A Protocol for the Elicitation of Imprecise Probabilities

Alane Alves Silva
Departamento de Engenharia de Produção
Universidade Federal de Pernambuco,
Recife–PE, Brazil
alaneaas@yahoo.com.br

Fernando Menezes Campello de Souza
Departamento de Eletrônica e Sistemas
Universidade Federal de Pernambuco
Recife–PE, Brazil
fmcs@hotlink.com.br

Abstract

A protocol for the elicitation of imprecise probabilities based on linear programming is applied to the case of two continuous variables. Two medical experts were elicited. The resulting convex set of probability distributions was compared with the results obtained by the application of an imprecise Dirichlet model to a database. An indicator is introduced to assess the inferential skill of the medical experts.

Keywords.

Elicitation, Imprecise Probabilities, Inferential Skill, Convex set, Imprecise Dirichlet Models.

1 Introduction

The method for the treatment of the a priori accumulated knowledge that one has about a state of the world comes from the so-called Bayesian inference. The a priori probability used in this setup, also called subjective or epistemic probability, represents the degree of belief that the individual has in the occurrence of an event which is represented in terms of the random variable θ. One of the main drawbacks of this approach, as pointed out in [6], is that a precise prior distribution on the random variable θ is required. Several approaches were proposed to try to overcome this disadvantage. These include upper and lower probability [4], upper and lower previsions [3] and the consideration of families of prior distributions. A whole new field of imprecise probabilities deals with this kind of problem.

The protocol presented in [6, 2, 5], and used here, is part of a method that provides a systematic procedure for the elicitation of a prior distribution of some random variable θ from an expert. The available evidence, in practical real world settings, is typically of a mixed, partial, nature, and the expert's knowledge has always a certain degree of vagueness. The new protocol is useful in these conditions. It avoids a source of confusion inherent in the commonly used betting schemes, where judgment would be elicited through preferences, involving thus two different psychological mechanisms. There is no need for a total precision on behalf of the specialist, and there are no errors to be treated statistically.

The method uses pairwise comparative probabilistic assertions involving events defined by the random variable θ. The general method does not require θ to be a random variable; it could be just a category.

The protocol for the elicitation of prior knowledge presented in [6, 5, 2] is used here for the case of two continuous variables. The result of this application was compared with the results obtained by the application of an imprecise Dirichlet model presented in [3].

2 Linear Programming Imprecise Probabilities Model

The Linear Programming Imprecise Probabilities Model (LPIPM) is an elicitation method that was presented in [5, 6] and further developed in [2]. It will be summarized here.

The method is a systematic procedure to elicit an expert prior distribution of some unknown real-valued continuous random variable θ. The expert announces minimum and maximum plausible values for θ. In his mind he evaluates that the probability that the true value of θ lies outside these two limits, θ_{min} and θ_{max}, is zero. It is assumed that θ is distributed in the interval $[\theta_{min}, \theta_{max}]$ according to a probability density π. This interval is partitioned into $2n$ subintervals of equal Lebesgue measure, $[\theta_{j-1}, \theta_j)$, $j = 1, 2, \ldots, 2n$. The value of n depends upon the intended precision. If one accepts that 5% is a good precision for an expert one then can adopt $n = 10$ (20 slices of 5%). It is the usual quantization procedure of a continuous variable. It is convenient to represent the interval $[\theta_{j-1}, \theta_j)$ by θ_j, for short. Define also $\pi_j = Pr\{\theta \in [\theta_{j-1}, \theta_j) = \pi(\theta_j)$, the probability that θ belongs to the jth subinterval. The probability that θ belongs to the interval $[\theta_j, \theta_{j+k})$ is $\sum_{i=0}^{k} \pi_{j+i}$ for $j + k \leq 2n$. It is clear that $\sum_{j=1}^{2n} \pi_j = 1$.

The questions posed to the specialist are of the following type: Which one is greater than the other,

$$Pr\big\{\theta \in [\theta_j, \theta_{j+k})\big\} \quad \text{or} \quad Pr\big\{\theta \in [\theta_l, \theta_{l+m})\big\} \text{ ?}$$

The superposition of intervals may cause confusion, so the two intervals in each question should not overlap. Use of this assumption was made in order to elaborate the indicators for the construction of the elicitation questionnaire [6]. The problem is treated then as if we were dealing with finitely many θ's. More details can be found in [5] and [6]. The general method was extended for the nonaleatory case in [1] and [2]. In this case one would have a set of finitely many categories, $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$, where θ_j, $j = 1, 2 \ldots, n$ is neither a number, nor represent an interval, as in the continuous case explained previously. In other words, θ_j would not be a random variable.

