Statistica Neerlandica (2001) Vol. 55, nr. 3, pp. 346–357

How to assign probabilities if you must

C. J. Albers[1] and W. Schaafsma

Department of Mathematics, University of Groningen, P.O. Box 800, NL-9700 AV Groningen, the Netherlands

Empirical evidence can sometimes be incorporated in a probabilistic analysis by conditioning with respect to the observations. Usually, the underlying probability distribution and also the conditional distribution are not completely known. The assignment of probabilities will then require a compromise. The making of such a compromise goes beyond mathematical theory: a statistical discussion is needed. It depends on the context whether the result of such discussion is almost compelling, reasonable, or not really agreeable. This is illustrated by means of a simple example from the area of predictive distributional inference.

Key Words and Phrases: epistemic probabilities, Wald's decision functions, proper loss functions.

1 Introduction

Most theories of probability, Bayesian statistics included, prescribe the incorporation of empirical evidence by computing conditional distributions. The availability of such prescriptions suggests that the approach is compelling. This may be the case if a fair die is rolled once and somebody has been instructed to tell us, without lying, whether the number of eyes y is even or odd, but the compellingness disappears if the instructions are less specific. If somebody provides us with the information that y is even, it can be very misleading to infer that the probabilities of the possibilities $y = 2$, 4 and 6 are equal to $1/3$.

The example. A fair die has been rolled once and the true number of eyes y has been made available to some person (or Nature), henceforth referred to as Player I. Player I has to provide Player II (the statistician) with true information x about y. He has to choose one of the statements made in Table 1. Note that there is a difficulty if $y = 6$. In that case, Player I has to choose between $x = 2$ and $x = 3$.

Table 1. The outcome space

$x = 1$:	y is neither even, nor a triple
$x = 2$:	y is even
$x = 3$:	y is a triple

[1] casper@math.rug.nl

Question. Suppose Player I provides us with the information x (either 1, 2, or 3). Which probabilities $q_x(\eta)$ should be assigned to the theoretical possibilities η for y?

Preliminary exploration. If $x = 1$, then the theoretical possibilities are 1 and 5 and nobody will criticize the specification $q_1(1) = q_1(5) = 1/2$ because the die is fair. If $x = 2$ we are in trouble because if we assume that Player I chooses $x = 2$ in case $y = 6$ (Strategy 1) then $q_2(2) = q_2(4) = q_2(6) = 1/3$ is logical, but if we assume that he would choose $x = 3$ (Strategy 0) then $x = 2$ implies that $y = 6$ is impossible and that $q_2(2) = q_2(4) = 1/2$ is appropriate. If $x = 3$, a similar discussion can be made.

2 Background information

The situation now obtained is characteristic for almost everything in statistics: the solution depends on an unknown true value, here strategy number $i \in \{0, 1\}$. We shall discuss a logical approach, a Bayesian approach and we give a preview of the Fisher–Neyman–Pearson–Wald approach we recommend. This section will be concluded with some additional comments.

The situation at the end of Section 1 is such that it is logical to specify that, e.g., $q_2(6) \in \{0, 1/3\}$ or, more precisely, that $q_2(6)$ is equal to 0 if $i = 0$ and equal to $1/3$ if $i = 1$. The latter replacement is logically valid, but useless because we do not know whether $i = 0$ or $i = 1$. (See, however, the end of this section.) Simply stating that $q_2(6)$ is either 0 or $1/3$ is a possibility which is in line with some theories in logic. The interesting book CLEAVE (1991) starts with Carnap's statement that, in logic, the science of valid inference, there are no morals. Indeed, if one refuses to choose between 0 and $1/3$ (or inbetween) then no moral principles will be needed. The statistician, however, will accept the task of specifying a number $q_2(6)$ at least if a reasonable compromise is possible.

The Bayesians adhere to the perspective that Player I chooses Strategy 1 with probability ρ. As a consequence $q_2(6)$ will be chosen equal to $P(Y = 6|X = 2) = \rho/(2 + \rho)$. The question now is '*which* ρ?'. In this problem, the choice $\rho = 1/2$ cannot be defended on the basis of symmetry arguments. The Bayesian might continue by simply using $\rho = 1/2$ because this is in the middle of the interval [0, 1] of possibilities, or he might continue by simply using ρ itself as the outcome of a uniformly distributed random variable. Such constructions are not necessarily reasonable. This is why we abandon the Bayesian approach and start from scratch in Section 3 with a 'classical' statistical approach where i is regarded as an unknown number. In Section 5 we accept the idea that i is the outcome of a random variable such that the possibility 1 has probability ρ. The discussion leads to the conclusion that this does not help much if one knows nothing about the true value of ρ.

