A Quantitative Measure of Relevance Based on Kelly Gambling Theory

Mathias Winther Madsen

ILLC, University of Amsterdam

Defining a good concept of relevance is a key problem in all disciplines that theorize about information, including information retrieval [3], epistemology [5], and the pragmatics of natural languages [12].

Shannon information theory [10] provides an interesting quantification of the notion of information, but it does not in itself provide any tools for distinguishing useless from useful facts. The microeconomic concept of value-of-information [1] does provide tools for doing so, but it is not easily combined with information theory, and is largely unable to exploit any of its tools or insights.

In this paper, I propose a framework that integrates information theory more natively with utility theory and thus tackles these problems. Specifically, I draw on John Kelly's application of information theory to gambling situations [7]. Kelly showed that when we take logarithmic capital growth as our measure of real utility, information theory can integrate seamlessly with classical Bernoulli gambling theory. My approach here is to turn this approach on its head and base a notion of information directly on the concept of utility.

The resulting measure coincides with Shannon information in situations in which any piece of information can be converted into a strategy improvement. When the environment provides both useful and useless information, the concept explains and quantifies the difference, and thus suggests a novel notion of value-of-information.

1 Doubling Rates and Kelly Gambling

In real gambling situations, people will often evaluate a strategy in terms of its effect on the **growth rate** of their capital, that is,

$$R = \frac{\text{Posterior capital}}{\text{Prior capital}}.$$

However, using growth rate as your utility measure suggests a gambling strategy in which you bet your entire capital on the single most likely event. Such a strategy assigns a non-zero probability to the event of losing the whole capital. If it is used in repeated plays of the same game, it thus leads to eventual bankruptcy with probability 1.

If we are instead interested in maximizing the long-term growth of a stock of capital through repeated investment and reinvestment, a better measure of strategy quality is the logarithm of the growth rate,

$$W = \log R.$$

When the logarithm is base two, this quantity is called the **doubling rate** of the capital, in analogy with the half-life of a radioactive material. W measures how many times your capital is expected to be doubled in a single game, and $1/W$ the average number of games it takes to double your capital once.

Because $\log 0^+ = -\infty$, your doubling rate will be $-\infty$ if there is even the slightest chance that you lose your whole capital by the strategy you are using. Consequently, using the doubling rate as your measure of utility will discourage strategies that can lead to bankruptcy and instead lead to a strategy that maximizes long-term exponential growth [4, ch. 6].

For the purposes of the present paper, however, the doubling rate is also interesting because of its seamless integration with Shannon information theory. To see this, consider a horse race in which horse x has probability $p(x)$ of winning. By the method of Lagrange multipliers, we can find that independently of the the odds on the horses, a gambler's doubling rate is maximized by **proportional betting** [4], i.e., betting a fraction of $p(x)$ of the total capital on horse x. If you know what these probabilities are, you thus know what the optimal strategy is.

In general, however, a gambler may have have only little or bad information about the horses, and thus use an inferior probability estimate $q \approx p$. Using the probability distribution q as a capital distribution scheme, the gambler's doubling rate will then be as follows, assuming that the odds are expressed as $c/r(x)$ for some constant c and some positive function r:

$$\begin{aligned} W(q) &= \sum_x p(x) \, \log \left(c \times \frac{q(x)}{r(x)} \right) \\ &= \sum_x p(x) \, \log \left(c \times \frac{p(x)}{r(x)} \times \frac{q(x)}{p(x)} \right) \\ &= \sum_x p(x) \, \log \left(\frac{p(x)}{r(x)} \right) - \sum_x p(x) \, \log \left(\frac{p(x)}{q(x)} \right) + \log c. \end{aligned}$$

The second term in this expression, $\sum_x p(x) \, \log (p(x)/r(x))$, is the **Kullback-Leibler divergence** [8] between p and r, and is also written $D(p \,||\, r)$. It is an measure of how big an error the probability estimate r induces in an environment with actual probabilities p. Similarly, the second term is $D(p \,||\, q)$, the divergence from p to q.