The input from a specialist consists then in answering a certain number of pairwise comparisons of the probabilities of events, and also to express the relative odds of these events. Two linear programming problems are posed:

$$\underset{\pi_j}{\text{Max}} \text{ (Min)} \sum_{j=1}^{2n} c_j \pi_j \tag{1}$$

subject to:

$$a_{jk} \sum_{i=0}^{k} \pi_{j+i} - a_{lm} \sum_{i=0}^{m} \pi_{l+i} \leq b_s \tag{2}$$

$$\pi_j \geq 0, \ j = 1, 2, \ldots, 2n \tag{3}$$

$$\sum_{j=1}^{2n} \pi_j = 1 \tag{4}$$

Let q be the number of questions posed to the specialist, so one will have q constraints similar to equation 2, where for each one, one choose $\leq b_s$ or $\geq b_s$, depending on the specialist's answer, $a_{jk} > 0$, $a_{lm} > 0$, and $j + k < l$ to avoid overlapping between two intervals.

The values used in this work were $a_{jk} = a_{lm} = 1$ and $b_s = 0$. Therefore one wants to know just which interval is more probable.

There are several possibilities for the choice of the c_j's. In the sequel they were chosen in such a way as to get a distribution with the minimum expected value for θ (the maximization problem) and a distribution with a maximum expected value for θ (the minimization problem). Hence

$$c_j = 2n - j + 1 \tag{5}$$

Note that to maximize

$$\sum_{j=1}^{2n} (2n - j + 1)\pi_j \tag{6}$$

is the same as to maximize

$$\frac{\theta_1 + \theta_{2n}}{2} - \sum_{j=1}^{2n} \theta_j \pi_j \tag{7}$$

as far as the choice of π_j, $j = 1, 2, \ldots, 2n$ is concerned. Since the values $\theta_1, \theta_2, \ldots, \theta_{2n}$ are in an arithmetic progression, one has $\theta_{j+1} - \theta_j = a$ for $j = 1, 2, \ldots, 2n - 1$ where $a > 0$. Then

$$\frac{\theta_1 + \theta_{2n}}{2} = \frac{\theta_1 + \theta_1 + (2n-1)\ a}{2} = \theta_1 + \frac{2n-1}{2}\ a$$

So

$$\frac{\theta_1 + \theta_{2n}}{2} - \sum_{j=1}^{2n} \theta_j \pi_j = a \left[\frac{2n+1}{2} - \sum_{j=1}^{2n} j \pi_j \right]$$

On the other hand,

$$\sum (2n - j + 1)\pi_j = 2n + 1 - \sum_{j=1}^{2n} j \pi_j$$

One sees then that to maximize

$$\sum (2n - j + 1)\pi_j$$

is the same as to minimize

$$\sum_{j=1}^{2n} \theta_j \pi_j$$

Clearly different $c_j's$ will produce different results.

The set of constraints guarantees that $\{\pi_j\}_{i=1}^{2n}$ is in fact a probability distribution, either for the maximum or for the minimum LP problem. If the $c_j's$ are the ones defined by expression 5, all the convex combinations of the two solutions (one corresponding to the distribution with minimum expected value for θ, i.e., the maximization problem, and the other to the distribution with the maximum expected value for θ, i.e., the minimization problem) will be consistent with the expert's answers. This convex set of probability distribution can then be used in inference or decision procedures. Of course this family is, in principle, smaller then the set of all possible probability distributions compatible with the specialist answers. To specify this last set is not a simple task. The “size” of this family could be estimated by the volume of the feasible set of the optimization problems.

One could use the same feasible set, but a different objective function. Another objective function could be the entropy of the distribution, defined by

$$H = -\sum_{j=1}^{2n} \pi_j \log \pi_j$$

the optimization problems would now be of the non-linear programming type, but the reasoning is the same.

Typically the two solution will be different, and one will obtain two distribution functions as depicted in Figure 1.