The latter situation is not very realistic in the sense that if one knows that i is the outcome of a random variable then one will usually also have some information about ρ. Anyway, computations lead to $\rho = 0.495$ (and, hence $q_2(6) = 0.198$) as the Bayesian solution, which corresponds to the minimax regret procedure. The minimax risk procedure is characterized by $\rho = 0.667$ (and $q_2(6) = 0.250$). A referee remarked that, without moral information, the choice of principle cannot be discussed. We agree to a certain extent: many statisticians have experienced the Wald's minimax risk principle is often too conservative and that it may even lead to degenerative results such that, e.g., the observations are completely ignored (see e.g. SCHAAFSMA, 1969). Such experiences suggest that the choice of ρ 'should be' closer to 0.495 than to 0.667.

The reader may wonder why so much attention is paid to such an elementary and impractical example. The reason is that we are interested in the foundations of predictive (distributional) inference. The problem considered is a very simple example of a problem where a predictive inference is required. The problem is related to the well known quiz-master problem (see e.g. KOOI, 1999). The earliest reference we found was SELVIN, 1975) and to the prisoner's dilemma (according to RAPOPORT, 1974, the first appearance of this problem was in 1952 by Flood). An advantage of our problem is that it is less vexatious and more in line with what we have to do in mathematical statistics at large and in distributional inference in particular.

Note that the whole perspective may change drastically if additional information is provided. If e.g. Player II observes that Player I flips a coin before issuing the statement $x = 2$ then he should *not* conclude that $\rho = 1/2$ and, hence $q_2(6) = \frac{1}{2}/(2 + \frac{1}{2}) = 0.20$ is appropriate. It is logical to conclude that Player I has seen a six, and hence $q_2(6) = 1$, because in any other case it makes no sense to flip a coin.

With these preliminarities in mind, the reader will, hopefully, appreciate the following discussions illustrating *(i)* Fisher's desire to create an inductive logic, *(ii)* Popper's statement that *induction is a myth*, *(iii)* the fact from life that *induction is a must*. Epistêmê (infallible knowledge about the universe) is beyond reach but we can do our best in providing 'approximations'.

3 A Fisher–Neyman–Pearson–Wald approach

Fisher proposed methods of inference while Neyman and Pearson, and especially Wald, tried to make comparative analyses of such methods in the hope that some method comes out best (from certain perspectives). The first and very essential step in this approach is to shift the attention from the concrete situation with a given statement x (either 1, 2 or 3) to that where probabilities $q_\xi(\eta)$ have to be assigned for each value ξ a priori possible for x. It seems reasonable to restrict the attention to the class

$$\mathscr{D} = \{Q_{a,b} | \tfrac{1}{3} \leqslant a \leqslant \tfrac{1}{2}, \tfrac{1}{2} \leqslant b \leqslant 1\}$$

of procedures $Q_{a,b}$ defined in Table 2. One of the arguments behind this restriction is that in a Bayesian context (Strategy 1 is chosen with probability ρ) the posterior probabilities $\mathrm{P}(Y = \eta | X = \xi)$ are of this kind, with $a = 1/(2 + \rho)$ and $b = 1/$

Table 2. Procedures $Q_{a,b}$

$q_\xi(\eta)$	$\eta = 1$	$\eta = 2$	$\eta = 3$	$\eta = 4$	$\eta = 5$	$\eta = 6$
$\xi = 1$	1/2	0	0	0	1/2	0
$\xi = 2$	0	a	0	a	0	$1 - 2a$
$\xi = 3$	0	0	b	0	0	$1 - b$

$(2 - \rho)$. In this section we do *not* make the assumption that i is the outcome of a random variable.

