KYBERNETIKA — VOLUME *51* (2015), NUMBER 3, PAGES 469–485

MODELING BIASED INFORMATION SEEKING WITH SECOND ORDER PROBABILITY DISTRIBUTIONS

Gernot D. Kleiter

Updating probabilities by information from only one hypothesis and thereby ignoring alternative hypotheses, is not only biased but leads to progressively imprecise conclusions. In psychology this phenomenon was studied in experiments with the "pseudodiagnosticity task". In probability logic the phenomenon that additional premises increase the imprecision of a conclusion is known as "degradation". The present contribution investigates degradation in the context of second order probability distributions. It uses beta distributions as marginals and copulae together with C-vines to represent dependence structures. It demonstrates that in Bayes' theorem the posterior distributions of the lower and upper probabilities approach 0 and 1 as more and more likelihoods belonging to only one hypothesis are included in the analysis.

Keywords: probability logic, Bayes' theorem, degradation, pseudodiagnosticity task, second order probability distributions

Classification: 03B48, 49N30, 62F15, 91E10

> If our object be to discover the effects of an agent A, we must procure A in some set of ascertained circumstances, as A B C, and having noted the effects produced, compare them with the effect of the remaining circumstances B C, when A is absent. (John Stuart Mill, Collected Works, Vol. 7, p. 391, Method of Difference)

1. INTRODUCTION

Imagine the following situation:

> A physician is 50 % sure that a patient is suffering from disease H, $P(H) = .5$. The physician knows that the probability that if the patient is suffering from H, the patient shows symptom E_1 is .7, $P(E_1|H) = .7$. The physician may obtain just one more piece of information out of the following three options:
>
> 1. $P(E_2|H)$, the probability of a second symptom given the presence of the disease,
> 2. $P(E_1|\neg H)$, the probability of the first symptom given the absence of the disease, or

DOI: 10.14736/kyb-2015-3-0469

3. $P(E_2|\neg H)$, the probability of the second symptom given the absence of the disease.

Which option is the best choice?

This is a classical task in the psychology of decision making under uncertainty. It is called the "pseudodiagnosticity task" and was introduced by Doherty, Mynatt, Tweney, and Schiavo [4, 8]. For short here I will call it the "Doherty task". In empirical studies the majority of the participants select the first option, $P(E_2|H)$. Doherty et al. however considered $P(E_1|\neg H)$, the second option, to be the best choice. Only with $P(E_1|\neg H)$ Bayes' theorem can be used to compute the posterior point probability

$$P(H|E_1) = \frac{P(H)P(E_1|H)}{P(H)P(E_1|H) + (1 - P(H)P(E_1|\neg H)}. \tag{1}$$

In the status quo situation, before any of the three options is selected, the posterior probability is constrained by the given values of the prior $P(H)$ and the likelihood $P(E_1|H)$. The unknown value of $P(E_1|\neg H)$ may have any value between 0 and 1 and the posterior probability is therefore in the interval [19]

$$P(H|E_1) \in \left[\frac{P(H)P(E_1|H)}{P(H)P(E_1|H) + 1 - P(H)},\ 1\right]. \tag{2}$$

The lower bound results from $P(E_1|\neg H) = 1$ and the upper bound from $P(E_1|\neg H) = 0$. If however, as most participants do, $P(E_2|H)$ is selected, then the interval is

$$P(H|E_1, E_2) \in \left[\frac{P(H)P(E_1, E_2|H)}{P(H)P(E_1, E_2|H) + 1 - P(H)},\ 1\right]. \tag{3}$$

The interval in (3) is wider than the interval in the status quo (2) because

$$P(E_1, E_2|H) \leq \min\{P(E_1|H), P(H_2|H)\} \leq P(E_1|H).$$

Thus, selecting $P(E_1|\neg H)$ with option 2. results in a precise point probability. Selecting $P(E_2|H)$ with option 1., however, results in an interval that is wider than the initial status quo interval.

The preference for $P(E_2|H)$ is seen as a confirmation bias: people do not consider alternative hypotheses, here the case in which the disease is absent. This results in conclusions that are less precise than in the initial situation. This phenomenon, the degradation of inferences in the light of new information, is not new. The Doherty task, however, is an especially prototypical example.

Assume $P(H) = .5$, $P(E_1|H) = .7$ and $P(E_2|H) = .7$. Then in the status quo the posterior interval is $[.5 \times .7/(.5 \times .7 + .5), 1] = [.4118, 1]$, whereas after selecting $P(E_2|H)$ the interval is $[.5 \times (.7 + .7 - 1)/(.5 \times (.7 + .7 - 1) + .5), 1] = [.2857, 1]$.

If we continue to select only the "affirmative " likelihoods given H and not those given the alternative hypothesis $\neg H$, then the interval gets wider and after a few more steps is completely noninformative, i.e, $[0, 1]$. The additional information makes the situation more uncertain. Selecting the likelihood under the alternative hypothesis, however, leads to a precise point value. It is easily seen that selecting the third option $P(E_2|\neg H)$ results in the noninformative interval $[0, 1]$.

Gilio [5] studied generalized probabilistic inference rules and observed that with rules like the modus ponens or the modus tollens more information leads to less precise inferences. He called this phenomenon *degradation*. Closely related is the property of *dilution* described by Seidenfeld and Wasserman [18]. A similar phenomenon was described in the context of continuous probability functions and called "bad news" [9]. Recently degradation was investigated in probability logic [10, 20, 22, 21].

While usually updating by Bayes' theorem is done by conditioning on a new event or new data, in the Doherty task updating is done by information about a probability. The new evidence is the value of a probability. In the context of Jeffrey's rule this is called updating by *soft evidence*. Below we will distinguish first and second order probabilities. We will update second order probabilities by the values of first order probabilities.