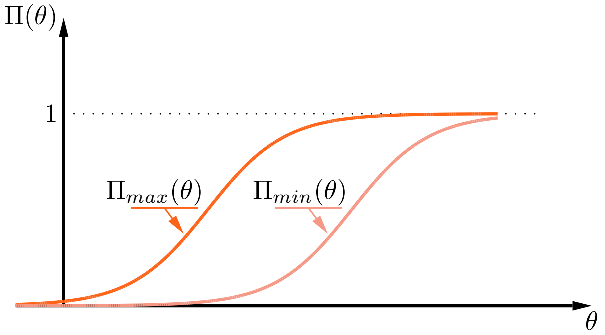

Figure 1: Example of convex set of probability distributions.

Let Π_{max} and Π_{min} be distribution functions on θ. They are constructed from the solutions of the respective linear programming problems. Notice that an area will form between the two curves. The ratio of this area to the total area $[\theta_{max} - \theta_{min}] \times 1$ of the rectangle was defined in [6, 2] as the **vagueness**, V, of the specialist. That is,

$$V = \frac{1}{2n} \sum_{j=1}^{2n} \left|\Pi_{max}(\theta_j) - \Pi_{min}(\theta_j)\right| \tag{8}$$

The vagueness will be minimum if the specialist answers all the questions consistently, i.e., the feasible set of the LP problems is nonempty. If the specialist does not answer any question, he will be totally consistent, and its vagueness will be maximum ($V = 1$). So, consistency, here, does not mean sharpness of the elicited family of probability distributions.

The **precision** is defined by

$$P = 1 - V \tag{9}$$

The more vague the specialist is, the less will be its precision.

3 The Imprecise Dirichlet Model

Following the reasoning presented in [3], the available data will be used to estimate the upper and lower posterior probability in an Imprecise Dirichlet Model (IDM). Walley [3] defined IDM as the set of all Dirichlet distributions (s, t) for a given parameter s.

The Dirichlet (s, t) prior distribution for θ, where $t = (t_1, t_2, \ldots, t_k)$, has probability density function.

$$\tau(\pi) \propto \prod \pi_j^{st_j - 1}$$

where $s > 0$, $t_j > 0$ for $j = 1, \ldots, k$, and $\sum_{j=1}^{k} t_j = 1$. In this parametrization t_j is the mean value of π_j according to the Dirichlet (s, t) and s determines the influence of the prior distribution on posterior probability. The IDM is the set of all Dirichlet (s, t) distributions The parameter s is considered to be a hyper-parameter typically chosen between 1 and 2 (see [3]).

Each Dirichlet prior has a posterior distribution obtained applying Bayes’ theorem. These prior distributions form the set of posterior Dirichlet distributions. In this set are defined the posterior upper and lower probabilities:

$$\overline{P}(\pi_j|x) = \frac{n_j + s}{N + s} \tag{10}$$

$$\underline{P}(\pi_j|x) = \frac{n_j}{N + s} \tag{11}$$

were n_j denote the number of observations falling in the j^{th} cell θ_j in N trials and π_j will denote the probability of the cell θ_j

The degree of imprecision in the posterior upper and lower probabilities is defined in [3] can be measured by

$$\overline{P}(\pi_j|n) - \underline{P}(\pi_j|n) = \frac{s}{N + s} \tag{12}$$

For more details, see [8] and [7].

4 Application

Two medical doctors were submitted to an elicitation protocol concerning the prior knowledge about the systolic blood pressure (SBP) and diastolic blood pressure (DBP) of an individual. The first doctor is a young cardiologist (expert 1), an M.Sc. student in Medicine. The second one is an experienced cardiologist (expert 2), specialized in the field of hypertension. Both were given the same evidence concerning an individual: male, aged 46 years, Body Mass Index of 27 kg/m^2 nonsmoker, policeman, educated up to the highschool level, has no health complaints, and was randomly chosen amongst all the individuals in his city with similar characteristics.

The questions in the elicitation questionnaire concerned the SBP and DBP of this individual. The elicitation protocol was of the same type as one presented in [6]. For the SBP a minimum of 90 mm Hg and a maximum of 190 mm Hg were established, and for the DBP the minimum and maximum values established were, respectively, 40 mm Hg and 100 mm Hg.

Differently from the elicitation presented in [6], where only the SBP was considered, now the unknown state of nature is a vector in a subset of the Euclidean space $\mathbb{R}^2$. In order to apply the linear programming based elicitation method, the SBP×DBP rectangle was divided into 20 cells labeled θ_j, $j = 1, 2, \ldots, 20$ as shown in Figure 2. For example $\theta_{10} = [170 - 190, 55 - 70]$.