Note that simply choosing $q_2(2) = q_2(4) = q_2(6) = 1/3$ and $q_3(3) = q_3(6) = 1/2$ is a possibility, but not a clever one. It corresponds to $Q_{\frac{1}{3},\frac{1}{2}}$ and is represented by the left-lower point A of the rectancle in the left graph of Figure 1. If Player I is known to act according to Strategy 1 then $Q_1 = Q_{\frac{1}{3},1}$ is the procedure to choose. It is represented as the left-upper point of the rectangle. If he would choose Strategy 0 then $Q_0 = Q_{\frac{1}{2},\frac{1}{2}}$ (the right-lower point) is 'logically valid' from the probabilistic viewpoint. As we are unaware of Player I's strategy and yet are forced to assign probabilities, a *compromise* will be needed. Averaging the parameters of Q_0 and Q_1 we obtain $Q_{\frac{5}{12},\frac{3}{4}}$ (point E). This solution is not satisfactory from an intellectual viewpoint: it is like cutting a Gordian knot without examining it. To improve this situation, we adopt the perspective of the theory of statistical decision functions. In

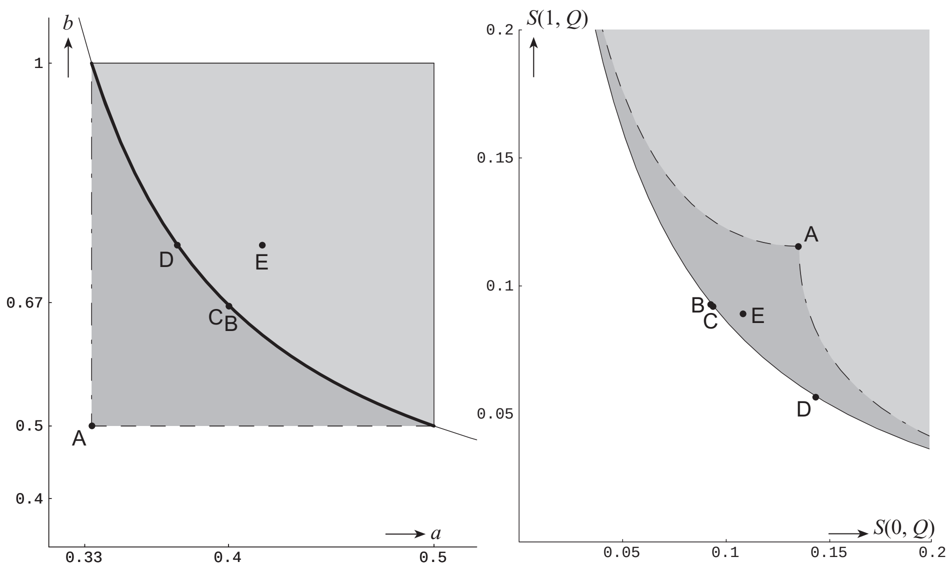

Fig. 1. Visualization of procedures in the (a, b)-plane (left) and the $(S(0, Q), S(1, Q))$-plane (right). The solid curves correspond to the procedures Q_ρ. Only the part inside the box in the figure to the left deserves consideration. The dashed lines correspond to procedures $Q_{a,\frac{1}{2}}$, the dot-dashed ones to $Q_{\frac{1}{3},b}$. Explanation of the points A, B, C, D and E will be given in the text, especially at the end of Section 3.

its usual form this theory tries to prescribe how statements should be made about true values of unknown parameters (here strategy number i). Our situation is different in the sense that a predictive statement is required, namely about the true value of some random variable Y. The distributional form of inference is more natural if such predictive inference is required than in the classical situation. On the other hand, it is more complicated to arrive at satisfactory results (unless a Bayesian approach is adopted). We for example, do not know whether the fundamental Wald–Lehmann minimal-complete class theorem is valid in predictive inference at large. In our special example it will follow from the concrete analysis that the minimal complete class corresponds to the class of all Bayes procedures. We, however, doubt whether this holds in general. Anyway, given the observation x, we have to choose a probability distribution $Q = Q(x)$ on $\{1, \ldots, 6\}$ with probabilities $q_x(1), \ldots, q_x(6)$ and think in terms of the loss $L(y, Q)$ to be incurred if the true value y is revealed. An important requirement is that the loss is *proper*: if y is the outcome of a random variable Y with its probabilities $p(\eta) = \mathrm{P}(Y = \eta)$ known, then L is said to be proper if

$$\mathbf{E}L(Y, Q) = \sum_{\eta=1}^{6} L(\eta, Q)p(\eta)$$

is minimum as a function of Q if the corresponding probabilities satisfy $q(\eta) = p(\eta)$.