Here we analyse a version of the Doherty task that is more general than the original one. We assume that the prior probability, $P(H)$, and the likelihoods, like $P(E_1|H)$ or $P(E_1|\neg H)$, are only imprecisely known. In real decision making problems this is clearly more realistic than to work with precise point probabilities. It is highly plausible for human representation and processing of uncertainty. Mathematically we express the imprecision by beta distributions. Moreover, we do not assume stochastic independence of the involved events and uncertain quantities. We present a new formal analysis of the Doherty task using continuous second order probability functions; we drop the independence assumption and use copulae and vines to model stochastic dependence structures. References on copulae are [7, 15], a reference on vines is the handbook by Kurowicka and Joe [13]. There is a close relationship between Bayesian networks and vines described in [6, 11, 12]. We will rely on simulation methods. For copulae and vines the simulation algorithms are described in [14]. We will use the R [16] package *VineCopula* provided by [17].

2. MODELING THE DOHERTY TASK

We consider a binary hypothesis, H and $\neg H$, and a set of n binary events, $E_1, \ldots, E_n$ and $\neg E_1, \ldots, \neg E_n$, respectively. First order probabilities are denoted by $P(H)$, $P(E_i|H)$, $P(E_i|\neg H)$ etc. The probability of conjunctions like $P(E_1 \wedge E_2|H)$ is written as $P(E_1, E_2|H)$.

First order probabilities are treated as continuous *uncertain quantities* (random variables) with values between 0 and 1:

$$\begin{aligned} X = P(H), \qquad & Y_1 = P(E_1|H), \ldots, Y_n = P(E_n|H), \quad \text{and} \\ & Z_1 = P(E_1|\neg H), \ldots, Z_n = P(E_n|\neg H)\,. \end{aligned}$$

The uncertain quantities span a $d = 1 + 2n$ dimensional unit-cube. Each uncertain quantity is characterized by a beta distribution,

$$\begin{aligned} X \sim Be(a_X, b_X), \qquad & Y_1 \sim Be(a_{Y_1}, b_{Y_1}), \ldots, Y_n \sim Be(a_{Y_n}, b_{Y_n}), \\ & Z_1 \sim Be(a_{Z_1}, b_{Z_1}), \ldots, Z_n \sim Be(a_{Z_n}, b_{Z_n})\,. \end{aligned}$$

The *events* $E_1, \ldots, E_n$ are not assumed to be conditionally independent given H and, likewise, $E_1, \ldots, E_n$ are not assumed to be conditionally independent given $\neg H$.

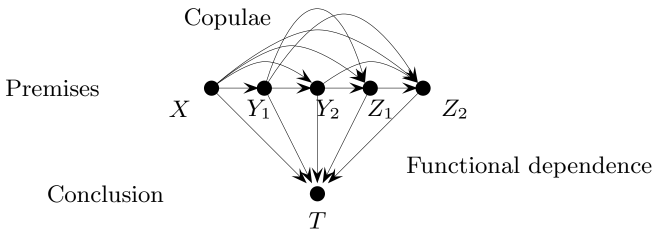


Fig. 1. Bayesian network of an inference with $d = 5$ $(n = 2)$ premises X, Y_1, Y_2, Z_1, and Z_2 and the target function T (Bayes' theorem, for example,) representing the conclusion. The probability of the conclusion is functionally dependent upon the joint cumulative probability distribution $F(X, Y_1, Y_2, Z_1, Z_2)$.

Correspondingly, the *uncertain quantities* $X, Y_1, \ldots, Y_n$, $Z_1, \ldots, Z_n$ are not assumed to be independent. Their dependence structure is characterized by an d-dimensional copula. This multivariate copula is decomposed by a sequence of $\binom{d}{2}$ pairwise copulae.

First order probabilities are sometimes interpreted as "objective" probabilities and second order probabilities as subjective probabilities, degrees of belief, or epistemic probabilities. Psychologically it is a big advantage to be able to express correlated degrees of belief by copulae.

We investigate the updating of $X = P(H)$ by a sequence of $Y_i = P(E_i|H)$ and $Z_i = P(E_i|\neg H)$, that is, we update second order probability densities by the values of first order probabilities, not by binary events. The updating is done by Bayes' theorem which acts as a *target function* of the uncertain quantities X, Y_i, and Z_i. The uncertain quantities, their dependence structure, and the target function may be represented by two kinds of graphical models, as a Bayesian network or as a vine. We assume that the density functions are absolutely continuous—a condition met by the family of beta distributions with shape parameters $0 < a, b < \infty$.

Figure 1 shows the complete graph of a Bayesian network with five variables. Independencies might be encoded by removing directed edges. Figure 2 shows a C-vine in which the pairwise copulae and the pairwise conditional copulae are represented by the edges of the graph. Vines are usually graphically represented by a sequence of trees in which the edges of one tree become the nodes of the next following tree. The representation in Figure 2 is more closely related to the factorization principle in Bayesian networks. In vines independencies are encoded by product copulae.

For point probabilities Bayes' theorem states

$$P(H|E_1, \ldots, E_n) = \frac{P(H)P(E_1, \ldots, E_n|H)}{P(H)P(E_1, \ldots, E_n|H) + (1 - P(H))P(E_1, \ldots, E_n|\neg H)}. \tag{4}$$

As we do not assume conditional independence of the events the probabilities of the con-

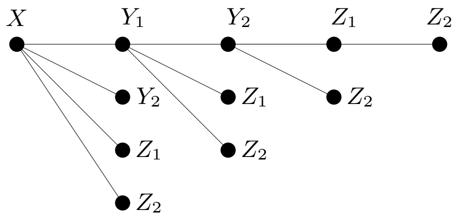


Fig. 2. C-vine encoding of the dependence structure of the five variables X, Y_1, Y_2, Z_1, and Z_2. The ten pairwise copulae specify $C(X, Y_1)$, $C(X, Y_2)$, $C(X, Z_1)$, (X, Z_2), $C(Y_1, Y_2|X)$, $C(Y_1, Z_1|X)$, $C(Y_1, Z_2|X)$, $C(Y_2, Z_1|X, Y_1)$, $C(Y_2, Z_2|X, Y_1)$ and $C(Z_1, Z_2|X, Y_1, Y_2)$.

junctions are interval probabilities with the bounds [1, p. 298ff.]

