Relevance of Qualitative Constraints in Diagnostic Processes

A. CAPOTORTI
Università di Perugia, Italy

Abstract

This paper reviews recent results obtained in the medical diagnosis field by adding to a coherent inference process qualitative constraints. Such further considerations turn out to be significant whenever a basic lower-upper conditional probability assessment induces extension bounds too vague to take any decision. Three general types of qualitative judgements are proposed and fully described. They do not constitute a "panacea" to solve any problematic situation, but their application can considerably improve inferences results in specific fields, as two practical applications show.

Keywords

coherent inference, conditional exchangeability, qualitative constraints, diagnosis procedures

1 Introduction

In many practical applications, and in particular in the medical field, there is the problem that the information at hand is not so fully detailed and sound to adopt sophisticated statistical tools. This happens especially whenever information is based on data collected from different sources or by heterogenous samples. In these cases a *genuinely* probabilistic reasoning can anyway help to reach considerable results about relevant statements. Of course, with such approach, answers differ from usual *uniquely determined* statistical results, having, in general, interval-based conclusions. Unluckily, there is the widespread bad habit of avoiding not unique answers by forcing in the model *artificial* assumptions, such as independence, and this can bring to misleading inferences. On the other hand, it is true that, especially if information is very limited , results could be so vague that it is impossible to make any reasonable decision. Hence, it is reasonable to search for further properties that can help us to reach sharper conclusions. This can be obtained by a deeper analysis of the problem and also by further qualitative judgements. Of particular importance are conditional exchangeability assumptions, which are more general and reasonable than those of independence, comparisons between conditional probabilities, which are apt to capture expert convictions not numerically expressible, and restrictions on the admissible class of agreeing conditional measures, which are induced by indirect considerations on some statement not considered at the beginning.

In this paper we will explicitly show how such further considerations can be formalized and operationally adopted in general inference processes. Moreover we will have an idea of their relevance by applying them on two medical diagnostic procedures: a *median decision* process for the asbestosis diagnosis based on X-ray film's readings and a reliability judgement of a *GIST (gastrointestinal stromal tumor)* diagnosis based on istochemical results.

2 Coherent Inferences with Limited Information

As already sketched out in the Introduction, whenever a problem does not allow a description by usual statistical models, a simple probabilistic approach can anyway be adopted to extrapolate which are the bounds induced by the available information. This is possible by embedding the problem at hand in a coherent setting, i.e. representing the relevant entities through conditional events endowed with numerical values or bounds and looking for some class of conditional measures agreeing with them. Once a class has been detected, it can be used to make inference on relevant quantities (usually called "indexes").

With such approach, we have, on one side, the peculiarity of a direct introduction of conditional probability assessments, hence they are not derived as sub-products of joints and marginal evaluations, on the other hand we are aware of working with imprecise tools (interval assessments, classes of distributions, bounds for conclusions, etc.). The wide range of subjects covered in the previous ISIPTAs symposia ([8, 9]) testifies of the meaningfulness and soundness of the last aspect, while appropriateness and usefulness, both from a theoretical and a practical point of view, of the first are contained in the work started in [6] and recently fully described in Coletti & Scozzafava's book [7].

2.1 Preliminaries

Let us now introduce a proper formalization to operate with the framework depicted before. For the sake of simplicity we will use conditional and unconditional events, but everything can be easily generalized to (finite) random variables, conditional or not (see for example what it has been done about conditional previsions in [4]). The initial information, usually a knowledge and/or rule base, is represented trough a conditional lower-upper probability assessment. Hence we will have a generic list of n conditional events $\mathcal{F} = (S_1|C_1, \dots, S_n|C_n)$, where each $S_i|C_i$ represents some macro-situation S_i (i.e. some combination of events) considered in some particular hypothetical circumstances C_i (usually the C_is represent different scenarios).

Incompleteness of the information can have two origins: the S_is do not describe all possible combinations and the different circumstances C_is can *overlap* or do not cover all possibilities. For this, it is crucial to know which are the relationships of incompatibility, implication, coincidence or whatever, among the events $\mathcal{U}_{\mathcal{F}} = \{C_1, \ldots, C_n, S_1, \ldots, S_n\}$ because they represent constraints that any model must fulfill. Moreover they limit which are the possible *atoms*. The atoms are elementary events obtained by full combinations of affirmed or negated events in $\mathcal{U}_{\mathcal{F}}$ [1].

