Expert consulting and information combining: a sequential model

Paola Monari, Patrizia Agati, and Luisa Stracqualursi

Dipartimento di Scienze Statistiche
Università di Bologna
via delle Belle Arti, 41
I-40126 Bologna, Italy
(e-mail: paola.monari@unibo.it)
(e-mail: patrizia.agati@unibo.it)
(e-mail: stracqualursi@stat.unibo.it)

Abstract. In many research fields, where valuable information about a random phenomenon may come from different, possibly heterogeneous sources of knowledge ("experts"), the combining of the available information is a powerful uncertainty-reducing process. As efficiency reasons often suggest to perform a *sequential* procedure, in this paper some informativeness-founded selecting and stopping rules are proposed; their performance is discussed in a case-study.
Keywords: sequential consulting, Kullback-Leibler divergence, curvature.

1 Introduction

In many research fields, particularly in decision making and risk analysis, valuable information about a random phenomenon may come from different, possibly heterogeneous, sources of knowledge: information systems (such as, for example, sensor fusion systems), theoretical or empirical models, privileged witnesses. In a single, conventional word: 'experts'. So, the combining of the available information — especially once they were modelled in form of probability distributions — become a powerful uncertainty-reducing process: for example, to assess the entity of an environmental risk or the probability of a space probe malfunctioning, or forecast hurricane track, or classify biological samples, such as fossils. The output of the process — a final probability distribution on the investigated random variable — can be viewed as representing a synthesis of the current state of knowledge regarding the uncertainty of interest: a 'sufficient' synthesis, which must not involve loss of any relevant information.

Numerous algorithms for *simultaneous* combining have been proposed in literature (for a critical review, [Genest and Zidek, 1986] and [Cooke, 1991]). It's not so about *sequential* algorithms. And it is a fact that the investigator often prefers to consult the experts in successive stages rather than simultaneously. So, s/he avoids wasting time and money by consulting a too large sample of experts: at each stage, depending on the amount of information reached, s/he can choose whether to stop or to continue the process and, depending on the answers obtained from the experts already contacted, s/he can select the 'best' expert to be consulted on the subsequent stage.

The aim of this work is to propose some selecting and stopping rules which can be suitable to be used in a sequential consulting process. The substance of such rules is almost independent of the procedure chosen for combining information from the experts; not so their mathematical form. The reference, in the present work, is the Bayesian aggregation model suggested by Morris (1977), reviewed in a recursive form.

The paper is organized as follows. Section 2, in writing Morris' aggregation algorithm in a recursive form, gives the notation for the successive sections. In Section 3, some stopping and selecting criteria are suggested. Their performance is discussed in a real data based case-study which, together with some concluding remarks, is presented in Section 4.

2 A recursive algorithm for the sequential knowledge updating

In a context of uncertainty about the value of a random quantity $\theta \in \Theta \subset \Re$, let's denote with $h_0(\theta)$ the prior probability distribution which reflects the initial state of information. With the aim to acquire knowledge (so reducing the uncertainty) about θ, an investigator A performs a sequential consulting of (at most n) experts Q_j: at each stage k ($k = 1, 2, \ldots, K$; $K \leq n$), the selected expert $Q^*_{j;k}$ (or, more briefly, Q_k) answers by giving his/her/its own density $g_k(\theta)$. Treating each expert's density as result of an experiment, the investigator can revise the initial distribution $h_0(\theta)$ via Bayes' theorem.

Assuming that [Morris, 1977]:

***a*)** each $g_k(\cdot)$ is parameterized with a location parameter m_k and a shape parameter v_k;

***b*)** for each k, the probability which A assigns to the event $v^{(k)} = \bigcap_{i=1}^{k} v_i$ — that is, the event "the shape parameter values the experts will give are $[v_1, ..., v_i, ..., v_k]' = \mathbf{v}$" — does not depend on θ: in symbols, $\ell\left(v^{(k)}|\theta\right) = \ell\left(v^{(k)}\right)$;

Morris shows that the posterior density can be written as[1],

$$h\left(\theta|m^{(k)}, v^{(k)}\right) = \frac{\ell\left(m^{(k)}|v^{(k)}, \theta\right) \cdot h_0(\theta)}{\int_\Theta \ell\left(m^{(k)}|v^{(k)}, \theta\right) \cdot h_0(\theta)\, d\theta} \tag{1}$$

where:

[1] It can be shown that these assumptions can be relaxed without changing substantially the results [Morris, 1977].