In this paper the elaborations are restricted to the logarithmic loss function

$$L(y, P(x)) = -\log\{q_x(y)\}$$

The properness of this loss function is an immediate consequence of the positiveness of the Kullback–Leibler information number because if P denotes the true distribution of Y with $\mathrm{P}(Y = \eta) = p(\eta)$, and Q is any other distribution with $Q(\{\eta\}) = q(\eta)$, then

$$\mathbf{E}L(Y, Q) - \mathbf{E}L(Y, P) = \sum_{\eta=1}^{6} \log\left\{\frac{p(\eta)}{q(\eta)}\right\} p(\eta)$$

is positive if $P \neq Q$ (GOOD, 1952). Using this loss function we can determine the risk (expected loss) of the procedure $Q_{a,b}$. As the distribution of (X, Y) depends on whether Player I chooses Strategy 0 (saying $x = 3$ if $y = 6$) or Strategy 1 (saying $x = 2$), there are two distributions involved. They are given in Table 3. We shall have to consider the risks (expected losses):

$$R(0, Q_{a,b}) = -\tfrac{1}{6}\log\{(\tfrac{1}{2})^2 a^2 b(1-b)\} \tag{1}$$

$$R(1, Q_{a,b}) = -\tfrac{1}{6}\log\{(\tfrac{1}{2})^2 a^2 b(1-2a)\}, \tag{2}$$

obtained from Tables 2 and 3. $R(0, Q_{a,b})$ is minimum if both a^2 and $b(1-b)$ ($a \in [\frac{1}{3}, \frac{1}{2}]$, $b \in [\frac{1}{2}, 1]$) are maximum, i.e. if $Q_0 = Q_{\frac{1}{2},\frac{1}{2}}$ is used. Similarly $R(1, Q_{a,b})$ is minimum if $Q_1 = Q_{\frac{1}{3},1}$ is used. The fact that the minimization of these risks as a function of a can be separated from the minimization as a function of b indicates that

Table 3. Distributions P_i

$P_i(X = \xi, Y = \eta)$	$\eta = 1$	$\eta = 2$	$\eta = 3$	$\eta = 4$	$\eta = 5$	$\eta = 6$
$\xi = 1$	$1/6$	0	0	0	$1/6$	0
$\xi = 2$	0	$1/6$	0	$1/6$	0	$i/6$
$\xi = 3$	0	0	$1/6$	0	0	$(1 - i)/6$

our example is not representative for predictive inference at large. The minimum risk achieved by using the best procedure is called the envelope risk. It depends on i and is given by

$$R^*(0) = R(0, Q_0) = \log 2 \tag{3}$$

$$R^*(1) = R(1, Q_1) = \tfrac{1}{3}\log 2 + \tfrac{1}{2}\log 3. \tag{4}$$

The risk of any procedure $Q_{a,b}$ is larger than the envelope risk. The difference is referred to as the regret or shortcoming. Using $S(i, Q_{a,b})$ as notation, we have

$$S(0, Q_{a,b}) = R(0, Q_{a,b}) - R^*(0) = -\tfrac{1}{6}\log\{16a^2b(1 - b)\}$$

$$S(1, Q_{a,b}) = R(1, Q_{a,b}) - R^*(1) = -\tfrac{1}{6}\log\{27a^2b(1 - 2a)\}$$

Note that $S(0, Q_{a,b})$ is a decreasing function of a and an increasing function of b if $(a, b) \in [\frac{1}{3}, \frac{1}{2}] \times [\frac{1}{2}, 1]$. For $S(1, Q_{a,b})$ the situation is reversed. In our example, the minimal complete class of procedures corresponds to the class of Bayes procedures. This class can be obtained by minimizing the convex combination

$$\begin{aligned}(1 - \rho)S(0, Q_{a,b}) &+ \rho S1, Q_{a,b}) \\ &= -\tfrac{1}{6}[2\log a + \rho\log(1 - 2a) + \log b + (1 - \rho)\log(1 - b) \\ &\qquad + (1 - \rho)\log 2^4 + \rho\log 3^3]\end{aligned}$$

of both shortcomings. With elementary calculus it can be seen that this linear combination is minimum as a function of a and b if $a = (2 + \rho)^{-1}$ and $b = (2 - \rho)^{-1}$. This is expressed by $b = a/(4a - 1)$ and by the boldfaced curve in Figure 1 (left). Henceforth we use the notation