We will generally denote by $\mathcal{L}_C$ the set of such logical constraints and we will refer only to atoms A_r, with $r = 1, \ldots, a$, spanned by $\mathcal{U}_{\mathcal{F}}$ and inside the disjunction $\bigvee_{i=1}^{n} C_i$. In the sequel we will also need to use the characteristic vectors of the events, i.e. vectors whose components are 1 or 0 depending if the corresponding atom implies or not the event, and we will denote them with the same letter of the event but in boldface lower-cases. Hence, for example, $\mathbf{s_i}$ and $\mathbf{c_i}$ will denote the characteristic vectors of S_i and C_i, respectively, while their juxtaposition $\mathbf{s_i c_i}$ will represent the characteristic vector of the conjunction S_iC_i (for the sake of simplicity we will omit the usual conjunction operator $\wedge$). To complete this notational parenthesis, in the following we will use the logical operator $\neg$ to denote negations.

The last component of an assessment is represented by numerical bounds $\mathbf{p} = ([lb_1, ub_1], \ldots, [lb_n, ub_n])$, each closed interval $[lb_i, ub_i]$ associated to the corresponding conditional event $S_i|C_i$, and usually estimated by expert believes, literature reports or by collected data.

Note that some $S_i|C_i$ could be actually unconditional (i.e. the situation S_i is considered independently from any specific circumstance) and in such case C_i will coincide with the sure event Ω. Moreover some of the numerical bounds $[lb_i, ub_i]$ could degenerate in a single value p_i, representing a precise assessment.

2.2 Coherence

If we don't want, or we cannot, adopt for the domain $(\mathcal{F}, \mathcal{L}_C, \mathbf{p})$ a unique probabilistic model, it is just possible to search for a class $\mathbb{P}_{\mathcal{F}}$ of conditional probability distributions, such that $\mathbf{p}$ coincides with the restriction to $\mathcal{F}$ of the closed envelop of $\mathbb{P}_{\mathcal{F}}$. This can be operationally checked by the satisfiability of a *class of sequences* of linear systems. *Sequences* of linear systems are necessary to allow the possibility that conditioning events C_is have induced probability not bounded away from 0. Hence there could be the need of classifying the conditional events in different *zero layers*. On the other hand, a *class* of linear systems is required because, to be sure $\mathbf{p}$ agrees with a *closed envelope*, each bound lb_j or ub_j must be cyclically forced to be strictly fulfilled as an equality (for a deeper exposition

[1] In some discipline atoms are called *possible worlds*.

of both aspects refer again to [7], in particular to chapt. 12 and 15).

Such linear systems will anyway have a common structure like

$$\left\{ \begin{array}{rcl} \mathbf{E}\cdot\mathbf{x} & = & 0 \\ \mathbf{L}\cdot\mathbf{x} & \geq & 0 \\ \mathbf{U}\cdot\mathbf{x} & \leq & 0 \\ \mathbf{x}\geq 0 & , & \mathbf{x}\neq 0 \end{array} \right. \tag{1}$$

where $\cdot$ represents the row-column matrix product, $\mathbf{x}$ is a column vector of unknowns, with each component x_r associated to an atom A_r, $r = 1,\ldots,a$, while $\mathbf{E}$, $\mathbf{L}$ and $\mathbf{U}$ are matrices that reflect the numerical constraints induced by $\mathbf{p}$. Hence in $\mathbf{E}$ a generic row is of the form

$$(\mathbf{s_i c_i} - p_i \mathbf{c_i})$$

for each $S_i|C_i$ with a precise assessment p_i and cyclically for one $S_k|C_k$ with an imprecise assessment and forcing p_k to be equal to lb_k or to ub_k. On the other hand, in $\mathbf{L}$ and $\mathbf{U}$ there are, respectively, rows like

$$(\mathbf{s_j c_j} - lb_j \mathbf{c_j})$$

and

$$(\mathbf{s_j c_j} - ub_j \mathbf{c_j})$$

for each $S_j|C_j$ with probability bounds lb_j and ub_j different from the chosen p_k.

Through the set of solutions $\mathbf{x}$, it is possible to represent the searched class $\mathbb{P}_{\mathcal{F}}$.

2.3 Extension

Once coherence of the assessment $(\mathcal{F},\mathcal{L}_C,\mathbf{p})$ has been assured, and in practical application this turns out to be a compulsory step whenever information comes from different sources, it is possible to perform inference on any conditional event $H|E$ judged important to reach conclusions on the problem. Usually H represents some hypothesis to test on the basis of some fact E.

In this context, inference reduces to compute the coherent extension of $\mathbf{p}$ to $H|E$, obtainable as the closed envelop $[lb_{H|E}, ub_{H|E}]$ of the values $P(H|E)$ with $P \in \mathbb{P}_{\mathcal{F}}$. Operationally we need to perform sequences of optimizations of the form

$$\begin{array}{l} \text{minimize/maximize } \mathbf{he}\cdot\mathbf{x} \\ \text{s.t.} \\ \quad \begin{array}{rcl} \mathbf{E}\cdot\mathbf{x} & = & 0 \\ \mathbf{L}\cdot\mathbf{x} & \geq & 0 \\ \mathbf{U}\cdot\mathbf{x} & \leq & 0 \\ \mathbf{e}\cdot\mathbf{x} & = & 1 \\ \mathbf{x}\geq 0 & & \end{array} \end{array} \tag{2}$$

where the normalization constraint $\mathbf{e} \cdot \mathbf{x} = 1$ permits the optimization problem to be linear instead of fractional.