– $\ell\left(m^{(k)}|v^{(k)},\theta\right)$, denoted in the following by $\ell_k(\theta)$ for notational convenience, indicates the conditioned likelihood function of θ for the data $m^{(k)} = \bigcap_{i=1}^{k} m_i$, given $v^{(k)}$: it represents — for θ varying — A's probabilities that the location parameter values the experts will provide are $\mathbf{m} = [m_i]'_{i=1,\ldots,k}$;
– the posterior $h\left(\theta|m^{(k)},v^{(k)}\right)$ or, more briefly, $h_k(\theta)$, represents the synthesis distribution at stage k.

If the following assumption holds too:

***c*)** for each k, the conditional probability which A assigns to the event "$g_k(\cdot)$ shape parameter will be v_k", given $m^{(k-1)}, v^{(k-1)}$ and θ, does not depend on θ — that is, $\ell\left(v_k|m^{(k-1)},v^{(k-1)},\theta\right) = \ell\left(v_k|m^{(k-1)},v^{(k-1)}\right)$—;

then Morris' (simultaneous) aggregation algorithm (1) can be written in a recursive form as,

$$h_k(\theta) = \frac{\ell\left(m_k|v_k,m^{(k-1)},v^{(k-1)},\theta\right)\cdot h_{k-1}(\theta)}{\int_{\Theta}\ell\left(m_k|v_k,m^{(k-1)},v^{(k-1)},\theta\right)\cdot h_{k-1}(\theta)\,d\theta} \tag{2}$$

where $\ell\left(m_k|v_k,m^{(k-1)},v^{(k-1)},\theta\right)$ is the conditioned likelihood function of θ for the only observation m_k, given v_k and also the location and shape values provided by the $k-1$ previously consulted experts.

As regards the arduous assessment of the function $\ell(\cdot)$ in (2), the relation $\ell\left(m_k|v_k,m^{(k-1)},v^{(k-1)},\theta\right) = \ell_k(\theta)/\ell_{k-1}(\theta)$ allows to use Morris' (simultaneous) result,

$$\ell_k(\theta) \propto C_k(\theta)\cdot\prod_{i=1}^{k} g_i(\theta) \tag{3}$$

where the *calibration function* $C_k(\theta)$ encapsulates the state of knowledge about each expert's performance and the degree of dependence among the k experts. Briefly [Morris, 1977], let τ_i denote the i-th *performance indicator*, defined as Q_i's cumulative function $G_i(\cdot|m_i,v_i)$ evaluated at the true value of θ: $C_k(\theta)$ expresses the admissibility degrees which the investigator assigns to each possible θ value looked at as the realization of the k-dimensional quantile vector $\tau = [\tau_i]'_{i=1,\ldots,k}$. Technically, $C_k(\cdot)$ is nothing but a subjectively assessed density $\phi_k(\cdot)$ of τ, conditioned on $\mathbf{v}$ and θ, looked at as a function of θ (for fixed $\mathbf{m}$): in symbols, the relation between the so-called *performance function* $\phi_k(\cdot)$ and the calibration function $C_k(\theta)$ is,

$$\phi_k(\tau|\mathbf{v},\theta) = \phi_k\left[\mathbf{G}(\theta|\mathbf{m},\mathbf{v})\,|\mathbf{v},\theta\right] = C_k(\theta) \tag{4}$$

where $\mathbf{G}(\theta|\mathbf{m},\mathbf{v})$ — briefly, $\mathbf{G}(\theta)$ — denotes the vector $[G_i(\theta|m_i,v_i)]'_{i=1,\ldots,k}$.