$$Q_\rho = Q_{\frac{1}{2+\rho},\frac{1}{2-\rho}}$$

to denote the procedure of the form $Q_{a,b}$ which minimizes $(1 - \rho)S(0, Q_{a,b}) + \rho S(1, Q_{a,b})$. Note that $Q_0 = Q_{\frac{1}{2},\frac{1}{2}}$ and $Q_1 = Q_{\frac{1}{3},1}$ are as before. Looking for a compromise $Q_{a,b}$ between Q_0 and Q_1, one should not go (too far) beyond the boldfaced curve in Figure 1 (left) which characterizes the procedures which are 'admissible' in the sense of Wald's theory. For the general case we have

$$S(0, Q_\rho) = -\tfrac{1}{6}\log\left\{16\frac{1 - \rho}{(2 + \rho)^2(2 - \rho)^2}\right\}$$

$$S(1,\, Q_\rho) = -\tfrac{1}{6}\log\left\{27\frac{\rho}{(2+\rho)^3(2-\rho)}\right\}.$$

The corresponding points constitute the left-lower bound of the regret set

$$\mathscr{S} = \{(S(0,\, Q_{a,b}),\, S(1,\, Q_{a,b}))|\tfrac{1}{3} \leqslant a \leqslant \tfrac{1}{2}, \tfrac{1}{2} \leqslant b \leqslant 1\}$$

which is, of course, nothing but the risk set when the origin is shifted to $(R^*(0),\, R^*(1))$. In this example the minimal complete class corresponds to the class of all admissible procedures as well as to the class of all Bayes procedures.

4 Detailed discussion of Figure 1 and Table 4

Figure 1 provides visualizations of the parameters $(a,\, b)$ of the procedures $Q_{a,b}$ (left) and of the corresponding points $S((0,\, Q_{a,b}),\, S(1,\, Q_{a,b}))$ (right). Table 4 gives details about some specific points, namely the following:

Point A. The naive procedure $Q_{\frac{1}{3},\frac{1}{2}}$ has already been discussed at the beginning of the previous section. It corresponds to the idea that the probabilities $q_2(2) = q_2(4) = q_2(6) = 1/3$ have to be assigned if $x = 2$ (because 2, 4 and 6 are equiprobable if Strategy 1 is chosen) and that the probabilities $q_3(3) = q_3(6) = 1/2$ have to be assigned if $x = 3$ (because 3 and 6 are possible and equiprobable if Strategy 0 is used). The snake in the grass is that if $x = 2$, and Strategy 0 would have been chosen in the case $y = 6$, then 2, 4 and 6 are not equiprobable at all, because $y = 6$ is impossible. A similar argument holds for $x = 3$, if Strategy 1 would have been chosen in the case $y = 6$. It is very difficult for probabilists to accept that the information '*the die is fair, the number of eyes is even*' does not imply that the outcomes 2, 4 and 6 are equiprobable. In the present context they will be less unwilling to deviate from that which they regard as the foundations of Probability Theory than in, e.g., the quiz-master's problem. The reason is that we can present a precise analysis of the situation. The crux is, of course, that *the source of the information ('the number of eyes is even') has to be made part of the probabilistic model*. In the present Fisher–Neyman–Pearson–Wald approach this is done by referring to the true but unknown number i of the strategy which Player I would

Table 4. Specialization of some regret points

Procedure	Name		$S(0,\, Q)$	$S(1,\, Q)$
$Q_{\frac{1}{3},\frac{1}{2}}$	naive conditional probabilities	A	0.135	0.116
$Q_{0.495} = Q_{0.4007,0.6648}$	minimax regret	B	0.093	0.093
$Q_{\frac{1}{2}} = Q_{0.4000,0.6667}$	minimum average regret	C	0.094	0.092
$Q_{\frac{2}{3}} = Q_{\frac{3}{8},\frac{3}{4}}$	minimax risk	D	0.144	0.057
$Q_{\frac{5}{12},\frac{3}{4}}$	naive compromise	E	0.109	0.089
$Q_{\frac{2}{5}} = Q_{\frac{5}{12},\frac{5}{8}}$			0.072	0.119
$Q_{0.55} = Q_{0.3922,0.6897}$			0.107	0.080

apply if he would have been confronted by $y = 6$. It follows from the discussion that the naive procedure $Q_{\frac{1}{3},\frac{1}{2}}$ represented by A is not appropriate: a should be larger than $1/3$ and b larger than $1/2$; how much larger, that is the question.

Point B. We are attracted by the idea of minimizing the maximum regret (see Section 2). This can be achieved by looking for the Bayes procedure Q_ρ that satisfies $S(0, Q_\rho) = S(1, Q_\rho)$. The computations provide $\rho = 0.495$, etc. See Table 4 and Figure 1 (Point B) and notice that both regrets 0.093 are smaller than those of the naive procedure.