The main difficulty of such procedure is the usually huge number a of atoms but, thanks to a smart use of null probabilities, in [2, 5] this complexity problem has been tackled and mainly solved for practical applications.

3 Results Improvement by Qualitative Constraints

Extension bounds $[lb_{H|E}, ub_{H|E}]$ are what, from a pure probabilistic point of view, our information implies on $H|E$ but, sometimes, they could result too wide to take any decision. Anyway, it is possible, maintaining a *model free* approach, to shrink the reference conditional probability class $\mathbb{P}_{\mathcal{F}}$ adding qualitative (i.e. not numerically expressed) considerations to the numerical constraints $\mathbf{p}$. Of course there are several possible different kinds of constraints to introduce, but we will focus on few of them, either because they are quite natural or because by them we have reached quite satisfactory results.

3.1 Conditional Exchangeability vs Independence

As already mentioned, a widespread tool for restricting the variability of the conclusions is to adopt some assumption of independence. And it is actually a powerful restriction, but usually it is a too strong assumption, not supported by the problem. It is in fact usually confused with the information that some evaluations are made *independently* (i.e. one given without knowing the others), while it should be used to model situations whose measure of uncertainty cannot be *modified* by simply taking into account some other aspect. Moreover its formalization and use in a context of partial information should be done with the awareness of all its implications, that are deeper then the simple factorization of some joint probabilities (for more details see once more [7], chapt. 17).

In the presence of strong symmetries, like for example assessment on the same statement made *independently* by different experts with similar skills (see for example Lad et al. [11]), it is more suitable to introduce some kind of *exchangeability*. This is opportune whenever it is relevant *how many* instead of *which* events realize, or, in other words, whenever it possible to identify a *sum* as a sufficient statistic (for a detailed explanation refer to [10], sect. 3.9). In particular, whenever the assessment is mainly conditional, the judgement of *conditional exchangeability* could be the more suitable and it is formulated as follows:
if there is a group of k events $E_1, \ldots, E_k$ regarded exchangeable *under a specific scenario* C_j, then any conjunction of the E_is with the same number of affirmed and negated events must be equally evaluated. In other words, for any fixed number $s \in \{0, \ldots, k\}$ there must be a constant c_s such that

$$P(E_{i_1} \ldots E_{i_s} \neg E_{i_{s+1}} \ldots \neg E_{i_k} | C_j) = c_s \tag{3}$$

for any permutation of the indexes $i_1, \dots, i_k$.

Conditions like (3) actually reduce the "degree of freedom" for the unknowns $\mathbf{x}$ respect the constraints (1) of the original assessment, restricting "de facto" the admissible class of conditional measures $\mathbb{P}_{\mathcal{F}}$ and, possibly, shrinking some extension bounds.

Since (3) refers to a fixed conditioning event C_j, restriction of this type are easily reported as linear constraints. In fact, denoting with π_s and π'_s the characteristic vectors of two different permutations of the combination $E_{i_1} \dots E_{i_s} \neg E_{i_{s+1}} \dots \neg E_{i_k}$, extensions with the further conditional exchangeability requirement obtain by adding to (2) pairwise equalities of the form

$$(\pi_s \mathbf{c_j} - \pi'_s \mathbf{c_j}) \cdot \mathbf{x} = 0 \tag{4}$$

for each pair of permutations $\pi_\mathbf{s}$ and $\pi'_\mathbf{s}$ and each $s = 1, \dots, k-1$ (note that extreme cases $s = 0$ and $s = k$ do not actually constitute any constraint).

3.2 Conditional Probabilities Comparison

Sometimes there are conditional events which an expert believes more than some other, but he/she cannot express neither precise nor imprecise probability assessments on them, being only capable to compare them.

This is immediately interpretable as

$$P(S_i|C_i) \geq k^+ P(S_j|C_j) \tag{5}$$

for some constant value k^+.

Anyway, if none of the conditional probabilities present in (5) is uniquely constrained, its direct representation by vectors would be

$$\mathbf{x}^T \cdot [(\mathbf{s_i c_i})^T \cdot \mathbf{c_j} - (k^+ \mathbf{s_j c_j})^T \cdot \mathbf{c_i}] \cdot \mathbf{x} \geq 0 \tag{6}$$

that has the drawback of being quadratic. This increases the difficulties for the computation of the extension bounds. In fact, to deal with quadratically constrained optimization problems there are specific Operational Research's techniques, like interior-point algorithms [13] or duality bound methods [14], but they are not so safe and confirmed like those for linear programming problems.