Whenever only some pieces of information about the experts are available — an 'information block' which is not adequate to construct an empirically founded probability distribution of their performance indicators — the fiducial argument [Fisher, 1956] can be used for inductively modelling the calibration function, enabling it to be specified with a relatively small number of assessments [Monari and Agati, 2001]. With the following notation:

- $\tilde{\mathbf{G}}(\theta) = \left[\tilde{G}_i(\theta)\right]'_{i=1,\dots,k}$, with $\tilde{G}_i(\theta) = \ln\left[G_i(\theta) / (1 - G_i(\theta))\right]$;
- $\tilde{\mathbf{t}} = \left[\tilde{t}_i\right]'_{i=1,\dots,k}$, with $\tilde{t}_i = \ln\left[t_i / (1 - t_i)\right]$;
- c as normalization constant;

the resulting fiducial calibration function can be written as,

$$C_k(\theta) = C_k(\theta; \mathbf{t}, \mathbf{S}) =$$

$$= c \cdot \prod_{i=1}^{k} \{G_i(\theta) \cdot [1 - G_i(\theta)]\}^{-1} \cdot \exp\left\{-\frac{1}{2}\left[\tilde{\mathbf{G}}(\theta) - \tilde{\mathbf{t}}\right]' \mathbf{S}^{-1}\left[\tilde{\mathbf{G}}(\theta) - \tilde{\mathbf{t}}\right]\right\} \quad (5)$$

It's worth noting that function (5) is univocally defined by the following two quantities:

- A's assessment $\mathbf{t} = [t_i]'_{i=1,\dots,k}$ of the performance indicator τ;
- the subjective variance-covariance matrix $\mathbf{S}$, reflecting A's information about the variability and the reciprocal dependence of the experts' performance indicators.

3 Selecting and stopping rules

The purpose of expert consulting is reducing the uncertainty about the unknown quantity θ. So, in designing and performing the sequential process, it is reasonable to found the selecting and stopping rules on some criterion of informativeness. In particular, though no single number can convey the amount of information encapsulated in a density function, a synthetic measure of the (*expected*) additional informative value of a not-yet-consulted expert $Q_{j;k}$ is indispensable for selecting the one to be consulted at stage k, especially when the investigator's calibration assessments, together with the shape parameters provided by the experts, lead to not-coinciding preference orderings. And, analogously, as likelihood functions and posterior densities can display a wide variety of form, a synthetic measure of the reached knowledge degree about θ is needed for picking out the 'optimal' stage k^* at which data acquiring can be stopped.

Let's suppose the investigator A is performing the process of revising beliefs in light of new data according to the algorithm described in Section 2. The prior $h_0(\theta)$ has already been specified; each of n contacted experts Q_j has revealed the variance v_j — assumed as uninformative about θ: see *b)* in Section 2) — of his/her/its own density $g_j(\theta)$, and A has already consulted $k-1$ of them, so obtaining the locations of $k-1$ expert densities: A is