$$\begin{aligned} P(E_1, \ldots, E_n|H) &\in \left[\max\left\{0, \sum_{i=1}^{n} P(E_i|H) - (n-1)\}, \min_{i=1}^{n}\{P(E_i|H)\right\}\right], \\ P(E_1, \ldots, E_n|\neg H) &\in \left[\max\left\{0, P(E_i|\neg H) - (n-1)\}, \min_{i=1}^{n}\{P(E_i|\neg H)\right\}\right]. \end{aligned}$$

Expressed by the values of the uncertain quantities the lower bound of Bayes' theorem is

$$P(H|E_1, \ldots, E_n) \geq \frac{\max\{0, x(\sum_{i=1}^{n} y_i - (n-1))\}}{\max\{0, x(\sum_{i=1}^{n} y_i - (n-1))\} + (1-x)\min\{z_i\}} \tag{5}$$

and the upper bound is

$$P(H|E_1, \ldots, E_n) \leq \frac{x \min\{y_i\}}{x\min\{y_i\} + (1-x)(\max\{0, \sum_{i=1}^{n} z_i - (n-1))\}} \tag{6}$$

[21, 22].

In terms of distributions imprecision about the value of a probability may be expressed by a beta distribution $Be(a, b)$. First, this choice is motivated by pragmatic reasons. Beta distributions are easy to handle and so rich to express not only symmetrical and non-symmetrical unimodal distributions, but also U- and J-shaped distributions. Indifference about the value of a first order probability is expressed by the uniform distribution $Be(1, 1)$. A second reason for the choice of beta distributions is their conjugate relationship to the frequency of independent binary events: Bayes' theorem updates prior beta distributions by the number of observed successes and failures so that the posterior is again a beta distribution. A third reason is the ease of the assessment of beta distributions, investigated and used in decision analysis, Bayesian statistics, and psychology.

2.1. Replacing an interval by a distribution

In the original Doherty task two probabilities are given, $P(H) = h$ and $P(E_1|H) = a$, but not the value of $P(E_1|\neg H)$. All one can say about the unknown value is that it is between 0 and 1. Do all values in $[0, 1]$ have "equal rights"? What if the values in $[0, 1]$ are weighted by a distribution? We will now express the lack of knowledge about the missing $P(E_1|\neg H)$ by a beta distribution, $Z_1 \sim Be(a_{Z_1}, b_{Z_1})$. We may choose the uniform distribution $Be(1, 1)$ or, as in the example below $Be(1.5, 1.5)$, which is flat, symmetric, and slightly favors probabilities around .5.

Before selecting any additional information Bayes' theorem gives

$$t = \frac{ha}{ha + (1-h)z_1}, \quad t \in \left[\frac{ha}{ha + 1 - h}, 1\right]. \tag{7}$$

We find the probability density function of t by a change of variable $z_1 \to t$. We solve (7) for z_1

$$z_1 = \frac{ha(1-t)}{(1-h)t}, \tag{8}$$

and use the Jacobian, i. e., the positive first derivative $|\mathrm{d}t/\mathrm{d}z_1|$, obtained by the quotient rule $(u/v)' = (vu' - uv')/v^2$,

$$|t'| = \frac{ha(1-h)}{(ha + (1-h)z_1)^2}. \tag{9}$$

We divide the beta distribution by the Jacobian and obtain

$$p(t) = Be(z_1; a_{Z_1}, b_{Z_1}) \frac{(ha + (1-h)z_1)^2}{ha(1-h)}, \quad \text{where} \quad z_1 = \frac{ha(1-t)}{(1-h)t}. \tag{10}$$

For the special case of a uniform distribution, $Be(1, 1)$, we have

$$p(t) = \frac{ha}{(1-h)t^2}, \quad t \in \left[\frac{ha}{ha + (1-h)}, 1\right]. \tag{11}$$

If we select option 3 (the value y_2), its value is c, and if we introduce some simplifying assumptions, then the situation is completely analog to the status quo. We assume that (i) E_1 and E_2 are conditionally independent given H, so that $P(E_1, E_2|H) = ac$ and that (ii) the lack of knowledge about $P(E_1, E_2|\neg H)$ is expressed by the uniform $Be(1, 1)$. Now only the product ha changes to hac. The lower bound becomes

$$\frac{hac}{hac + 1 - h},$$

that is, it decreases:

$$\frac{hac}{hac + 1 - h} \leq \frac{ha}{ha + 1 - h}.$$

More information makes the situation less informative.

One might argue that in the status quo one should also include the ignorance about the likelihoods of $P(E_2|H)$ and $P(E_2|\neg H)$. But as nothing is known about both of them

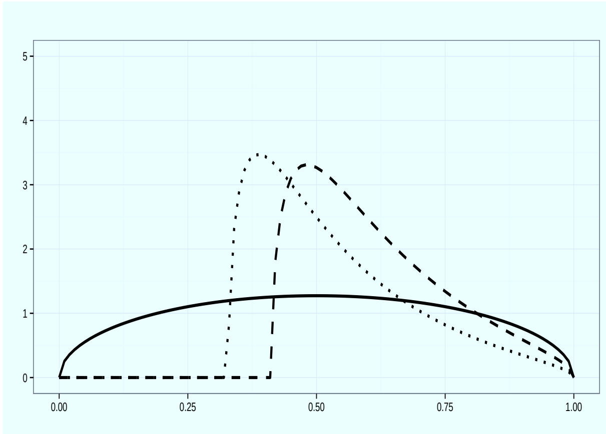


Fig. 3. (a) Posterior distribution (dashed) assuming the prior probability $P(H) = h = .5$, the likelihood $P(E_1|H) = a = .7$, and the non-informative beta distribution for the unknown likelihood, $Z_1 = P(E_1|\neg H) \sim Be(1.5, 1.5)$ (solid). (b) Posterior (dotted) after including in addition the information y_2 with the value $P(E_2|H) = c = .7$.

this would open the door for including an infinite sequence of arbitrary variables about which nothing is known.