That is why we propose an approximation of (5) that, even being a weaker constraint, has the advantage of leaving the extension problem in a linear form. The idea is of expressing (5) in a parametric way and introducing further unknowns that can capture the basic structure of the parameterization.

If we focus our attention on one of the two conditional probabilities in (5), let us say $P(S_j|C_j)$, we can take it as an *inference target* and compute its extension bounds $[lb_{S_j|C_j}, ub_{S_j|C_j}]$ as it has been illustrated in Subsection 2.3. We can now introduce new variables y_i, $i = 1, \dots, a$, representing the quantities $P(S_j|C_j)x_i$, so that the inequality (5) can be represented by

$$\mathbf{s_i c_i} \cdot \mathbf{x} - k^{+}\mathbf{c_i} \cdot \mathbf{y} \geq 0; \tag{7}$$

the link by new and old variables by

$$\mathbf{s_j c_j} \cdot \mathbf{x} - \mathbf{c_j} \cdot \mathbf{y} = 0; \tag{8}$$

while the variability bounds for $P(S_j|C_j)$ imply the constraints

$$lb_{S_j|C_j} x_i \leq y_i \leq ub_{S_j|C_j} x_i \qquad \text{for } i = 1, \ldots, a. \tag{9}$$

These constraints are all implied by (5), while the vice versa does not hold in general. Hence, if the minimization/maximization of $\mathbf{he} \cdot \mathbf{x}$ is performed with constraints (2), (7), (8) and (9) we are not guaranteed to have obtained the coherent extension for $P(H|E)$ of $\mathbf{p}$ plus (5), but just an interval containing it. However, once such optimal solutions $\mathbf{x}$ are obtained, they can be substituted in (6) to check if the interval $[lb_{H|E}, ub_{H|E}]$ is coherent. If not, the left-hand-side of (6) will result a negative value that can be adopted as a *measure of violation* of (5).

Of course it is not needed to add sequences of optimizations to cyclically impose equalities in (7) and (9) because they must be fulfilled as they are by each $P \in \mathbb{P}_{\mathcal{F}}$.

Anyway, (7), (8) and (9) increase significantly the space complexity of the optimization procedure. Hence, before to adopt them it would be better to check if the optimal solutions of the original linear program (2) already satisfy (6). If it is the case, it means that the qualitative comparison (5) is redundant because it actually does not restrict the class $\mathbb{P}_{\mathcal{F}}$.

3.3 Selectors Restriction

We introduce now a consideration that will result more technical than the previous ones. It will be less intuitive and also more debatable, hence it should be used more carefully and it will anyway need an *interpretation process* before being presented to a field's expert for its acceptance.

Analyzing the inference procedure for some conditional event $H|E$, it could happen to notice that results are mainly influenced by the possible variability of some other $K|F$. As usual $K|F$ can be conditioned to a proper F or unconditional, i.e. with $F = \Omega$. If $K|F$ does not belong to the initial list of conditional events $\mathcal{F}$, the induced bounds $[lb_{K|F}, ub_{K|F}]$ for its conditional probability could be extremely vague, and usually this is not noted at the beginning because $K|F$ could be of no direct interest.

However, it could be impossible to assess bounds for $P(K|F)$ either because the data on which $\mathbf{p}$ was built are not available anymore or because there is not direct information on $K|F$. Anyway, an *indirect* consideration is possible.

Variability range $[lb_{K|F}, ub_{K|F}]$ results from the union of all the extensions, say $[lb^j_{K|F}, ub^j_{K|F}]$, with $1 \leq j \leq n$, of the *extreme* conditional distributions $\mathcal{P}_j \subset \mathbb{P}_{\mathcal{F}}$. With *extreme distribution* we mean those $P \in \mathbb{P}_{\mathcal{F}}$ that reach at least one the lower or upper bounds (lb_j or ub_j) of the assessment **p**. It could happen that some of the $[lb^j_{K|F}, ub^j_{K|F}]$ is narrow enough to drastically influence $P(H|E)$, showing that not all the admissible distributions play the same role for the inference.

Hence, adopting a more *restrictive* attitude and thanks also to some extra consideration, it is possible to select only some of the admissible $\mathcal{P}_j \subset \mathbb{P}_{\mathcal{F}}$ by choosing more informative lower-upper bounds for $P(K|F)$ (possibly coinciding with the narrower interval $[lb^j_{K|F}, ub^j_{K|F}]$) so that the initial assessment can be updated and a new inference on $H|E$ performed.