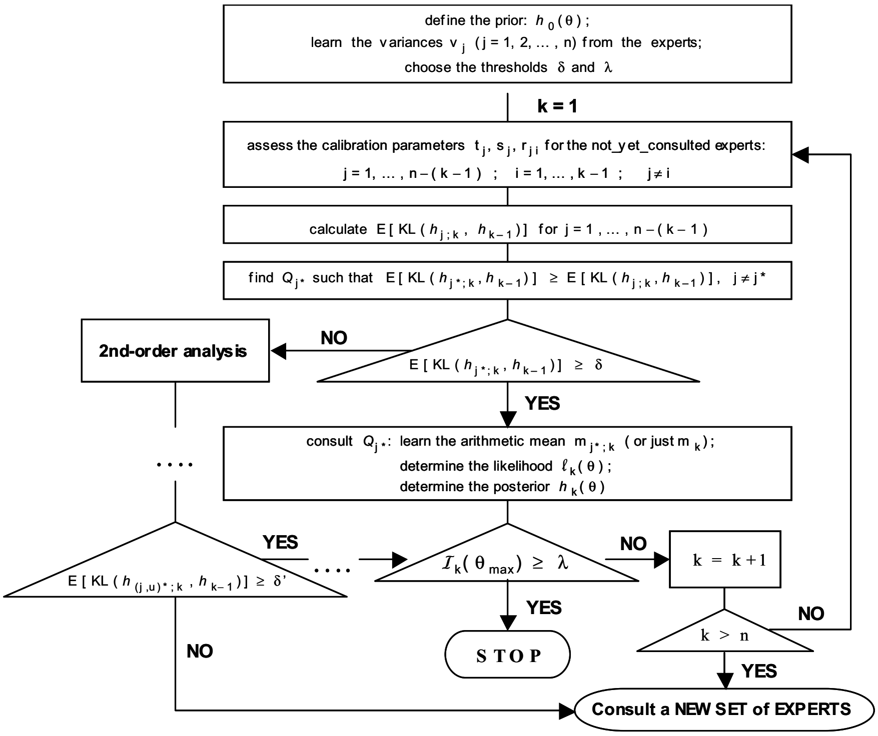

Fig. 1. Flow-chart of the sequential procedure.

now at stage k of the process (figure 1), and must select one among the not-yet-consulted experts $Q_{j;k}$ $(j = 1, 2, \ldots, n-k+1)$.

For each $Q_{j;k}$, the investigator A assesses — conditionally on v_j, on the basis of the information at his disposal (including all the expert locations m_i revealed up to stage $k-1$) — the parameters of the k-stage calibration function $C_{j;k}(\theta)$: that is, t_j, s_{jj} and the covariances s_{ji} (or the linear correlations r_{ji}) between $Q_{j;k}$ and each already-consulted expert Q_i, $i = 1, 2, \ldots, k-1$. At this point of the procedure, no $Q_{j;k}$ has revealed the location value m_j of his own $g_j(\theta)$: the several 'answers' m_j which each can virtually give are not all equally informative, so the (informative) value of each expert at the k-stage — to be measured with regard to A's current knowledge[2] of θ reflected in the posterior density $h_{k-1}(\theta)$ of the previous stage — is an *expected* value,

[2] In fact, all the other elements being equal, the more A is uncertain about θ, the more an answer m_j is worthy.

calculated by averaging a selected measure of relevant information about θ in $Q_{j;k}$'s answer over the space M_j of the virtually possible m_j values.

By reasoning in a *knowledge* context — which is an *inductive* context, where an expert opinion is more relevant the more it is able to modify the posterior distribution on the unknown quantity — a suitable measure of $Q_{j;k}$'s informative value can be the *expected Kullback-Leibler divergence* of the density $h_{j;k}(\theta)$ with respect to the previous stage posterior $h_{k-1}(\theta)$,

$$\mathrm{E}\left[KL\left(h_{j;k}, h_{k-1}\right)\right] := \int_{M_j} f\left(m_{j;k} | v_{j;k}, m^{(k-1)}, v^{(k-1)}\right) \cdot KL\left(h_{j;k}, h_{k-1}\right) dm_j \tag{6}$$

where the KL-divergence [Kullback, 1959],

$$KL\left(h_{j;k}, h_{k-1}\right) := \int_{\Theta} h_{j;k}(\theta) \cdot \ln\left[h_{j;k}(\theta) / h_{k-1}(\theta)\right] d\theta \tag{7}$$

measures indirectly the information provided by an answer $m_{j;k}$ in terms of the changes it yields on the density $h_{k-1}(\theta)$. The conditional density $f(\cdot)$ in (6) is equal to the denominator of (2) read as a function of $m_{j;k}$ and normalized; when assumptions *a)*, *b)* and *c)* hold, it can be determined as

$$f\left(m_{j;k} | v_{j;k}, m^{(k-1)}, v^{(k-1)}\right) = f\left(m^{(j;k)} |, v^{(j;k)}\right) / f\left(m^{(k-1)} | v^{(k-1)}\right) \tag{8}$$

where the density $f\left(m^{(j;k)} |, v^{(j;k)}\right)$ — and analogously $f\left(m^{(k-1)} | v^{(k-1)}\right)$ — is equal, up to the normalization term, to the denominator $\int_{\Theta} \ell\left(m^{(k)} | v^{(k)}, \theta\right) \cdot h_0(\theta)\, d\theta$ of (1), read as a function of $m^{(k)}$.