Figure 3 shows a numerical example. For the status quo condition the prior probability is $P(H) = h = .5$, the first likelihood is $P(E_1|H) = a = .7$, and the unknown second likelihood $P(E_1|\neg H)$ is beta distributed, $Z_1 \sim Be(1.5, 1.5)$ with mean .5 and sd .25, shown by the continuous curve. The dashed curve shows the distribution of $t = P(H|E_1)$ with mean .61 and sd .14. We now select the information $P(E_2|H) = y_2$ and assume that its value is also $c = .7$. Moreover, we assume (i) that $P(E_1|H)$ and $P(E_2|H)$ are conditionally independent so that $P(E_1, E_2|H) = ac = .49$, and (ii) that the lack of knowledge about $P(E_1, E_2|\neg H)$ is expressed by the beta $Be(1.5, 1.5)$. The resulting posterior distribution with mean .53 and sd .15 is shown by the dotted curve in Figure 3. It is located left of the curve for the status quo and favors smaller probabilities.

If we select $P(E_1|\neg H)$ (Doherty's rational choice) the uncertain quantity Z_1 is instantiated and becomes a constant. As a consequence, the posterior probability becomes a precise point probability.

The preceding analysis was based on three assumptions:

1. The probabilities $P(H) = h$, $P(E_1|H) = a$, and $P(E_2|H) = c$ are precise.

2. The events E_1 and E_2 are conditionally independent given H.

3. The lack of knowledge about the likelihood $P(E_1, E_2|\neg H)$ is expressed by a non-informative distribution.

	X	Y_1	Y_2
X	1	1	1
Y_1		2	2
Z_1			3

Tab. 1. Regular vine matrix representation of a C-vine with three variables. Note that the shown upper matrix follows [14] and is different from the encoding in the VineCopula package [17].

In real problems especially the first two assumptions are unrealistic. For a more general approach we drop the requirement of precise point probabilities and independence.

2.2. Representing the joint distribution by marginals and copulae

Instead of precise values for $P(H)$ and $P(E_1|H)$, we now use the uncertain quantities X and Y_1 and express our imprecise knowledge by two betas, $X \sim Be(a_X, b_X)$ and $Y_1 \sim Be(a_{Y_1}, b_{Y_1})$. The shape parameters of these distributions will be clearly greater than 1. Accordingly, we express the lack of knowledge of the likelihoods $P(E_1|\neg H)$, $P(E_2|H)$, $P(E_2|\neg H)$ by beta distributions with shape parameters close to 1.

Modeling the knowledge by these marginals only would be incomplete without specifying how they combine in a joint distribution. The five probabilities (!) might be independent. But this seems unrealistic and unnecessarily restrictive. It is more satisfactory to invoke a general dependence structure for the relationships between the degrees of belief. It may easily be the case, for example, that knowing that the probability of a disease H is .9, that I believe that the probability of a symptom is also high. In this case the probability of the disease and the probability of a symptom are correlated.

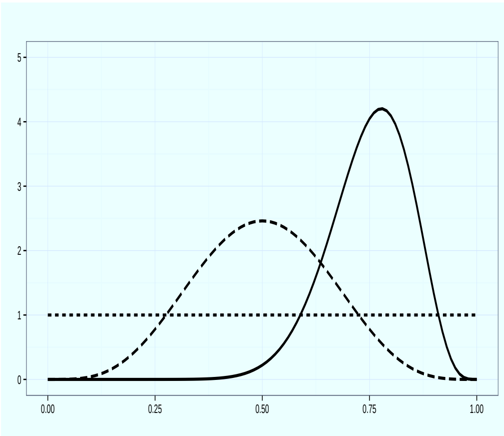
Finding the posterior distribution in an analysis of by now five dimensions with flexible marginals and flexible copulae is analytically not tractable. For the analysis of such problems simulation methods are used in the literature. We rely on the simulation methods provided in the VineCopula package of R [17].

Assume that in the status quo we have the following marginals

$$\begin{aligned} P(H) = X \sim Be(5,5), \qquad & P(E_1|H) = Y_1 \sim Be(15,5), \\ & P(E_1|\neg H) = Z_1 \sim Be(1,1)\,. \end{aligned}$$

The dependence structure of the C-vine with three variables is specified in the regular vine matrix in Table 1. We assume Gaussian copulae with parameters $\rho = 0$. The left panel of Figure 4 shows a summary of the status quo: The prior distribution and the distribution of the two liklihoods. The right hand panel of Figure 4 shows a histogram of the posterior distribution of the target function $T = P(H|E_1)$ obtained by simulation with the VineCopula package and $N = 100.000$.

We now select Y_2. We assume $Y_2 = P(E_2|H) \sim Be(15,5)$, $Z_2 \sim Be(1,1)$, and use again Gaussian copulae with $\rho = 0$. The C-vine is encoded by the regular vine matrix in Table 2. The copulae are $C(1,2)$, $C(1,3)$, $C(1,4)$, $C(1,5)$, $C(2,3|1)$, $C(2,4|1)$, $C(2,5|1)$,



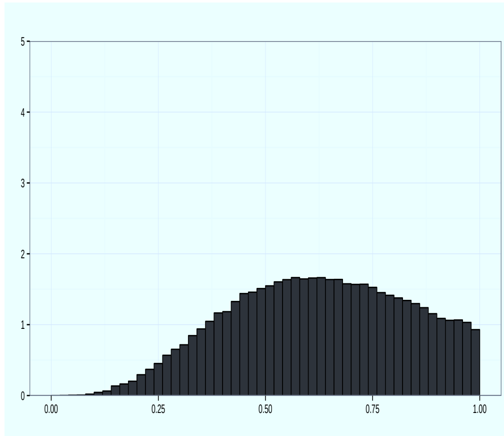


Fig. 4. One binary event (three variables). Left: Beta distributions of the prior $Be(5,5)$, the likelihood $Y_1 = P(E_1|H) \sim Be(15,5)$, and $Z_1 = Be(E_1|\neg H) \sim Be(1,1)$. Right: Histogram of the posterior distribution. Simulation with $N = 100.000$.

$C(3,4|1,2)$, $C(3,5|1,2)$, and $C(4,5|1,2,3)$. The hierarchical structure is depicted in Figure 2. It is closely related to the principle of factorization in a Bayesian network.