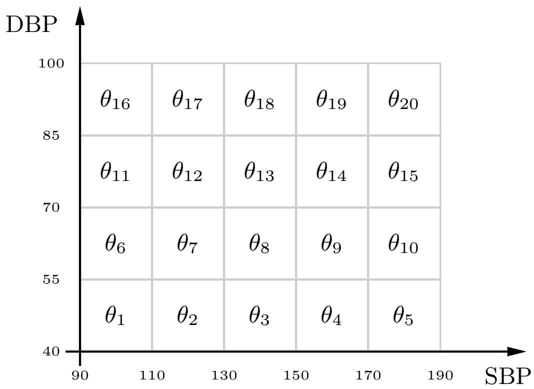

Figure 2: States of Nature

Notice that those θ_j's are not naturally ordered as in the monodimensional case.

For the attribution of the c_j's, the notion of pulse blood pressure (PBP) was used. The pulse pressure is the difference between the SBP and DBP. It is known from physiology, that SBP is always greater than DBP. In each cell θ_j (with the exception of the left half side of θ_{16}), the value of SBP is always greater than the value of DBP. This is a logical necessary condition of the cardiovascular dynamics. By considering the PBP one gets then a one-dimensional random variable that is, thus, naturally orderable. Notice that each θ_j there corresponds an average pulse pressure.

As far as the reduction to one dimension (PBP) is concerned, that is, from the original two (DPB and SBP), it should be mentioned that the medical doctors are aware that there is a correlation between DBP and SBP, and they took this into consideration when answering the questionnaire (this correlation varies with age). Moreover, PBP is an important marker of cardiovascular health. The PBP was discretized by taking the mean values of DBP and SBP ranges. For example, θ_8 correspond to a pulse pressure of $140.0 - 62.5 = 77.5$ mm Hg (see Figure 2).

Table 1: The elicitation questionnaire.

	[SBP, DBP]	1|0	[SBP, DBP]
1	$[90 - 190, 40 - 70]$		$[90 - 190, 70 - 100]$
2	$[90 - 150, 40 - 100]$		$[150 - 190, 40 - 100]$
3	$[90 - 130, 40 - 100]$		$[130 - 190, 40 - 100]$
4	$[90 - 190, 40 - 70]$		$[110 - 190, 70 - 100]$
5	$[90 - 170, 40 - 70]$		$[90 - 190, 70 - 100]$
$\vdots$	$\vdots$		$\vdots$
42	$[110 - 130, 55 - 70]$		$[150 - 170, 70 - 85]$

The cardiologists answered to the questionnaire independently of each other. This experiment was performed and its results were analyzed in [1], where many applications of decision theory in cardiology are presented.

Some questions of the questionnaire presented to the cardiologists are shown in table 1. The first question, for instance, is the following:

> What is more like, that this individual's SBP is between 90 mm Hg and 190 mm Hg and DBP is between 40 mm Hg e 70 mm Hg, or that SBP is between 90 mm Hg and 190 mm Hg and DBP is between 70 mm Hg and 100 mm Hg?

4.1 Results

The evidence provided to the cardiologists were purposedly scarce. In [6], indicators for the construction of the elicitation questionnaire were presented. Besides guaranteeing symmetry and avoiding bias, a questionnaire constructed based on the mentioned indicators, guarantees the expert's gradual and smoothly progressive perception of the parameter (state of nature; the random variable θ), distributed along with the questionnaire. The idea in the construction presented in [6] was not to confound the specialist with reasoning retrocessions. So one should take the first consistent answers, instead of a through revision. The first expert (the young cardiologist) was consistent only in the 17 first questions, and its vagueness was 18.75%. The experienced cardiologist was able to be consistent in 31 of the 42 questions of the elicitation questionnaire. Its vagueness was 33.33%. The results of the LPIPM are shown in Figure 3 and Figure 4,

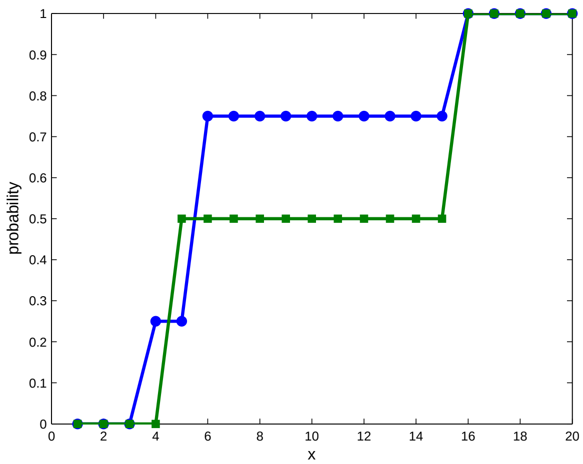

Figure 3: Elicitation results of expert 1.