The expert $Q^*_{j;k}$ presenting the greatest expected KL-divergence is, at stage k, the most informative: but is he/she/it an expert worth consulting? The answer is yes, if the information he provides is, on average, *enough* different from what A already knows about θ, *i.e.* if the expected divergence of $h_{j^*;k}(\theta)$ with respect to $h_{k-1}(\theta)$ is not less than a predetermined value δ ($0 \leq \delta < \infty$). About the choose of the threshold δ, a very useful tool is the scheme proposed by McCulloch for deciding whether a KL-divergence value is a large or a small one [McCulloch, 1989].

So the *selecting rule* can be expressed as follows. *Consult the expert* $Q^*_{j;k}$ *such that*

$$\mathrm{E}\left[KL\left(h_{j^*;k}, h_{k-1}\right)\right] \geq \mathrm{E}\left[KL\left(h_{j;k}, h_{k-1}\right)\right] \qquad j \neq j^* \tag{9}$$

on condition that

$$\mathrm{E}\left[KL\left(h_{j^*;k}, h_{k-1}\right)\right] \geq \delta \tag{10}$$

If $Q^*_{j;k}$ *does not satisfy* (10), *then proceed to a 2nd order analysis: that is, consult the pair* $\left(Q_{j;k}, Q_{u;k}\right)^*$ *presenting the greatest expected KL-divergence, provided that it is* $\mathrm{E}\left[KL\left(h_{(j,u)^*;k}, h_{k-1}\right)\right] \geq \delta$*; otherwise contact a new set of experts and perform a new process by using the posterior* $h_{k-1}(\theta)$ *as a new prior* $h_0^{'}(\theta)$.

The expert $Q^*_{j;k}$ satisfying (10) becomes just Q_k, the "k-stage expert". By consulting him, A learns the location m_k of the density $g_k(\cdot)$: now, the k-stage calibration function $C_k(\theta)$ is univocally defined, and consequently, the likelihood function $\ell_k(\theta)$ and the posterior density $h_k(\theta)$ too.

In theory, the investigator should stop the process only when the knowledge about θ, reflected in the posterior density, is 'inertially stable': *i.e.*, only when additional experts, even if jointly considered, are not able to modify appreciably the synthesis distribution, on the contrary they contribute to its inertness. But too many experts could be needed for realizing such a stopping condition. It can be weakened by requiring just the knowledge about θ deriving from expert answers to be enough for A's purposes. A measure encapsulating the strength of the experimental data in determining a preference ordering among 'infinitesimally close' values of θ is Fisher's notion of information. The value of the *observed information* $I(\cdot)$ at the maximum of the log-likelihood function,

$$I_k(\theta_{\max}) := -\partial^2/\partial\theta^2 \ln \ell_k(\theta_{\max}) \tag{11}$$

is a second-order estimate of the spherical curvature of the function at its maximum: within a second-order approximation, it corresponds to the KL-divergence between two distributions that belong to the same parametric family and differ infinitesimally over the parameter space.

So, the *stopping rule* may be defined as follows. *Stop the consulting at stage k^* at which a pre-selected observed curvature λ of the log-likelihood valued at $\theta := \theta_{\max}$ has been reached,*

$$I_{k^*}(\theta_{\max}) \geq \lambda \tag{12}$$

For deciding whether a curvature value $I(\theta_{\max}) = w$ is a large or a small one, a device could be the following. Let's think of a binomial experiment where a number $x = n/2$ of successes is observed in n trials and find x such that $I(\hat{p}_{ML} = 0.5) = w$, where $\hat{p}_{ML} = 0.5$ is the maximum likelihood estimate of the binomial parameter p. Table 1 shows a range of x values with the corresponding w curvature values. The simple relation $x = w/8$ holds: so, for example, if $w = 120$, the width of the curve $\ln \ell_k(\theta)$ near $\theta := \theta_{\max}$ is the same as the curve $\ln \ell(p)$ at $\hat{p}_{ML} = 0.5$ when $x = 15$ and $n = 30$.

x	1	2	5	10	15	20	25	30	40	50
w	8	16	40	80	120	160	200	240	320	400

Table 1. Large or small curvature values? Relation between x and w values.