Point C. The Bayesian approach has a considerable appeal but it requires the choice of the prior probability ρ which, in the context of this section, is a fictitious construct (see Section 5 for a different possibility). We can ignore probabilistic terminology by simple stating that we want to minimize some weighted average of the risks, or, equivalently, of the regrets, $\frac{1}{2}S(0, Q_{a,b}) + \frac{1}{2}S(1, Q_{a,b})$. This is then achieved by the procedure $Q_{\frac{1}{2}}$ represented by the point C. Note that one coordinate of the corresponding regret point is larger, and one is smaller than that of B.

Point D. Wald was fascinated by the theory of games as presented in Von Neumann–Morgenstern (see WALD, 1947). This leads to minimizing the maximum risk procedure, which is obtained by equating the two risks. This provides $b = 2a$ and corresponds to point D.

Point E. At the beginning of this section we argued that a compromise will be needed and we suggested the algebraically natural candidate $Q_{\frac{5}{12},\frac{3}{4}}$ is not satisfactory from an intellectual viewpoint. Indeed, $(\frac{5}{12}, \frac{3}{4})$ is beyond the boldfaced curve in Figure 1 (left) which corresponds to the admissible procedures. This generates the task to construct a procedure with both regrets decreased. Our first try was to solve $1/(2 + \rho) = 5/12$ which provides $\rho = 2/5$, the regret $S(1, Q_{\frac{2}{5}})$ being larger than that of the naive compromise. Solving $1/(2 - \rho) = 3/4$ leads to the minimax risk procedure. Our final try was to take $\rho = 0.55$, which indeed has smaller regrets than the naive compromise.

A vexed issue at the end. The parameter ρ was used as a technical device to generate the minimal complete class $\{Q_\rho | 0 \leqslant \rho \leqslant 1\}$ of Bayes procedures. The procedure $Q_{\frac{1}{3},\frac{1}{2}}$ was much too naive in the sense that it assumes that Strategy 1 has been chosen if $x = 2$ and Strategy 0 if $x = 3$. This does not correspond to the facts. Nevertheless it cannot be denied that the outcome x contains some information with respect to Player I's choice of strategy. If we try to exploit this information by choosing ρ depending on x, then the procedure obtained may be reasonable but it will not be admissible in the sense of Wald discussed, the admissibility requires that the procedure is in the class $\{Q_\rho | 0 \leqslant \rho \leqslant 1\}$.

5 What if Player I uses a randomized strategy?

As complete consensus can obviously not be achieved, 'additional knowledge' would be welcome. In this section we restrict the attention to the idea that Player 1 chooses Strategy 1 with probability ρ, this 'physical' probability being either fully known (Situation 1) or fully unknown (Situation 2). See Section 6 for the other possibility.

Situation 1 may appear if we really are involved in a two-person zero-sum game. Knowledge of the pay-off matrix will then not only affect the choice of ρ but also the restriction to the class $\mathscr{D} = \{Q_{a,b} | \frac{1}{3} \leqslant a \leqslant \frac{1}{2}, \frac{1}{2} \leqslant b \leqslant 1\}$. This makes it clear that the 'solutions' presented at the end of Section 3 are only reasonable if further information is absent.

Henceforth the attention is concentrated on Situation 2: the randomization probability ρ will then appear as the unknown true value of the parameter $\theta \in \Theta = [0, 1]$, and the factual pair (x, y) is the outcome of a pair (X, Y) of random variables with distribution uniquely determined by ρ. As nothing is known about ρ, it is intuitively clear that the additional information is not worth much. We introduce random variables (X_θ, Y_θ) having the distribution P_θ which (X, Y) would have had, given $\rho = \theta$. If i is replaced by ρ in Table 3 then one obtains a table of $\mathrm{P}(X_\theta = \xi, Y_\theta = \eta)$ values.

The risk $-\mathbf{E}\log(q_X(Y))$ depends on the true value ρ of θ. In general we have

$$R(\theta, Q_{a,b}) = -\mathbf{E}\log\{q_{X_\theta}(Y_\theta)\}$$
$$= -\tfrac{1}{6}[\log(2^{-2}a^2b) - \theta\log(1-2a) - (1-\theta)\log(1-b)].$$

Note that this risk is equal to $(1-\theta)R(0, Q_{a,b}) + \theta R(1, Q_{a,b})$ and, hence, is given by (1) and (2) in the end points.