With two pieces of evidence, E_1 and E_2, the (first order) probability of their conjunction is not known. Thus, $P(E_1, E_2|H) \in [\max\{0, y_1 + y_2 - 1\}, \min\{y_1, y_2\}]$ and the posterior probability is an interval probability obtained by equations (5) and (6). As a consequence, we have *two* target functions, one for the lower and one for the upper bounds of the posterior probabilities.

The three histograms on the top of Figure 5 show the lower and upper posterior probabilities for $X \sim Be(5,5), Y_1 \sim Be(15,5)$, and $Y_2 \sim Be(15,5)$, $Z_1 \sim Be(1,1)$, and $Z_2 \sim Be(1,1)$; all copulae are Gaussian with $\rho = 0$. On the left is the histogram for the lower bound, on the right for the upper bound, and in the middle the histogram for independent events. Comparing the right hand panel of Figure 4 with the left and

	X	Y_1	Y_2	Z_1	Z_2
X	1	1	1	1	1
Y_1		2	2	2	2
Y_2			3	3	3
Z_1				4	4
Z_2					5

Tab. 2. Regular vine matrix representation of a C-vine with five variables. Note that the shown upper matrix follows [14] and is different from the encoding in the VineCopula package [17].

the right histograms in the first row of Figure 5 shows that the selection of Y_2 increases the imprecision of the posterior distribution. As long as we do not know the kind of dependence or independence of the events only a lower and an upper distribution can be inferred.

We next extend the number of options in the Doherty task and include a third event E_3 with $Y_3 = P(E_3|H) \sim Be(15,5)$ and $Z_3 = P(E_3|\neg H) \sim Be(1,1)$. The three histograms in the second row from the top of Figure 5 show the resulting distributions. The modes of the histograms of the lower and upper bounds are at 0 and 1. Note that the ordinates are truncated at 0 and 1. As we continue to include more and more information from only one hypothesis the precision of the conclusion degradates.

2.3. Degradation

In probability logic an inference rule degrades if adding more probabilistic premises (of the same kind) leads to interval probabilities of the conclusion that are contained as subsets in the next following interval [21]. While important inference rules like the and-introduction or the modus ponens degrade, Bayes' theorem does not always degrade (for an example see [21]). If however the $P(E_i|H)$ do have identical or increasing values, then the upper bounds increase monotonically. In this case the min-terms in Eq. (6) are constant whereas the probability of the conjunction in the second term in the denominator decreases monotonically. Similarly, if the $P(E_i|\neg H)$ in Eq. (5) are identical or increasing, then the lower bounds decrease monotonically.

To apply the concept of degradation to inferences with distributions we borrow a concept from statistics. In statistics the condition of independent and identically distributed sampling is of eminent importance. We drop the independence and keep the identical distributions. An inference is an inference with *completely identical distributed premises* (completely i.d. premises) if (i) the distributions of all premises with the same logical form are identical and if (ii) all pairwise copulae belong to the same family. If only the premises belonging to one class with the same logical form have identical distributions and all the pairwise copulae of these premises belong to the same family, we say the premises are *partially identical distributed* (partially i.d).

Consider, for example, the $2n+1$ premises entering Bayes' theorem with the distributions

$$X \sim Be(a_X, b_X), \quad Y_i \sim Be(a_{Y_i}, b_{Y_i}), \quad Z_i \sim Be(a_{Z_i}, b_{Z_i}), \quad i = 1, \ldots, n\,.$$

If $a_{Y_i} = a_{Y_j}$, $b_{Y_i} = b_{Y_j}$ and $a_{Z_i} = a_{Z_j}$, $b_{Z_i} = b_{Z_j}$, $i, j = 1, \ldots, n$ and $i \neq j$, and if all pairwise copulae are Gaussian, for example, then the premises are completely i.d. If, for example, all distributions of the Z_i have uniform distributions $Be(1,1)$ and are independent, but the Y_i have beta distributions with different shape parameters, then the premises are partially i.d.

We now define the concept of degradation by a dominance relation. Let X and Y be two first order probabilities with the cumulative distribution functions $F_X(u)$ and $F_Y(u)$. X stochastically dominates Y, for short $X \succeq Y$, iff

$$F_X(u) \leq F_Y(u), \quad \forall u \in [0,1] \quad \text{and} \quad \exists u \text{ such that} \quad F_X(u) < F_Y(u)\,.$$