It thus turns out that the bookmaker and the gambler are in a symmetric situation: Both the distribution of bets (q) and the size of the odds (r) implicitly express subjective probability estimates. The payoffs for the gambler and the bookmaker are determined by the quality of these estimates.

In particular, if $c = 1$, the player with the probability estimate closest to p in informational terms will make money at the expense of the other. Further, if one of the two players acquire 1 bit of information about the real winner of the race, this signal can be converted into an increase of 1 capital doubling per game. In the horse race model, information thus translates directly into utility.

However, this correspondence rests on assumptions that are particular to the horse race model, including the fact that the situation involves only one random

variable, and that the gambler can bet on this variable without any restrictions on the capital distribution. In more complex and more realistic situations, an agent's representation of the environment may contain more variables, and not all of them will provide opportunities for capital growth in the same way.

In fact, even if a random variable affects an agent's payoff very strongly, it can easily be the case that it contains no useful information. For instance, if you don't own an umbrella, your optimal strategy might be the same before and after you learn that it's raining, even if your expected utility drops drastically.

The suggestion I want to make is that we take the notions of utility and strategy as primitives and derive a notion of relevance from those. This contrasts with Shannon information theory, which defines information independently of the agents using that information. It also contrasts with arithmetic value-of-information in using a logarithmic target statistic, rather than the nominal size of the capital. Both of these assumptions lead to a number of unique features which are illustrated by several examples in section 3.

2 Expected Relevant Information and Relevance Rates

Relevant information is a notion of information defined in terms of utility. The notion of utility itself only makes sense in the context of agents faced with choices, so I first need to define a notion of a decision problem.

Definition 1. *A* ***decision problem*** $D = (S, \Omega, p, u)$ *consists of a strategy set* S*, a sample space* Ω*, a probability measure* $p : \Omega \to \mathbb{R}$*, and a utility function* $u : S \times \Omega \to \mathbb{R}$*. When* u *is bounded and non-negative, we further define the* ***(expected) doubling rate*** *of the strategy* s *as*

$$W(s) = \int p(x) \log u(s, x)\, dx$$

$W^* = \sup_s W(s)$ *is the* ***optimal (expected) doubling rate*** *of the decision problem.*

Definition 2. *Let a decision problem* $D = (S, \Omega, p, u)$ *be given as above, and let* Y *be a stochastic variable. Then the* ***posterior decision problem given the event*** $Y = y$ *is* $D' = (S, \Omega, p', u)$*, where* $p'(x) = p(x \mid y)$*. The amount of* ***relevant information*** *in* $Y = y$ *is the increase in optimal doubling rate,*

$$K(y) = \sup_{s \in S} W'(s) - \sup_{s \in S} W(s),$$

where W' *is the doubling rate in* D'*. Further,* $K(Y) = E\,[K(y)]$ *is the* ***expected relevant information*** *contained in* Y*.*

Theorem 1. *Expected relevant information is non-negative.*

Proof. With respect to the marginal distribution of Y, we have the expectations

$$\begin{aligned}
E\left[\sup_{s\in S} W'(s)\right] &= \int p(y)\left(\sup_{s\in S}\int p(x\,|\,y)\,\log u(s,x)\,dx\right)\,dy \\
&\geq \sup_{s\in S}\left(\int p(y)p(x\,|\,y)\,\log u(s,x)\,dx\,dy\right) \\
&= \sup_{s\in S}\left(\int p(x)\,\log u(s,x)\,dx\right) \\
&= \sup_{s\in S} W(s) \\
&= E\left[\sup_{s\in S} W(s)\right],
\end{aligned}$$

where the last equality follows from the fact that W does not depend on Y. The expected posterior doubling rate is thus higher than the expected prior, and so

$$E\left[K(Y)\right] = E\left[\sup_{s\in S} W'(s) - \sup_{s\in S} W(s)\right] = E\left[\sup_{s\in S} W'(s)\right] - E\left[\sup_{s\in S} W(s)\right]$$

is non-negative.