4 Two Medical Applications

We will show now how the procedures described before can be applied on practical problems. In particular we will illustrate the results we recently attained for two different medical diagnostic processes. The first problem will show how to apply and the relevance of the conditional exchangeability assumptions and of the conditional probabilities comparisons as depicted in subsections 3.1 and 3.2. On the contrary, with the second one we will show the importance of a preliminary check of coherence whenever information comes from different sources and the influence in the results of selector restrictions, in line with subsections 2.2 and 3.3, respectively.

4.1 Accuracy Rates for an Asbestosis Median Decision Procedure

In [3] we re-examined the procedure of median decision making in the context of radiological determination of asbestosis. Median decision applies whenever there is a pool of experts, usually equivalent in skill, examining the same patients and each single case is finally diagnosed on the basis of the agreement of the majority of judgments.

In particular, in a recent paper [12], Tweedie and Mergersen analyze a previous case-report about incidence of asbestosis among a group of people with a similar history of asbestos exposure. Opinions of three radiologists are based on X-ray films readings, and the authors have rather limited information about the median decision procedure. Anyway, they are able to propose a tricky methodology to retrieve some conclusion about the probability of the diagnosis being correct.

However, the authors' analysis deeply relies on a assumption of independence for the experts' assessments and they adopt it because X-ray films are read *in-*

dependently by the radiologists. But this consideration should pertain to *experts' assessment procedure*, not to *our belief* about information's influence one expert opinion *could* have on an other. Actually, since the experts have similar skills, the response of one of them is already a significant indicator of what we could expect from an other.

Tweedie and Mergersen are aware of the inadequacy of the independence presumption, but they wonder how it could be replaced. The fact is that they "need" to introduce independence to maintain uniqueness of the agreeing conditional probability distribution. On the other hand, the information that the three experts are judged equivalently because of their similar skill cannot be ignored. As we have underlined in Subsection 3.1, assumptions of conditional exchangeabilities could be an appropriate answer to this need.

To make a synthesis (a full description can be obviously found in the cited papers), we can formalize the problem as it follows.

First of all we introduce events that refer to a generic patient with a X-ray film available:

label	*description*
F	asbestosis (*fibrosis*) presence
$D_i \quad , i = 1,2,3$	i-th expert positive asbestosis judgment
D^*	positive median decision diagnosis
S^*	positive median decision with a splitting vote

Since the similarity among radiologists, their *sensitivities* for the films' reading process $P(D_i|F)$, $i = 1,2,3$, are thought to be equal.

On the basis of recorded data on 642 patients and of specific literature references, the following conditional probability assessment **p** on $\mathcal{F} = (D_1|F, D_2|F, D_3|F, D^*, S^*|D^*)$ is considered[2]:

$$P(D_i|F) = .82 \qquad i = 1,2,3$$
$$P(D^*) = .12$$
$$P(S^*|D^*) = .42$$

The first probability $P(D_i|F)$ comes from literature results on sensitivity analyses performed by comparing radiological and histopathological evaluations. The other two $P(D^*)$ and $P(S^*|D^*)$ derives from the only data reported in [12]. In particular, $P(D^*)$ is directly estimated by the ratio $77/642$ of positive median diagnoses, while $P(S^*|D^*)$ is attained indirectly by the three individual 82%, 86% and 90% positive assessments through the formula

$$P(S^*|D^*) = (100-82)\% + (100-86)\% + (100-90)\% = 42\%.$$

To complete the assessment we must explicitly give which are the possible logical relations $\mathcal{L}_C$ among the unconditional events $\mathcal{U}_\mathcal{F} = \{F, D_1, D_2, D_3, D^*, S^*\}$.

[2] In [12] and [3] several assessments with different sensitivity values are examined, here we report only the first one as prototype

By the problem description we can pick out logical dependencies among the median decisions, with or without splitting vote, and individual experts' diagnosis

$$\begin{aligned} S^* &= (D_1 D_2 \neg D_3) \vee (D_1 \neg D_2 D_3) \vee (\neg D_1 D_2 D_3) \\ D^* &= S^* \vee (D_1 D_2 D_3) \end{aligned}$$

It is easy to check that the numeric assessment $\mathbf{p}$ is coherent and that, even being a precise conditional probability assessment, the admissible class $\mathbb{P}_{\mathcal{F}}$ is not a single conditional distribution, as it will appear in the sequel.