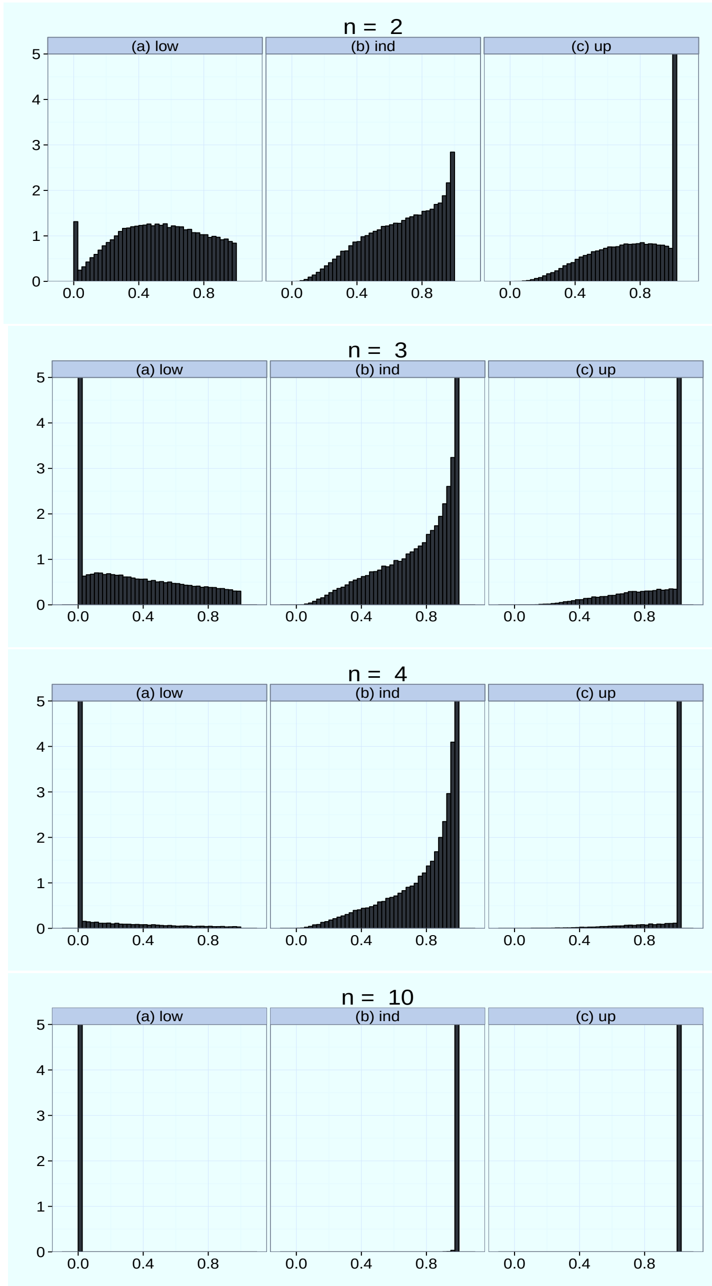


Fig. 5. Posterior histograms. From top to bottom: n= 2, 3, 4, and 10 events; left: lower bounds, right: upper bounds, middle: independence. Prior $Be(5,5)$, same likelihoods $Y_i = P(E_i|H) \sim Be(15,5)$, and $Z_i = Be(E_i|\neg H) \sim Be(1,1)$ for $i = 2,3,4,10$ events. Simulation with $N = 100.000$.

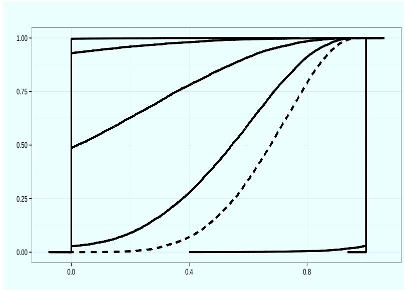


Fig. 6. Stochastic dominance and degradation in Bayes' theorem. Cumulative posterior distribution for the status quo with $n = 1$ event (dotted line). Cumulative posterior distributions of the lower bounds for 2, 3, 4 and 5 events (from right to left of the dotted line) and for the upper bounds (step function on the right of the dotted line). The four upper bound cdfs are so close that they are visually indistinguishable. $X \sim Be(5, 5)$, $Y_i \sim Be(10, 5)$, and $Z_i \sim Be(1, 1)$, $i = 1, \ldots, 5$, Gaussian copulae with $\rho = 0$ for $i = 2, \ldots, 5$.

A sequence of distributions $F_1, \ldots, F_n$ is increasingly dominant iff $F_k \succeq F_{k+1}$ and decreasingly dominant iff $F_k \preceq F_{k+1}$, $k = 1, \ldots, n-1$.

An inference rule degrades, if the distributions of its lower bounds are decreasingly dominant and those of its upper bounds are increasingly dominant.

We state the following *conjecture*: Bayes' theorem with partially identical premises degrades. We have, however, no proof for this conjecture. Because of the involved copulae and conditional copulae the proof seems to be difficult.

In the Doherty task it is reasonable to model the unknown Z_i by identical uniform or close to uniform distributions and by a multidimensional product copula. We thus have partially i.d. premises inducing degradation of the lower bound. Lower probabilities get more plausible. If, in addition, also the distributions of the Y_i are identical, this leads to completely i.d. premises and to degradation of the distributions of the lower and the upper bounds. In the example shown in Figure 5 the premises are completely i.d., thus the distributions of the lower and the upper bounds degrade. The example in Figure 6 illustrates stochastic dominance of the posterior distributions of the lower bounds for an increasing number of events and completely i.d. distributions.

The histograms in Figure 5 show the posterior distributions for the lower (left) and upper (right) bounds and for independence (middle) for $n = 2, 3, 4, 10$ five events (corresponding to $d = 5, 7, 9, 11$ variables). The distributions of the lower and upper probabilities approach distributions with sharp peaks with probability mass equal to one at 0 and 1, respectively.

If the Y_i and the Z_i are independent, numerical investigations show several interesting results. If all distributions are uniform $Be(1, 1)$, for example, then the posterior distributions are U-shaped with maximums at 0 and 1. As n increases 50 % of the probability mass concentrates at 0 and 50 % at 1—the most extreme form of "don't

	X	Y_1	Y_2	Y_3	Z_1	Z_2	Z_3
X		.8	.7	.6	-.5	-.4	-.3
Y_1			.7	.7	-.7	-.7	-.7
Y_2				.7	-.5	-.5	-.5
Y_3					-.3	-.3	-.3
Z_1						.3	.3
Z_2							.3
Z_3							

Tab. 3. Parameters (ρ) for Gaussian copulae. In the first row correlations $C(X, Y_1), \ldots, C(X, Z_3)$, in the second row conditional correlations $C(Y_1, Y_2|X), \ldots, C(Y_1, Z_3|X)$ up to $C(Z_2, Z_3|X, Y_1, Y_2, Y_3, Z_1)$ in the last row.

know". If the shape parameters of the beta distributions of Y_i and Z_i favor H, then the means of the sequence of distributions increase and the posterior distributions becomes more and more J-shaped with modes at 1.

The effects of the different kinds of copulae and copula parameters on the posterior distributions in a Doherty task is illustrated by the following example.

We consider the three binary event E_1, E_2, and E_3 with the marginals

$$X \sim Be(5,5), \qquad Y_1 \sim Be(15,5), Y_2 \sim Be(15,5), Y_3 \sim Be(15,5), \tag{12}$$

$$Z_1 \sim Be(1,1), Z_2 \sim Be(1,1), Z_3 \sim Be(1,1). \tag{13}$$

We encode the dependence structure again by a C-vine and by Gaussian copulae. We use, however, correlations as shown in Table 3.

The probability of the disease H correlates positively with the presence of the symptoms E_1, E_2, and E_3 ($\rho = .7$) and negatively with their absence ($\rho = -.7$). The probabilities of the presence of the symptoms are positively inter-correlated given the probability of the disease. Likewise, given the probability of the absence of the disease the probabilities of the absence of the symptoms are positively inter-correlated ($\rho = .3$).