For fixed θ, the risk $R(\theta, Q_{a,b})$ is minimum if $(a, b) = (1/(2+\theta), b = 1/(2-\theta))$. The procedure Q_θ thus obtained is Bayes with respect to all prior distributions τ on $[0, 1]$ which have θ as their expectation. The envelope risk

$$R^*(\theta) = R(\theta, Q_\theta)$$
$$= (1-\theta)R(0, Q_\theta) + \theta R(1, Q_\theta)$$

where $R(0, Q_\theta)$ is obtained from (1) by substituting $(a, b) = (1/(2+\theta), 1/(2-\theta))$ and $R(1, Q_\theta)$ follows from (2). The envelope risk displayed in Figure 2, is such that $R^*(0)$ and $R^*(1)$ correspond to (3) and (4). The maximum is reached for $\theta = 2/3$, the minimax risk procedure. The shortcoming $S(\theta, Q_{a,b}) = R(\theta, Q_{a,b}) - R^*(\theta)$ is a convex function of θ because $R^*(\theta)$ is concave and $R(\theta, Q_{a,b})$ is linear in θ. As a consequence the minimax regret and the minimax procedure are exactly the same as in Section 3. The area under the regret function is minimized by using $Q_{\frac{1}{2}} = Q_{\frac{2}{5},\frac{2}{3}}$ which, thus, is the minimal average regret procedure. The main difference with Figure 1 is that shortcoming functions are now visualized as functions of θ.

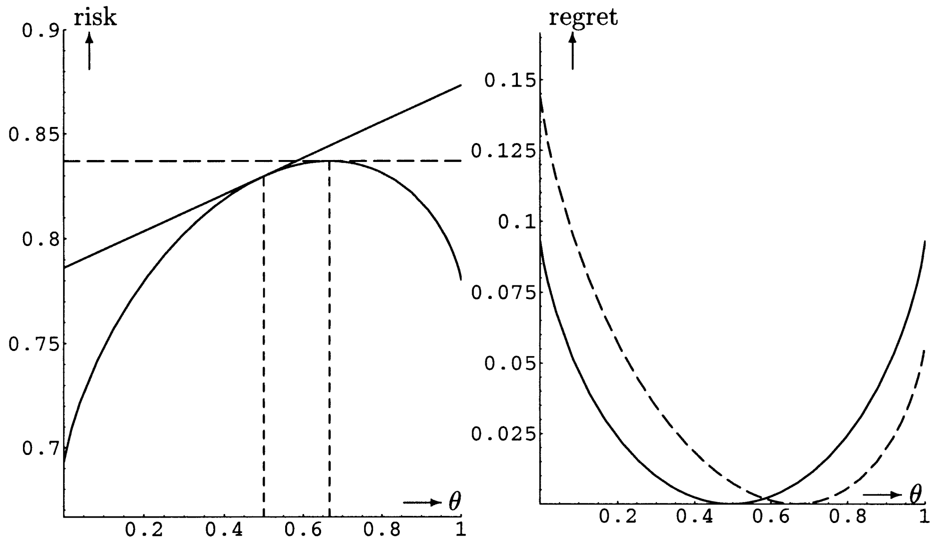

Fig. 2. The randomized strategy game. On the left: envelope risk $R^*(\theta)$ (curve), minimax risk procedure (dashed line) and minimax regret procedure (solid line). On the right, the same procedures, but now with the regret along the vertical axis.

6 Discussion

Section 5 suggests that nothing is gained if we know that Player I chooses Strategy 1 with some completely unknown probability ρ. One might argue that this state of ignorance changes as soon as the outcome x is available. The underlying random variable X assumes the values 1, 2, or 3 with probabilities $\frac{1}{3}, \frac{1}{3}+\frac{1}{6}\rho, \frac{1}{3}-\frac{1}{6}\rho$ respectively. Shouldn't this information be used to replace the a priori choice of $\rho = 1/2$ and $a = 2/5$, $b = 2/3$ by an a posteriori choice of $\rho = \frac{1}{2}, \frac{1}{2}+\epsilon, \frac{1}{2}-\epsilon$ if $x = 1, 2, 3$? In practice, Bayesian statisticians adapt their prior if it is in conflict with actual data. In the present example it is easy to see that this approach leads to a procedure that is 'inadmissible' from a theoretical viewpoint: the x-dependent choice of ρ suggested corresponds to a $Q_{a,b}$ with $(a, b) = ((\frac{5}{2}+\epsilon)^{-1}, (\frac{3}{2}+\epsilon)^{-1})$ not in the arc of admissible procedures.