We can consider the assessment $(\mathcal{F}, \mathcal{L}_{\mathcal{C}}, \mathbf{p})$ as a partial knowledge base whose main "lack" is the absence of an estimate for the expert's *specificity* $P(\neg D_i|\neg F)$. Anyhow, thanks to the conditional independence assumptions

$$P(D_i|D_j F) = P(D_i|F) \quad \text{and} \quad P(D_i|D_j \neg F) = P(D_i|\neg F) \tag{10}$$

and thanks to some algebraic manipulation involving Bayes' Theorem, Tweedie and Mergersen uniquely determine probability values for the usual accuracy indexes *specificity, positive predictive value, negative predictive value* and estimate the *true positive proportion*. We can compare their results with what we obtained firstly without any assumption, secondly adopting the method of Subsection 3.1 to incorporate the following conditions of conditional exchangeability[3]

$$\begin{aligned} P(D_1 D_2 \neg D_3|F) = \quad & P(D_1 \neg D_2 D_3|F) && = P(\neg D_1 D_2 D_3|F) \\ P(D_1 \neg D_2 \neg D_3|F) = \quad & P(\neg D_1 \neg D_2 D_3|F) && = P(\neg D_1 D_2 \neg D_3|F) \\ \\ P(D_1 D_2 \neg D_3|\neg F) = \quad & P(D_1 \neg D_2 D_3|\neg F) && = P(\neg D_1 D_2 D_3|\neg F) \\ P(D_1 \neg D_2 \neg D_3|\neg F) = \quad & P(\neg D_1 \neg D_2 D_3|\neg F) && = P(\neg D_1 D_2 \neg D_3|\neg F) \end{aligned} \tag{11}$$

and finally using considerations of Subsection 3.2 to consider the following conditional probabilities' comparisons that arise from the formalization of an interview with a further physician[4]:

[3] With respect to the notation of Subs.3.1 we have $k = 3$, $E_i = D_i$ and C_j equal at first to F and after to $\neg F$

[4] These comparisons are the result of the formalization of a long and detailed analysis of the influence of the knowledge of the answers of some expert on the behaviors of the others. It has been performed with a physician extraneous to the rest of the work

$$\frac{P(D_3|D_1D_2F)}{P(\neg D_3|D_1D_2F)} \geq 3/2\frac{P(D_1|F)}{P(\neg D_1|F)}$$ (linear)

$$\frac{P(D_3|\neg D_1\neg D_2F)}{P(\neg D_3|\neg D_1\neg D_2F)} \leq 2/3\frac{P(D_3|F)}{P(\neg D_3|F)}$$ (linear)

$$\frac{P(D_3|\neg D_1\neg D_2\neg F)}{P(\neg D_3|\neg D_1\neg D_2\neg F)} \leq 2/3\frac{P(D_3|\neg F)}{P(\neg D_3|\neg F)}$$ (quadratic)

$$P(D_3|D_1\neg D_2F) \in [.5, .5+(P(D_3|F)-.5)]$$ (linear)

$$P(D_3|D_1\neg D_2\neg F) \in [.5-(P(D_3|F)-.5), .5]$$ (linear)

$$P(D_2|D_1F) \geq P(D_2|F)$$ (linear)

$$P(D_3|D_2D_1F) \geq P(D_3|D_1F)$$ (quadratic)

$$P(D_2|\neg D_1\neg F) \leq P(D_2|\neg F)$$ (quadratic)

$$P(D_3|\neg D_2\neg D_1\neg F) \leq P(D_3|\neg D_1\neg F)$$ (quadratic)

Note that such relations, even being similar in structure (the first three actually reflect odds ratios comparisons), are distinguished, by labels, between those of them that are actually linear constraints since some quantity is uniquely determined and those that are properly quadratic and need the proposed linear approximation.

We cannot go into technical details, but it is important to mention just one computational feature: the number of atoms in this problem is 16, but conditional independence assumptions (10) reduce at two the degrees of freedom for their probabilities, i.e. everything is fully determined once the experts' sensitivity $P(D_i|F)$ and specificity $P(\neg D_i|\neg F)$ could be selected, while with conditional exchangeabilities (11) we have only a reduction at 8 degrees of freedom.

Here we report the different inferences performed on several accuracy indexes, specifying the particular assumptions adopted

index	*description*	extension bounds under			
		cond. idep.	*no ass.*	*cond. exch.*	*qual. comp.*
$P(\neg D_i\|\neg F)$	experts' specificity	.957	[0 , 1]	[.603 , 1]	[.820 , .970]
$P(F\|D^*)$	positive predict. val.	.961	[0 , 1]	[0 , 1]	[0 , .779]
$P(\neg F\|\neg D^*)$	negative predict. val.	.988	[.970 , 1]	[.971 , 1]	[.979 , 1]
$P(F)$	asbestosis incidence	.126	[0 , .130]	[0 , .130]	[0 , .106]
$P(D^*\|F)$	med. dec. sensitivity	.994	[.730 , 1]	[.730 , 1]	[.820 , .878]
$P(\neg D^*\|\neg F)$	med. dec. specificity	.995	[.880 , 1]	[.880 , 1]	[.954 , .970]

Whenever conditional exchangeability cannot help on limiting vague inference bounds, the further qualitative probabilistic comparisons are determinant. In fact, apart from the positive predictive value, all the intervals in the last column are tight enough to judge the procedure. About the only "vague" interval $[0, .779]$, even it does not bound from below the positive predictive value, it gives an interesting upper limitation for such index.