In the example the effect of introducing correlations different from 0 leads to a bimodal U-shaped posterior distribution of the lower bound (Figure 7). The conclusion becomes more imprecise. A loss of independence among the premises induces a loss of information transmitted to the conclusion. A similar effect results if the Gaussian copulae are replaced by copulae that assign higher joint probabilities to the tails.

One method to investigate degradation in the distributional approach is to study the (second order)probability mass at (first order) probabilities 0 and 1. In the examples we have seen that after introducing only a very few more events the lower and upper posterior distributions concentrate all their mass at 0 and 1. Table 4 shows the frequencies of the posterior probabilities at 0 and at 1 as the number of events increases from 1 to 9. The counts are based on the simulation of $N = 100.000$ cases. The probability mass of the lower bound converges at 0, of the upper bound at 1.

Also included are the frequencies of the posterior probabilities at 0 and 1 under the independence assumption. The posterior distribution gets more and more J-shaped with its mode at 1. It behaves similar as the upper bound in the analysis without the

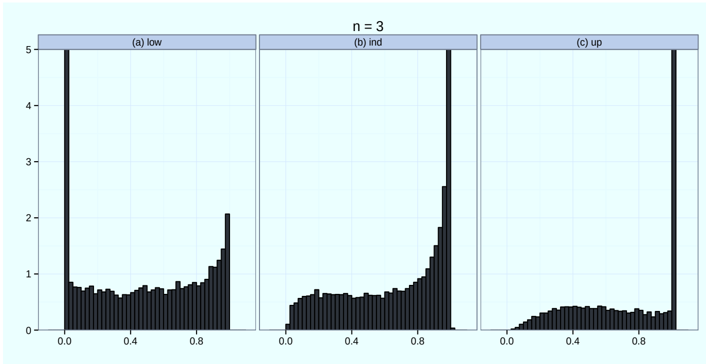


Fig. 7. Three events (7 variables) same prior and likelihood distributions as in the example in Figure 5, but with the correlations as shown in Table 3.

independence assumption. This demonstrates how strong the independence assumption is. It gives rise to the following conjecture: *With independent and identically distributed premises the posterior distribution converges, without independence it quickly diverges.*

The independence assumption may be a justified for statistical samples. If the events are heterogeneous, as in the context of medical diagnosis or financial risk assessment, the assumption is not justified. Usually markers of diseases or indicators of financial risks are not conditionally independent.

It is interesting to compare the behavior of the lower and upper posterior distributions with the behavior of upper and lower probabilities for inferences with exact point probabilities. The lower and upper point probabilities are obtained by Eq. (5) and Eq. (6), respectively.

n events	1	2	3	4	5	6	7	8	9
% at 0									
Lower bound	0	2.94	49.02	93.03	99.75	100.00	100.00	100.00	100.00
% at 100									
Upper bound	.80	50.06	83.24	95.70	99.23	99.86	99.98	100.00	
Independence	.80	2.96	7.40	13.64	21.36	29.98	38.38	46.02	54.21

Tab. 4. Percentages of posterior probabilities equal to 0 (upper part) and equal to 1 (lower part) out of 100.000 simulations. As the number of binary events n increases the lower and upper bounds of the posterior probability converge 0 and 1. All marginal distributions of $Y_i = P(E_i|H)$ are beta distributions $Be(10, 5)$, all marginals of $Z_i = P(E_i|\neg H)$ are noninformative $Be(1, 1)$, $i = 1, \ldots, n$. All copulae are Gaussian with $\rho = 0$.

We take the values $P(H) = .5$, $P(E_i|H) = 2/3$, $P(E_i|\neg H) = .5)$, $i = 1, \ldots, n$, which correspond to the means of the distributions in the previous example. For $n = 1$ $P(H|E_1) = .57$, for $n = 2$ we have $P(H|E_1, E_2) \in [.4, 1]$, and for $n = 3$ or more events the posterior interval is already noninformative $[0, 1]$. Degradation is slower in the distributional approach.

3. CONCLUSIONS

We have seen that in a setting with distributionally imprecise probabilities biased information seeking induces highly inconclusive inferences. The Doherty task is a prototypical example. A *diagnostic choice*, that is, selecting likelihood information about the alternative hypothesis is crucial. A *pseudodiagnostic* choice, that is, selecting further likelihood information about only one hypothesis leads to highly imprecise posterior distributions.

The difference between a diagnostic and a pseudodiagnostic choice becomes even more obvious as we add more options, so that n more probabilities may be selected. Should we select only likelihoods given H or should our selection be balanced by selecting in parallel likelihoods given $\neg H$? As n increases the posterior interval approaches $[0, 1]$ and the second order distributions of the lower and upper posterior probabilities approach δ-functions at 0 and 1.

We modelled uncertain knowledge by beta distributions, copulae, and vines. This allows to represent correlated beliefs and the updating of beliefs (distributions) by soft evidence (probabilities). We defined degradation in the context of distributions by the property of stochastic dominance. We stated the following conjectures

- Biased information seeking leads to conclusions that are less informative than the original situation with no information seeking at all.

- If Y_i and Z_i are identical and independently distributed and $a_Y > b_Y$ or $a_Z < b_Z$, then the posterior distribution converges at 1 as the number of premises increases. If $a_Y < b_Y$ or $a_Z > b_Z$, then the distribution convergences at 0.

- With identically (but not independently) distributed premises Y_i and Z_i the distributions of the lower and upper bounds of the posterior distributions degrade rapidly to 0 and 1. This seems to hold for a wide spectrum of copulae.

- Precise premises induce stronger degradation than imprecise ones. Point probabilities induce maximum degradation.

Especially the conjectures involving copulae are difficult to proof.

We did not consider structural zeroes or logical constraints among the variables (see, for example, [3]). Moreover, we did not consider the problem of degradation from the perspective of the Dempster-Shafer belief function approach. Prior beliefs may be expressed by belief functions and combined with the likelihoods to infer the lower envelope of the possible posterior probabilities [2, 23]. The lower and upper envelopes correspond to the coherent lower and upper conditional probabilities [2]. Belief functions may be appropriate in the case in which "...the prior knowledge could be only partially specified or, even worse, it could refer to a different space of hypothesis." [2].

