius depth η in (22) we obtain a different posterior. Hence, a further refinement of the proposed approach lets the data determine the optimal bandwidth, by minimizing the relative entropy in (23) as a function of $\mathbf{q}$ as well as (ϵ, η). We leave this step to the reader.

The kernel-based posterior (23) can be compared with alternative uses of Entropy Pooling to overlay a prior with partial information. For instance, Meucci (2010) obtains the posterior by imposing that the expected value of the conditioning variables be the same as the current value, i.e. $\mathrm{E}_{\mathbf{q}}\{\mathbf{X}\} = \mathbf{x}_T$. This approach is reasonable in a univariate context. However, when the number of conditioning variables $\mathbf{X}$ is larger than one, due to the curse of dimensionality, we can obtain the undesirable result that the posterior probabilities are highly concentrated in a few extreme scenarios. This does not happen if we condition through a pseudo-Gaussian kernel as in (22)-(23)

In Figure 2 we plot the Entropy Pooling posterior probabilities (23) in our example. We can appreciate the hybrid nature of these probabilities, which share similarities with both the prior (21) and the conditioning kernel (22).

	Time decay	Kernel	Entropy Pooling
St. dev.	1.31	1.30	1.28
Skew.	-2.55	-2.55	-2.52
VaR 99%	-5.35	-5.53	-5.31
CVaR 99%	-6.71	-6.70	-6.40
Eff. Num.	2,098	1,646	1,518

(24)

Using the Entropy Pooling posterior probabilities (23) we can perform all sorts of risk computations, as in Meucci (2012a). In (24) we present a few significant statistics for our historical portfolio (20), and we compare such statistics with those stemming from the standard exponential decay (21) and the standard conditioning kernel (22). For more details, we refer the reader to the code available at http://symmys.com/node/353.

On the last row of (24) we also report the effective number of scenarios, a practical measure of the predictive power of the above choices of probabilities, discussed in detail in Meucci (2012b).

References

Black, F., and R. Litterman, 1990, Asset allocation: combining investor views with market equilibrium, *Goldman Sachs Fixed Income Research*.

Meucci, A., 2008, Fully Flexible Views: Theory and practice, *Risk* 21, 97–102 Article and code available at http://symmys.com/node/158.

——— , 2010, Historical scenarios with Fully Flexible Probabilities, *GARP Risk Professional* December, 40–43 Article and code available at http://symmys.com/node/150.

——— , 2011, The Prayer: Ten-step checklist for advanced risk and portfolio management, *GARP Risk Professional* April/June, 54–60/55–59 Available at http://symmys.com/node/63.

——— , 2012a, Computing all risk numbers and risk decompositions with Fully Flexible Probabilities, *Working Paper* Article and code available at http://symmys.com.

——— , 2012b, Effective number of scenarios with Fully Flexible Probabilities, *Working Paper* Article and code available at http://symmys.com/node/362.

.

A Appendix

In this appendix we present technical results that can be skipped at first reading.

A.1 Time conditioning versus state conditioning

Crisp state conditioning (4) includes the rolling window approach (2) as a special case. To see this, it suffices to add to $\mathbf{X}$ a "random" variable X_0 with deterministic scenarios $x_{0,1} = t_1, x_{0,2} = t_2, \ldots$, and to define the domain $\mathcal{X}$ in (4) as $\mathcal{X} \equiv \{[T - \tau, T] \times \mathbb{R} \times \cdots \times \mathbb{R}\}$.

The time decayed exponential smoothing (3) is a special case of kernel smoothing (7). To see this, it suffices again to add to $\mathbf{X}$ a "random" variable X_0 with deterministic scenarios $x_{0,1} = t_1, x_{0,2} = t_2, \ldots$ that represent time, and to define a kernel as follows

$$k_{\epsilon,\boldsymbol{\mu}}(\mathbf{x}) \equiv e^{-\frac{|x_0-\mu_0|}{\epsilon}}, \tag{25}$$

where

$$\epsilon \equiv \tau / \ln 2, \quad \mu_0 \equiv T. \tag{26}$$

A.2 Entropy Pooling and fuzzy membership

First, let us define

$$\mathbf{v} \equiv (v(\mathbf{x}_1), \ldots, v(\mathbf{x}_J))'. \tag{27}$$

Now, let us express the view (12) reads

$$\mathcal{V} : \mathbf{p}'\mathbf{v} \geq \underline{v}. \tag{28}$$

The value $\underline{v}$ in (12) is a threshold which determines the intensity of the view. Two notable values for the intensity $\underline{v}$ are the following

$$\begin{aligned} \underline{v}_0 &\equiv \mathrm{E}_{\mathbf{p}_0}\{v(\mathbf{X})\} = \mathbf{p}_0'\mathbf{v} & (29)\\ \overline{v} &\equiv \max_{\mathbf{p}} \mathrm{E}_{\mathbf{p}}\{v(\mathbf{X})\} = \max_j v_j. & (30) \end{aligned}$$

When $\underline{v} \leq \underline{v}_0$ the view (14) is not binding; when $\underline{v} \in (\underline{v}_0, \overline{v}]$ the view (14) is binding; when $\underline{v} > \overline{v}$ the view (14) does not admit a solution.