This proposition closely mirrors the well-known fact that Shannon information content is non-negative on average. So while bad news may occasionally represent a setback, on average, information cannot hurt you. Notice also that the proof can be read as saying that an irrationally risk-averse agent can secure an unchanged average doubling rate by ignoring all incoming information.

Theorem 2. $1 - 2^{-K(Y)}$ *is the greatest fraction of future capital that an agent can trade for the value of* Y *without expected loss.*

Proof. Let $D = (S, \Omega, p, u)$ be the original decision problem, and let its prior and posterior doubling rates be W and W', respectively. Trading a fraction f of your future capital for the value of Y will modify this problem so that the utility function in the posterior decision problem is downscaled by a factor of $1 - f$.

Let $K'(Y)$ denote the expected amount of relevant information contained in Y in this modified problem. We then have

$$\begin{aligned}
K'(Y) &= \int p(y)\left(\int p(x|y)\,\log\left((1-f)u(s,x)\right)\,dx\right)dy \\
&= \int p(y)\left(\int p(x|y)\,\log u(s,x)\,dx\right)dy \;+\; \log(1-f) \\
&= K(Y) \;+\; \log(1-f).
\end{aligned}$$

The agent can thus expect an on-average loss in the modified problem if and only if $K(Y) + \log(1-f) < 0$. Taking powers of 2 on both sides if this inequality and rearranging the terms gives the desired result.

To distinguish relevant information from Shannon information in the usual sense, we further define a concept of "raw" information:

Definition 3. *Let p be a probability measure on Ω, and let X be a random variable distributed according to p. For any random variable Y, the expected amount of **raw information** contained in Y is*

$$G(Y) = I(X;Y) = H(X) - H(X \mid Y).$$

Raw information is thus not a measure of dependence between two random variables in particular, but rather a measure of global decrease in uncertainty. Any source of uncertainty you have in your environment is a potential source of raw information, but not necessarily of relevant information.

These two measures of information suggest a natural measure of relevance:

Definition 4. *Let $D = (S, \Omega, p, u)$ be a decision problem, and Y a random variable on Ω. Then the **relevance rate** of Y is $K(Y)/G(Y)$.*

In general, the relevance rate of a random variable can be both larger than and smaller than 1. However, the following theorem shows that the two coincide when an agent can bet with fair odds on the outcome of any random event whatsoever:

Theorem 3. *Let $D = (S, \Omega, p, u)$ be a decision problem in which the strategy space is the set of probability distributions on Ω, and Y is a random variable on Ω. Suppose further that the utility function u has the form $u(s,x) = s(x)v(x)$ for some non-negative, real-valued function v. Then $K(Y) = G(Y)$.*

Proof. This observation is due to Kelly [7]. We prove it by noting that a utility function of the form $u(s,x) = s(x)v(x)$ leads to a doubling rate of the form

$$\begin{aligned} W(s) &= \int p(x) \log s(x)v(x)\, dx \\ &= \int p(x) \log \left(\frac{p(x)}{b(x)} \right) dx - \int p(x) \log \left(\frac{p(x)}{s(x)} \right) dx \\ &= D(p \,||\, b) - D(p \,||\, s), \end{aligned}$$

where $b = 1/v$ can be interpreted as the odds placed on horse x being the winner of a race, and $s(x)$ the bet placed on that horse.