Moreover, note that some interval of the last column do not contain the corresponding values obtained by Tweedie and Mergersen. This because the further constraints go in the opposite direction of independence, bringing some kind of correlation but leaving ”untouched” the conditional exchangeability framework.

Our computations needed to solve several liner programming problems, but what we obtained is really based on reasonable probabilistic statements and not on tricky manipulation that have the only justification of bringing to single values instead of intervals.

4.2 Reliability of GIST Diagnosis Based on Partial Information

Other prototypes of applications of inference with a not fully detailed model are the medical diagnostic procedures where there is not a *golden standard* protocol to follow. This happens when new advances in the understanding of the biology are done or new techniques are discovered. In such situations, different opinions appear in scientific literature and they are based on disparate case studies, each one with its peculiarity and heterogeneity of data.

In particular, in [1] we analysed a diagnostic process for *gastrointestinal stromal tumors* (GISTs) where only recently a new and reliable phenotypic marker (the KIT protein CD117) for these neoplasm has been introduced.

The diagnosis path consist mainly of two stages: at first a histological analysis is done and later an immunohistochemical schema is adopted to confirm cases previously suspected to be GISTs. What we have done was to numerically evaluate the quality of the first discrimination and it was possible by matching information from a personal case study[5] and immunohistochemical behaviors reported in the relevant literature.

The problem can be synthesized as it follows: we have selected as relevant for a lesion the events

label	*description*
DIAGNOSIS	lesion is histologically suspected to be a GIST
GIST	lesion is really a GIST
CD117	KIT protein expression
CD34	Hematopoietic progenitor cell antigen expression
SMA	Muscle actin expression
DESM	Desmin expression
S100	S-100 protein expression

where the first two distinguish the suspected tumors by those actually belonging to the GIST's family, while the others represent the positivity for specific immunohistochemical markers.

[5]Data was collected at *Istituto di Anatomia e Istologia Patologica - Divisione di ricerca sul cancro - Universit degli Studi di Perugia - Italy* during the period Jan.1998–Sept.2002

We had only the following logical restriction due to the extreme specificity of the KIT marker

$$\text{CD117} \subseteq \text{GIST}.$$

By the personal case study we estimated (by observed frequencies) the following "knowledge base"

statement	*cond. prob.*
DIAGNOSIS	.510
CD117 CD34 ¬DESM ¬S100 | DIAGNOSIS	.308
¬SMA ¬CD117 CD34 DESM ¬S100 | DIAGNOSIS	.077
¬SMA CD117 CD34 ¬DESM S100 | DIAGNOSIS	.077
SMA ¬CD117 CD34 ¬DESM ¬S100 | DIAGNOSIS	.077
SMA CD117 ¬CD34 ¬S100 | DIAGNOSIS	.231
SMA CD117 ¬CD34 ¬DESM S100 | DIAGNOSIS	.077
¬SMA CD117 ¬CD34 ¬DESM S100 | DIAGNOSIS	.077

but it turned out to be incoherent with the "rule base" we derived by collecting different literature sources

statement	*expected frequencies bounds*
CD34 | CD117	[.60 , .70]
SMA | CD117	[.30 , .40]
S100 | CD117	[.096 , .105]
DESM | CD117	[.01 , .02]

A deeper analysis of the observed results has shown that there were two cases with dubious S100 positivity and they have judged as the cause of the inconsistency. In fact, performing an inference based only on the knowledge base, we obtain that the percentage for S100 | CD117 results between 13% and 70%, while it should be around 10% as indicated in the rule base.

Revising these two judgements, we have obtained a different knowledge base consistent with the literature rule base

statement	*cond. prob.*
DIAGNOSIS	.510
CD117 CD34 ¬DESM ¬S100 | DIAGNOSIS	.380
¬SMA ¬CD117 CD34 DESM ¬S100 | DIAGNOSIS	.077
SMA ¬CD117 CD34 ¬DESM ¬S100 | DIAGNOSIS	.077
SMA CD117 ¬CD34 ¬S100 | DIAGNOSIS	.077
SMA CD117 ¬CD34 ¬DESM S100 | DIAGNOSIS	.077
¬SMA CD117 ¬CD34 ¬DESM ¬S100 | DIAGNOSIS	.077

Further considerations has induced us to add the further constraint $P(\text{CD117}|\text{GIST}) \in [0.95, 0.99]$ for the sensitivity of the KIT marker.