As discussed in Meucci (2008), the solution of the Entropy Pooling optimization (14) can be computed explicitly and reads

$$\mathbf{p} \propto \mathbf{p}_0 e^{\mathbf{v}\lambda(\underline{v})}, \tag{31}$$

where $\lambda(\underline{v})$ is a suitable Lagrange multiplier, and where the product and the exponential of vectors are meant entry-by-entry. The Lagrange multiplier $\lambda(v)$ is fully determined by the view intensity v in (14). When $v \leq \underline{v}_0$ the view (14) is not binding and thus $\lambda(\underline{v}) = 0$; when $\underline{v} \in (\underline{v}_0, \overline{v}]$ the view (14) is binding and $\lambda(\underline{v})$ is an increasing function of $\underline{v}$.

Assume that the prior is uninformative as in (15) and the views are expressed on a log-membership as in (16). Then $\mathbf{v} = \ln \mathbf{m}$, where $\mathbf{m} \equiv (m_{\mathcal{X}}(\mathbf{x}^{(1)}), \ldots, m_{\mathcal{X}}(\mathbf{x}^{(J)}))'$ and thus

$$\mathbf{p} \propto \mathbf{m}^{\lambda(\underline{v})}. \tag{32}$$

Hence, when the view intensity $\underline{v}$ in (28) is v^* such that $\lambda(v^*) = 1$, the Entropy Pooling posterior probabilities (32) are the same as the conditional fuzzy membership probabilities (8). The intensity v^* follows from computing the expectation on the left-hand-side of the inequality (28), which is always binding

$$\begin{aligned} v^* &= \mathbf{p}' \ln(\mathbf{m}) = \mathrm{E}_{\mathbf{u}}\{\ln m_X(\mathbf{X})\} + \ln J + \mathbf{p}' \ln(\mathbf{p}) && (33)\\ &\geq \mathrm{E}_{\mathbf{u}}\{\ln m_X(\mathbf{X})\} && (34) \end{aligned}$$

A.3 Examples of fuzzy membership as Entropy Pooling

To better understand how fuzzy membership is associated with Entropy Pooling, let us consider specific cases. In the case of crisp conditioning (4), the membership function is the indicator $m_{\mathcal{X}} = 1_{\mathcal{X}}$. Therefore the view function (16) reads

$$v(\mathbf{x}) \equiv \begin{cases} 0 & \text{if } \mathbf{x} \in \mathcal{X} \\ -\infty & \text{otherwise.} \end{cases} \tag{35}$$

The view (32) is always binding, because the binding threshold $\underline{v}_0$ defined in (29) is $-\infty$. Thus the Lagrange multiplier is non-null for any view intensity $\underline{v}$. As a result, the Entropy Pooling posterior probabilities (18) read

$$p^{(j)} \propto \begin{cases} 1 & \text{if } \mathbf{x}_j \in \mathcal{X} \\ 0 & \text{otherwise.} \end{cases} \tag{36}$$

As expected, the Entropy Pooling posterior probabilities (36) are the crisp conditioning probabilities (4).

Now, let us turn to kernel smoothing (7), which, as highlighted in (10), is a special case of fuzzy membership $m_{\boldsymbol{\mu}} = k_{\epsilon,\boldsymbol{\mu}}$. In particular, let us focus on the Gaussian kernel (6). Then the view function (16) is proportional to the square Mahalanobis distance

$$v(\mathbf{x}) \equiv -\tfrac{1}{2\epsilon^2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}). \tag{37}$$

Using the general identity $(e^x)^\alpha = e^{x\alpha}$, we obtain that the Entropy Pooling posterior probabilities (18) read

$$p^{(j)} \propto e^{-\frac{1}{2\underline{\epsilon}^2} (\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}, \tag{38}$$

where $\underline{\epsilon} \equiv \epsilon / \sqrt{2\lambda(\underline{v})}$. These are the Gaussian kernel probabilities (7). Notice that the intensity $\underline{v}$ in the view (32) determines the width of the Gaussian kernel: as $\underline{v} \leq 0$ approaches 0, the view (32) becomes more binding and thus the Lagrange multiplier increases, causing the kernel radius $\underline{\epsilon}$ to shrink; as $\underline{v}$ decreases, the view (32) becomes more lax and thus the kernel radius $\underline{\epsilon}$ increases.

A similar phenomenon occurs in the case of exponential smoothing (3). In this case the kernel (25) induces as in (16) the view function

$$v(x_0, x_1 \ldots, x_N) \equiv -\tfrac{\ln 2}{\tau} |x_0 - T| . \tag{39}$$

The Entropy Pooling posterior probabilities (18) become

$$p^{(j)} \propto e^{-\frac{\ln 2}{\underline{\tau}} |t_j - T|}, \tag{40}$$

where $\underline{\tau} \equiv \tau/\lambda(\underline{v})$. These are the exponential smoothing probabilities (3). The intensity $\underline{v}$ in the view (32) determines the decay rate of the exponential smoothing: as $\underline{v} \leq 0$ approaches 0, the view (32) becomes more binding and thus the Lagrange multiplier increases, causing the half-life $\underline{\tau}$ to shrink; as $\underline{v}$ decreases, the opposite effect occurs.