4 Case-study and concluding remarks

The behavior of the algorithms proposed in the previous section — and implemented [Agati and Stracqualursi, 2001] in MATHEMATICA — has been investigated in simulation and experimental studies. In this section, the results from medical data are synthetically presented to exemplify how the selecting and stopping rules work. Particularly, data in table 2 regard a sequential consulting process of $n = 4$ orthopaedists, performed by an Italian research laboratory about the long-term failure log-odds θ of a new hip prosthesis. A fifth surgeon has assessed the calibration parameters, without modifying them in proceeding from a stage to the successive one. He has also (subjectively) chosen the following thresholds:

- $\delta = 0.02$: by reading this value in McCulloch's scale, at stage k the most informative expert $Q^*_{j;k}$ is consulted only if the expected KL-divergence of $h_{j^*;k}(\theta)$ with respect to $h_{k-1}(\theta)$ is not less than the KL-divergence of a Bernoulli distribution $B(p)$ with $p = 0{,}5$ from a Bernoulli distribution with $p = 0.65$; or, in other words, only if stopping the process at stage $k-1$ instead of proceeding to stage k involves, on average, an information loss larger than that one yielded by using a $B(0{,}65)$ instead of a $B(0{,}5)$;
- $\lambda = 120$: by using the scale proposed in Section 3, the consulting process is stopped at stage k^* at which the observed curvature of the log-likelihood function $\ln \ell(\theta)$ valued at $\theta := \theta_{\max}$ is the same as the function $\ln \ell(p)$ at $\hat{p}_{ML} = 0.5$ when, in a binomial experiment, $n = 30$ and $x = 15$.

Q_j	v_j	t_j	s_{jj}	r_{j1}	r_{j2}	r_{j3}	r_{j4}
Q_1	0.150	0.45	1.20	1			
Q_2	0.145	0.65	1.50	+0.20	1		
Q_3	0.120	0.75	1.70	−0.05	+0.50	1	
Q_4	0.110	0.45	1.10	+0.10	+0.10	+0.10	1

Table 2. Input data for the sequential consulting of four orthopaedists about long-term failure log-odds of a new hip prosthesis.

In this study, the conditions *a)*, *b)* and *c)* mentioned in Section 2 can be held to be satisfied. In fact: *a)* it rests on empirical evidence — and the experts confirm it — that the failure log-odds θ can be supposed as Gaussian; *b)* it is reasonable to think the probability the fifth orthopaedist assigns to the event "the experts will give the variances $[v_1, ..., v_4]' = \mathbf{v}$" is the same for all θ values: so the surgeons' stated variances alone give no information able to change the investigator's beliefs about θ; *c)* it is reasonable as well to assume the conditional probability the investigator assigns to the event "the expert $Q_{j;k}$ will give the variance v_k", given the shape and location values provided by the $k-1$ previously consulted experts, is the same for all θ values. So the combining algorithm outlined in Section 2 has been applied, as well as the selecting and stopping rules suggested in Section 3.

Q_j	*Stage* $k=1$ $\mathrm{E}\left[KL\left(h_{j;1},h_0\right)\right]$	*Stage* $k=2$ $\mathrm{E}\left[KL\left(h_{j;2},h_1\right)\right]$	*Stage* $k=3$ $\mathrm{E}\left[KL\left(h_{j;3},h_2\right)\right]$
Q_1	1.41487	**1.92935**	—
Q_2	1.35582	1.52293	1.42842
Q_3	1.42427	1.72981	**1.93624**
Q_4	**1.60348**	—	—
	↓	↓	↓
$Q^*_{j;k}$	Q_4	Q_1	Q_3
m_k	-1.208	-1.992	-2.752
$I_k\left(\theta_{\max}\right)$	18.713	53.492	138.984 ($>120=\lambda$)

Table 3. Output of the proposed sequential procedure in the consulting of four orthopaedists about long-term failure log-odds of a new hip prosthesis.

Table 3 summarizes the results of the sequential process, while figure 2 shows the posterior distributions $h_k\left(\theta\right)$ at each stage.

For $k=1$, the selecting rule proposed in Section 2 chooses the expert Q_4: really he offers the smallest variance ($v_4=0.110$), and also the investigator's uncertainty about his performance indicator is assessed to be the smallest ($s_{44}=1.10$). Q_4's answer ($m_{4;1}=-1,208$) leads to a curvature value $I_1\left(\theta_{\max}\right)=18.713<120=\lambda$: so the process goes on.