Putting together all these assessments, they force the usual accuracy indexes to be in the following bounds

index	*description*	*extension bounds*
P(DIAGNOSIS | GIST)	sensitivity	[.47 , .76]
P(¬DIAGNOSIS | ¬GIST)	specificity	[0 , .88]
P(GIST | DIAGNOSIS)	positive predictive value	[.85 , .94]
P(¬GIST | ¬DIAGNOSIS)	negative predictive value	[0 , 69]

that, apart from the positive predictive value, reflect a weak ”influence” of the constraint considered.

Adding to the assessment the probabilistic comparison $P(\text{DIAGNOSIS} \mid \text{GIST}) \geq P(\text{DIAGNOSIS} \mid \neg\text{GIST})$ we have not obtained appreciable improvements.

On the contrary, reasoning as described in Subsection 3.3, we have focused the attention on the ”a priori” values of GIST's incidence. In fact, its coherent bounds result $P(\text{GIST}) \in [.59, .97]$ while one extreme sub-class of the admissible conditional probabilities induce the more restrictive lower bound of .81. Since the pathologist judged as reasonable a variability around 81% of the GISTS's incidence, we have added to the whole assessment the restriction $P(\text{GIST}) \in [.806, .815]$ obtaining the more relevant results

index	*description*	*extension bounds*
P(DIAGNOSIS | GIST)	sensitivity	[.53 , .59]
P(¬DIAGNOSIS | ¬GIST)	specificity	[.58 , .80]
P(GIST | DIAGNOSIS)	positive predictive value	[.85 , .93]
P(¬GIST | ¬DIAGNOSIS)	negative predictive value	[.22 , .32]

that confirm a good positive predictive performance of the diagnostic procedure, while they express a really bad reliability in the case of a negative diagnosis. This, in a way, reverses the role that the KIT marker should have. Instead of being used as a *confirmatory* tool in already suspected cases, it should have a crucial role for the right diagnosis of lesion at first not suspected to be GISTs.

References

[1] A. Capotorti, S. Fagundes Leite. Reliability of GIST diagnosis based on partial information. accepted for the pubblication on *Statistical Inference in Human Biology*, (M. Di Bacco, G. D'Amore, F. Scalfari editors), Kluwer Academic Publishers, Norwell, MA (USA), 2003.

[2] A. Capotorti, L. Galli and B. Vantaggi. How to use locally strong coherence in an inferential process based on upper-lower probabilities. *Soft Computing*, 7(5), 280-87, 2003.

[3] A. Capotorti, F. Lad. Reassessing accuracy rates from median decision procedures. submitted to *The American Statistician*, 2003.

[4] A. Capotorti, T. Paneni. An operational view of coherent conditional previsions. *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Benferhat, S., Besnard, P. (eds): ECSQUARU 2001, LNAI 2143, 132–143, 2001.

[5] A. Capotorti, B. Vantaggi. Locally strong coherence in an inference process. *Annals of Mathematics and Artificial Intelligence*, 35(1-4), 125- 149, 2002.

[6] G. Coletti, R. Scozzafava. Characterization of Coherent Conditional Probabilities as a Tool for their Assessment and Extension. *Int. Journ. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 4(2): 103-127, 1996.

[7] G. Coletti, R. Scozzafava. *Probabilistic Logic in a Coherent Setting*, Dordrecht: Kluwer, Series "Trends in Logic", 2002.

[8] ISIPTA'99. *Proocedings of the First International Symposium on Imprecise Probabilities and Their Applications.* de Cooman, G., Cozman, F.G., Moral, S., Walley, P. (eds.), Imprecise Probabilities Project, Universiteit Gent, Belgium, 1999.

[9] ISIPTA'01. *Proocedings of the Second International Symposium on Imprecise Probabilities and Their Applications.* held at the Robert Purcell Community Center of Cornell University Ithaca, NY, USA, 26-29 June 2001, de Cooman, G., Fine, T. L., Seinfeld, T. (eds.), The Nederlands: Shaker Publishing.

[10] F. Lad. *Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction*, New York: John Wiley, 1996.

[11] F. Lad, M. Di Bacco. Assessing the value of a second opinion: the role and structure of exchangeability. *Annals of Mathematics and Artificial Intelligence*, 35:227–252, 2002.

[12] R. Tweedie, K. Mengersen. Calculating accuracy rates from multiple assessors with limited information. *The American Statistician* , 53, 233–238, 1999.

[13] R.J. Vanderbei, D.F. Shanno. An interior-point algorithm for nonconvex nonlinear programming. Tecnical report SOR-97-21, *Statistics and Operations Research*, Princeton University, 1997.

[14] N.V. Thoai. Duality bound method for the general quadratic programming problem with quadratic constraints. *Jour. of Optimization Theory and Applications*, 107(2): 331–354, 2000.

Andrea Capotorti is with the Dipartimento di Matematica e Informatica at Università degli Studi di Perugia, via Vanvitelli 1, 06123 Perugia, Italy. E-mail: capot@dipmat.unipg.it