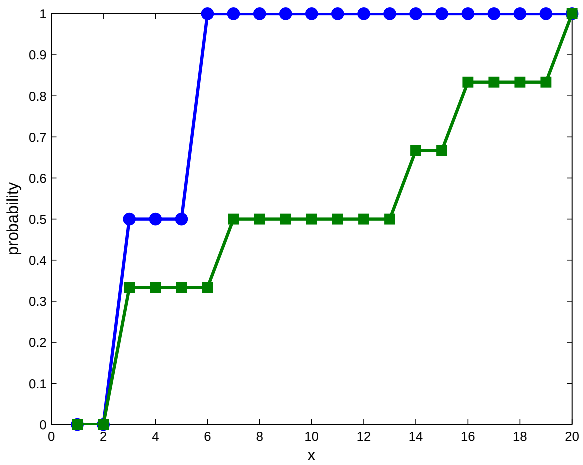

Figure 4: Elicitation results of expert 2.

4.2 The Inferential Skill

It could appear at first that the experienced cardiologist should have a smaller vagueness as compared to the young cardiologist, and that the numerical results obtained contradicted the common sense. But this is not so. The scarce evidence provided to the specialists could not be a basis for obtaining a "thin" family of probability distributions. The young cardiologist thought it could do so and its vagueness was indeed smaller than that of the experienced cardiologist. But in doing this he was able to answer consistently only to 17 questions, while the other was consistent in 32 questions, although with a larger vagueness.

Three parameters should be taken into account in order to ascertain the **inferential skill** of the expert:

1. The difference of the averages of the maximum and minimum distributions ($\Delta\mu$), normalized by $\Delta\mu_{max}$, which is the maximum value of $\Delta\mu$;

2. The fraction of consistently answered questions (R);

3. The vagueness (V).

An overall indicator, S, of inferential skill would be then:

$$S \triangleq \frac{\Delta\mu}{\Delta\mu_{max}} + R - V \tag{13}$$

If the expert is a good one, the scarcity of the available evidence (input to him), should be reflected in a compatible vagueness (V) and difference between the means of the two distributions ($\frac{\Delta\mu}{\Delta\mu_{max}}$). Given that, the larger the value of R, the better the expert. If the expert is not that good, R will tend to be smaller, indicating a sharpness unwarranted by the scarce evidence. If the expert has a fixed value in mind, for example, the vagueness around this value will tend to be smaller, contributing to a decrease in the overall vagueness. Also, in this case, ($\frac{\Delta\mu}{\Delta\mu_{max}}$) will tend to be smaller. This is an indication that the expert is not so good. The experienced cardiologist had $S = 0.83$ and the less experienced one had $S = 0.36$. The fact that the two distributions for the young cardiologist touched each other, may be indicative that he fixed his attention on a certain θ_j, creating a psychological mechanism of anchorage.

5 Comparisons with a Database

The two elicitation results were compared with data collected in a Brazilian sample of 2129 subjects. The two specialists were not aware of the existence of the data. By selecting a subset of the sample containing individuals with roughly the same characteristics as the ones of the case presented to the two specialists, one obtains the probability distribution function shown in Figure 5, alongwith the mean distribution for each expert for comparison.

It is important to point out that a specific piece of input was given to the two experts, namely, that the individual was a policeman. In the database this was not specified.

6 The Imprecise Dirichlet Model

Table 2 shows the results of the IDM method, for two values of the hyperparameter s, namely, $s = 1$ and $s = 2$. The database used is the same.