In practice, the additional information will not be of the form of the two extremes ('ρ fully known' or 'ρ completely unknown') suggested in the beginning of Section 5. A confidence interval or distributional inference about ρ may be available. This will affect the conclusion that $Q_{a,b}$ with $(a, b) \approx (0.40, 0.67)$ is the rule to choose. That a compromise solution is affected if additional information is provided, is completely natural though it illustrates the hazards involved in any discussion of the types presented.

Our theory started from the class $\mathscr{D}$ of procedures and concentrated the attention on risks and regrets based on logarithmic loss. It resulted in the opinion that (a, b) should be close to the arc $\{(a, b)|(a-\frac{1}{4})(b-\frac{1}{4}) = \frac{1}{16}\}$ because these points generate the class $\{Q_\rho | 0 \leqslant \rho \leqslant 1\}$ of all Bayes procedures which corresponds to the minimal complete class. An extensive discussion suggested that the procedure to be chosen should be of the form Q_ρ where ρ is not too much different from $1/2$. The agreement among the probabilities actually assigned is such that none of these probabilities is completely compelling in the cases $x = 2$ or $x = 3$. They, however, are all very reasonable because the agreement between, e.g., $q_2(6) = 0.198$ (minimax average risk or regret) and $q_2(6) = 0.25$ (minimax risk) is quite satisfactory.

The question arises whether the agreement will be affected if the class of procedures is extended or the loss function is replaced by another one, e.g. that due to BRIER (1950) or that due to EPSTEIN (1969). The answer is easy: if the loss function is proper, then the class $\{Q_\rho | \rho : 0 \leqslant \rho \leqslant 1\}$ will not be affected, $Q_{\frac{1}{2}}$ will minimize the average risk or regret (if Section 5 is considered, integration should be with respect to the Lebesgue measure). The position of the ρ values corresponding to the minimax regret or the minimax risk procedures will become somewhat different (see ALBERS, 2000).

7 Epilogue

This paper illustrates that straightforward conditioning to incorporate empirical evidence can be misleading. This phenomenon is well-known from other problems, such as the quiz-master paradox and the prisoners' dilemma. The source of the information should be formalized and made part of the probabilistic model, which will then become 'statistical' in the sense that the unknown true value of a parameter appears. Similar issues are involved elsewhere though they often go unnoticed. The following problem is ill-posed because information is neither provided about the player's set of alternatives nor about the rules he has to obey: *Example 'A bridge player announces that his hand (of 13 cards) contains (i) an ace (that is, at least 1 ace), (ii) the ace of hearts. What is the probability that it will contain another ace?'* (PARZEN 1960, page 75).

Acknowledgement

The authors thank the referees for their helpful comments.

References

ALBERS, C. J. (2000), How to assign probabilities if you must, IWI-preprint 2000-5-04 and http://www.math.rug.nl/~casper.

BRIER, G. W. (1950), Verification of weather forecasts expressed in terms of probabilities, *Monthly Weather Review* **78**, 1–3.

CLEAVE, J. P. (1991), *Study of logics*, Oxford Science Publications.

EPSTEIN, E. S. (1989), A scoring system for probability forecasts of ranked categories, *Journal of Applied Meteorology* **8**, 985–987.

GOOD, I. J. (1952), Rational decisions, *Journal of the Royal Statistical Society* **14**, 107–114.

KOOI, B. P. (1999), *The Monty Hall dilemma*, Master's Thesis, Department of Philosophy, University of Groningen.

PARZEN, E. (1960), Modern probability theory and its applications, *John Wiley and Sons*, *New York*.

RAPOPORT, A. (ED.) (1974), *Game theory as a theory of conflict reasoning*, Reidel, Dordrecht.

SCHAAFSMA, W. (1969), Minimax risk and unbiasedness for multiple decision problems of type I, *Annals of Statistics* **5**, 1684–1720.

SELVIN, S. (1975), Letter to the editor: A problem is probability. *The American Statistician* **29**, 67.

WALD, A. (1947), Book review of J. von Neumann and O. Morgenstern: 'Theory of games and economic behaviour', Princeton University Press, 1944, *Reprinted from The Review of Economic Statistics* **24**, no. 1, 47–52.

Received: October, 1999. Revised: June 2000.