At stage 2, the selecting rule shows its usefulness: in fact, the v_j, t_j and s_{jj} values [3] don't lead to a unique preference ordering. The most informative expert Q_1 is selected by the algorithm[4] and $m_{1;2}$ is observed. The curvature value is $I_2\left(\theta_{\max}\right)=53.492<120=\lambda$: the consulting proceeds.

At stage 3, the preference for Q_3 instead of Q_2 is also (but not only) motivated by the correlations with Q_1: a negative correlation ($r_{31}=-0.05$) is more informative than a weak positive one ($r_{21}=0.20$). The observed $m_{3,3}$ leads to $I_3\left(\theta_{\max}\right)=138.984>120=\lambda$. The process is stopped: the expert Q_2 is left out of the consulting and stage-3 posterior $h_3\left(\theta\right)$ — whose location and shape values are, respectively, -1.873 (the median, here coinciding with the arithmetic mean and the mode) and 0.084 (the standard deviation) — can be regarded as the synthesis expression of the expert knowledge about the long-term failure log-odds θ of the new hip prostheses.

[3] The correlations between Q_4 and the other experts are all equals: so they don't come into play.

[4] It's worth noting that the value $m_{4;1}$ observed at stage 1 has modified, at stage 2, the previous-stage preference ordering: for this reason, the selecting at each stage one only expert is to be preferred to selecting a set of experts (simultaneously.

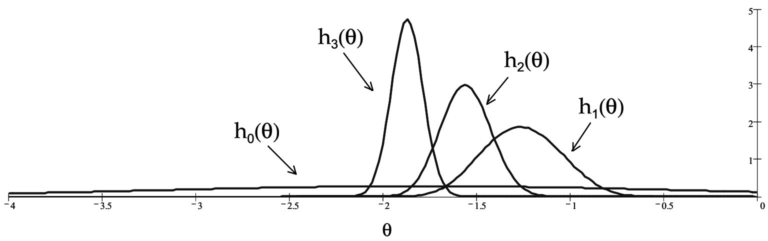

Fig. 2. Posterior distributions at stages 0 (*i.e.*, the prior), 1, 2 and 3 of the sequential procedure.

By looking at this selecting and stopping output, the behavior of the informativeness criteria appears to be coherent with the intuition, so giving an empirical support about the soundness of the proposed selecting and stopping algorithms in performing an efficient sequential consulting process. At present, our research efforts are focused on the combining of information from hurricane track prediction models: so, with the aim of assessing the calibration parameters for each model (an 'expert', in our framework), simulations were performed on a training-set of North Atlantic historical hurricane data regarding the location of specific storms at prefixed time intervals. Successively, separately for each time interval, each track prediction model with its own parameters entered in the informativeness-founded sequential algorithm and, on the basis of the selecting and stopping output, a Bayesian combined track prediction model for each prefixed time interval was proposed: the research — still in progress — promises interesting results.

References

[Agati and Stracqualursi, 2001]P. Agati and L. Stracqualursi. *Algoritmi computazionali per l'aggregazione di opinioni esperte.* Clueb, Bologna, 2001.

[Cooke, 1991]R.M. Cooke. *Experts in Uncertainty. Opinion and subjective probability in science.* Oxford University Press, Oxford, 1991.

[Fisher, 1956]R.A. Fisher. *Statistical Methods and Scientific Inference.* Oliver & Boyd, Edinburgh, 1956.

[Genest and Zidek, 1986]C. Genest and J. V. Zidek. Combining probability distributions: a critique and an annotated bibliography. *Statistical Science*, pages 114–148, 1986.

[Kullback, 1959]S. Kullback. *Information Theory and Statistics.* Wiley, New York, 1959.

[McCulloch, 1989]R.W. McCulloch. Local model influence. In *Journal of the American Statistical Association*, pages 473–478, 1989.

[Monari and Agati, 2001]P. Monari and P. Agati. Fiducial inference in combining expert judgements. *Journal of the Italian Statistical Society*, pages 81–97, 2001.

[Morris, 1977]P. A. Morris. Combining expert judgments: a bayesian approach. *Management Science*, pages 679–693, 1977.