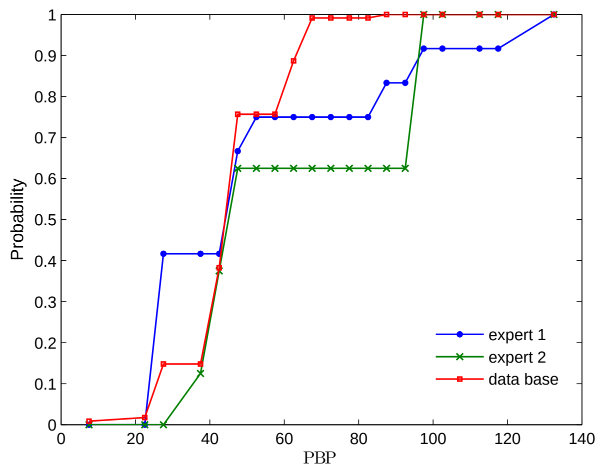

Figure 5: Probabilities obtained from a database.

According to the interpretations forwarded in reference [3], $s = 1$ corresponds to "frequentists" and "objective Bayesians", and $s = 2$, to an "overly cautious Bayesian". As the value of s increases, the expert becomes more cautious.

7 The LPIPM and the IDM

Table 3 summarizes the results of the application of the two models. The values of the probabilities $\underline{P}$ and $\overline{P}$ were computed from expression 11 and 10, respectively, for the IDM. For the LPIPM, those values were computed from π_{max} and π_{min}, respectively.

The experienced cardiologist, expert 2, has the largest degree of imprecision, according to the terminology introduced by Walley (1996) [3]. For $s = 1$ (a "frequentist") and $s = 2$ (an "overly cautious Bayesian") the lower probabilities, using the IDM (only the data was considered) were roughly the same, but the upper probability for the "overly cautious Bayesian" was larger than the one of the frequentist. Both cardiologists and the "overly cautious Bayesian" (IDM) had essentially the same upper probabilities. The two smallest degrees of imprecision, close to each other, were the one of expert 1 (the young cardiologist), and the IDM with $s = 1$ ("frequentist").

It was not told to the experts if the individual were in good health or not. It was told only that the individual had no health complaints.

The individuals in the database were randomly selected in public places, and had no explicit heart problems. The database was used just to check if there would be no large discrepancies between the probabilities obtained from the experts and the one obtained from the row data. The values shown in Figures 6, 7 e 8 do not point to large discrepancies.

Table 2: Results of the IDM.

	$s=1$		$s=2$	
	$\overline{P}$	$\underline{P}$	$\overline{P}$	$\underline{P}$
θ_1	0.008621	0	0.017094	0
θ_2	0.008621	0	0.017094	0
θ_3	0.008621	0	0.017094	0
θ_4	0.008621	0	0.017094	0
θ_5	0.008621	0	0.017094	0
θ_6	0.008621	0	0.017094	0
θ_7	0.008621	0	0.017094	0
θ_8	0.008621	0	0.017094	0
θ_9	0.008621	0	0.017094	0
θ_{10}	0.008621	0	0.017094	0
θ_{11}	0.017246	0.008625	0.025645	0.008551
θ_{12}	0.241397	0.232776	0.24788	0.230786
θ_{13}	0.137897	0.129276	0.145265	0.128171
θ_{14}	0.008621	0	0.017094	0
θ_{15}	0.008621	0	0.017094	0
θ_{16}	0.017246	0.008625	0.025645	0.008551
θ_{17}	0.137897	0.129276	0.145265	0.128171
θ_{18}	0.379397	0.370776	0.384701	0.367607
θ_{19}	0.112022	0.103401	0.119611	0.102517
θ_{20}	0.017246	0.008625	0.025645	0.008551

Table 3: Results of the two models.

	$\underline{P}$	$\overline{P}$	Degree of Imprecision
$s = 1$	47.13	59.20	12.07
$s = 2$	46.73	70.06	23.33
expert 1	57.50	70.00	12.50
expert 2	37.50	70.83	33.33

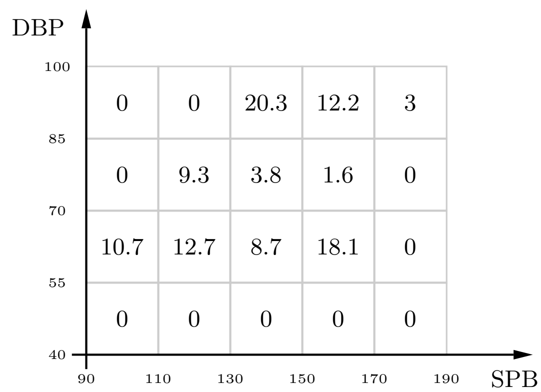

Figure 6: Expert 1

It seems reasonable to admit that the knowledge of the experienced cardiologist is larger than the one