Sampling from only one condition violates a fundamental methodological principle, the use of control conditions. The Doherty task demonstrates the importance of John Stuart Mill's Method of Difference, cited as dictum at the beginning of the paper. The Doherty task is a typical example in which degradation is observed. It demonstrates how easily the quality of inferences can break down with unbalanced and nonindependent information. Seidenfeld and Wasserman [18] remark that one might consider to pay some extra money to protect oneself from information leading to dilution. Psychologically degradation can be notoriously counterintuitive.

There are several related problems in statistics, such as excluding noisy variables in model selection (e. g. by introducing a penalty for the inclusion of additional variables), missing data problems, robustness, or the handling of unbalanced designs in linear models. The methodology of experimental design, however, protects against unbalanced information seeking and degradation by demanding control groups, independent sampling, and randomization.

ACKNOWLEDGEMENTS

Supported by the Austrian Science Foundation (I-141-G15) within the LogICCC programme of the European Science Foundation. Many thanks to Wolfgang Trutschnig! He helped me to bring more order into my thoughts and to avoid a number of errors. The remaining ones are completely my fault.

(Received April 29, 2014)

REFERENCES

[1] G. Boole: An Investigation of the Laws of Thought. Macmillan/Dover Publication, New York 1854/1958.

[2] G. Coletti, D. Petturiti, and B. Vantaggi: Bayesian inference: the role of coherence to deal with a prior belief function. Statist. Methods Appl., online, 2014.

[3] G. Coletti and R. Scozzafava: Probabilistic Logic in a Coherent Setting. Kluwer, Dordrecht 2002.

[4] M. E. Doherty, C. R. Mynatt, R. D. Tweney, and M. D. Schiavo: Pseudodiagnosticity. Acta Psychologica *43* (1979), 111–121. DOI:10.1016/0001-6918(79)90017-9

[5] A. Gilio: Generalization of inference rules in coherence-based probabilistic default reasoning. Int. J. Approx. Reasoning *53* (2012), 413–434. DOI:10.1016/j.ijar.2011.08.004

[6] A. Hanea: Dependence modeling. Vine copula handbook. In: Dependence Modeling. Vine Copula Handbook (D. Kurowicka and H. Joe, eds.), chapter Non-parametric Bayesian belief nets versus vines, World Scientific, New Jersey 2011, pp. 281–303. DOI:10.1142/9789814299886_0014

[7] H. Joe: Dependence Modeling with Copulas. Chapman and Hall/CRC, Boca Raton 2015.

[8] L. Kern and M. E. Doherty: "Pseudodiagnosticity" in an idealized medical problem-solving environment. J. Medical Education *57* (1982), 100–104.

[9] G. D. Kleiter: Propagating imprecise probabilities in Bayesian networks. Artificial Intelligence *88* (1996), 143–161. DOI:10.1016/s0004-3702(96)00021-5

[10] G. D. Kleiter: Ockham's razor in probability logic. In: Synergies of Soft Computing and Statistics for Intelligent Data Analysis (R. Kruse, M.xQ,R. Berthold, C. Moewes, M. A. Gil, P. Grzegorzewski, and O. Hryniewicz, eds.), Advances in Intelligent Systems and Computation *190*, Springer, Heidelberg 2012. pp. 409–417. DOI:10.1007/978-3-642-33042-1_44

[11] D. Kurowicka and R. Cooke: Distribution-free continuous Bayesian belief nets. In: Proc. Fourth International Conference on Mathematical Methods in Reliability Methodology and Practice, Santa Fe 2004.

[12] D. Kurowicka and R. Cooke: Uncertainty Analysis with High Dimension Dependence Modelling. Wiley, Chichester, 2006.

[13] D. Kurowicka and R. Joe: Dependence Modeling: Vine Copula Handbook. World Scientific, Singapure 2011.

[14] J.-F. Mai and M. Scherer: Simulating Copulas. Stochastic Models, Sampling Algorithms, and Applications. Imperial College Press, London 2012.

[15] R. B. Nelsen: An introduction to Copulas. Springer, Berlin 2006.

[16] R Development Core Team, Vienna, Austria: R: A Language and Environment for Statistical Computing, 2014.

[17] U. Schepsmeier, J. Stoeber, E. C. Brechmann, and B. Graeler: Statistical inference of vine copulas. Version 1.2 edition, 2013.

[18] T. Seidenfeld and L. Wasserman: Dilation for sets of probabilities. Ann. Statist. *21* (1993), 1139-1154. DOI:10.1214/aos/1176349254

[19] R. D. Tweney, M. E. Doherty, and G. D. Kleiter: The pseudodiagnosticity trap. Should subjects consider alternative hypotheses? Thinking and Reasoning *16* (2010), 332–345. DOI:10.1080/13546783.2010.525860

[20] C. Wallmann and G. D. Kleiter: Exchangeability in probability logic. In: Communications in Computer and Information Science (S. Greco, B. Bouchon-Meunier, G. Coletti, M. Fedrizzi, B. Matarazzo, and R. R. Yager, eds.), IPMU (4) *300*, Springer, Berlin 2012, pp. 157–167. DOI:10.1007/978-3-642-31724-8_17

[21] C. Wallmann and G. D. Kleiter: Degradation in probability logic: When more information leads to less precise conclusions. Kybernetika *50* (2014), 268–283. DOI:10.14736/kyb-2014-2-0268

[22] C. Wallmann and G. D. Kleiter: Probability propagation in generalized inference forms. Studia Logica *102* (2014), 913–929. DOI:10.1007/s11225-013-9513-4

[23] L. A. Wasserman: Prior envelopes based on belief functions. Annals Statist. *18* (1990), 454–464.

Gernot D. Kleiter, Department of Psychology, University of Salzburg, Hellbrunnerstr. 34, A-5020 Salzburg. Austria.
e-mail: gernot.kleiter@sbg.ac.at