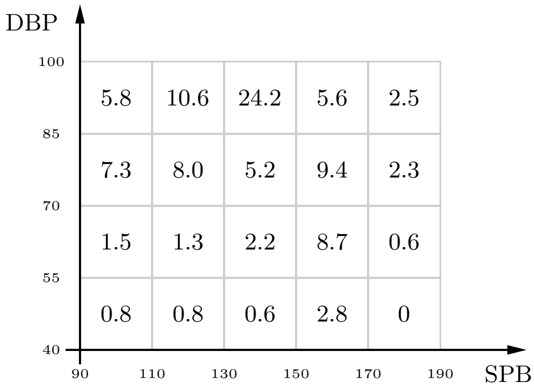

Figure 7: Expert 2

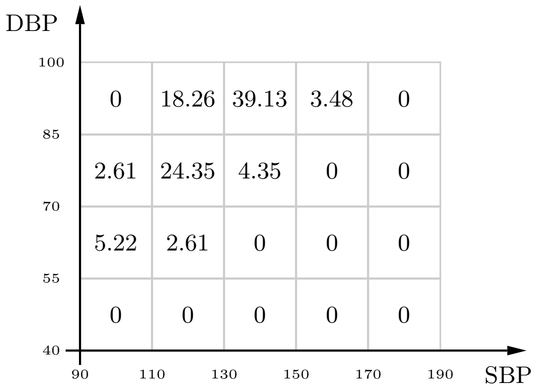

Figure 8: Data Base

that could be obtained from a sample of 115 individuals. Thus he can assign probability masses to a broader range of blood pressures. It should be remembered that it was mentioned explicitly in the written specifications of the individual whose blood pressure profile was to be assessed that he is a policeman. The database does not explicit that. This is not an additional information for the expert. It was all the time in the available evidence presented to him, and automatically included in the LPIPM. The professions of the 115 individuals in the database were unknown; they were randomly selected from the population, and their professions were not registered in the database.

8 Final Remarks

The elicitation protocol presented in [6] and [2] was useful in the assessement of the prior knowledge of two cardiologists concerning the blood pressure profile of an individual. An indicator of the inferential skill of the experts, based on the constructs of the elicitation method, was introduced, which can discriminate the two assessed medical doctors. The experienced cardiologist, in a comparison with the results of an IDM, could be considered as a cautious Bayesian.

References

[1] A.A. Silva. *A Teoria da Decisão em Cardiologia.* Dissertação de Mestrado em Engenharia de Produção, Universidade Federal de Pernambuco, Recife, Brasil, 2002. 2002.

[2] F.M. Campello de Souza, 2002. *Decisões Racionais em Situações de Incerteza.* Editora Universitária da Universidade Federal de Pernambuco, Recife, PE, Brasil, 2001, ISBN-85-7315-178-1.

[3] P. Walley, 1996. *Inferences from Multinomial Data: Learning about a Bag of Marbles* Journal of the Statistical Society, Ser. **B**., 58, 3-57.

[4] A.P. Dempster, 1966. *New methods for reasoning towards posterior distributions based on sample data.* Ann. Math. Statist., **37**, 355-374.

[5] G.C. Nadler Lins. *Contribuições a um protocolo de edução do conhecimento a priori.* Tese de Mestrado em Engenharia de Produção, Universidade Federal de Pernambuco 2000.

[6] G.C. Nadler Lins, F.M. Campello de Souza. A protocol for the elicitation of prior distributions. *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, ISIPTA'01, Cornell University, Ithaca, New York, 265-272, 2001.

[7] F.P. A. Coolen. 1997 *An imprecise Dirichlet model for Bayesian analysis of failure data including rightcensored observations.* Reliability Engineering and System Safety, **56**, 61-68.

[8] J.M. Bernard 2002. *Implicative analysis for multivariate binary data using an imprecise Dirichlet model* Journal of Statistical Planning and inference, **105**, 83-103.

[9] J.M. Bernard 2001. Non-parametric inference about an unknown mean using the Imprecise Dirichlet Model. *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, ISIPTA'01, Cornell University, Ithaca, New York, 2001.