As a consequence of Jensen's inequality, the unique minimum of $D(p \,||\, s)$ is $s = p$. The doubling rate is thus maximized by proportional betting ($s = p$), regardless of the probability environment and the odds. The optimal doubling rate under the distribution p is thus

$$\begin{aligned} \sup_{s \in S} W(s) &= D(p \,||\, b) - D(p \,||\, p) \\ &= D(p \,||\, b) \\ &= \int p(x) \log \left(\frac{p(x)}{b(x)} \right) dx. \end{aligned}$$

Similarly, the optimal doubling rate given the condition $Y = y$ is

$$\sup_{s \in S} W'(s) = \int p(x|y) \log \left(\frac{p(x|y)}{b(x)} \right) dx$$

Taking expectations with respect to Y and subtracting the prior doubling rate from the posterior, we find

$$\begin{aligned} K(Y) &= E\left[\sup_{s \in S} W'(s)\right] - E\left[\sup_{s \in S} W(s)\right] \\ &= E\left[\int p(x|y) \log p(x|y)\, dx\right] - E\left[\int p(x) \log p(x)\, dx\right] \\ &= -H(X|Y) - (-H(X)) \\ &= G(X). \end{aligned}$$

This establishes the desired equality.

3 Examples of Relevance Measurements

Having now introduced the notion of relevance, I would like present a series of examples that illustrate the use and utility of the concept.

Code-breaking with Optional Investment Suppose you can invest any fraction of your capital in a lottery defined as follows: If you can guess the four binary digits of my credit card code in a single try, you get your investment back 16-fold; otherwise, you lose it.

Suppose you invest a fraction p of your capital in this lottery and keep a fraction of $1 - p$ in your pocket. This will give you an expected doubling rate of

$$E\left[W(p)\right] = \left(\frac{1}{2^4}\right) \log(1 - p + 16p) + \left(1 - \frac{1}{2^4}\right) \log(1 - p).$$

This is a decreasing function in p, and your optimal strategy is $p^* = 0$, i.e., not putting any of your money into the lottery.

However, suppose that you have an inside source that can supply you with some of the digits of the credit card code. For each digit you receive, your chance of guessing the code in your first attempt obviously increases, leading to a doubling rate of

$$E\left[W_i(p)\right] = \left(\frac{2^i}{2^4}\right) \log(1 + 15p) + \left(1 - \frac{2^i}{2^4}\right) \log(1 - p)$$

after you have received i of the four digits. This function attains its maximum on the unit interval at $p^* = {}^0\!/_{15}, {}^1\!/_{15}, {}^3\!/_{15}, {}^7\!/_{15}, {}^{15}\!/_{15}$ for $i = 0, 1, 2, 3, 4$, as illustrated in Figure 1. The optimal expected doubling rates in these cases are

$$W_i^* \;=\; 0.00,\ 0.04,\ 0.26,\ 1.05,\ 4.00, \qquad \text{for } i = 0, 1, 2, 3, 4.$$

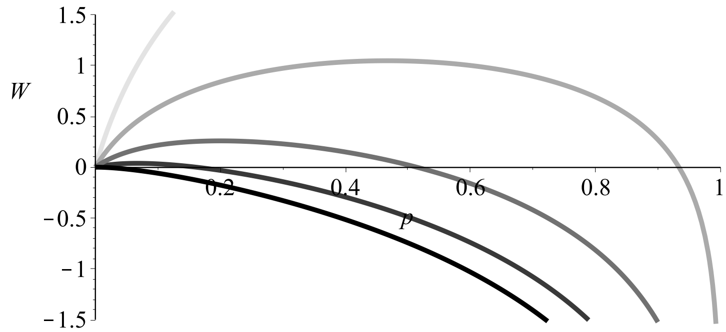


Fig. 1. Doubling rate as a function of investment level in the code-breaking example. Brighter gray lines represent doubling rates after more messages have been received.

It thus turns out that the four digits you receive are not equally relevant to you. The first contains only 0.04 bits of relevant information; further, since it decreases your uncertainty about the credit card code by exactly one bit, its relevance rate is 0.04 bits of relevant information per bit of raw information. The second contains 0.22 bits of relevant information per bit of raw information, and the third 0.79, and the fourth 2.95.

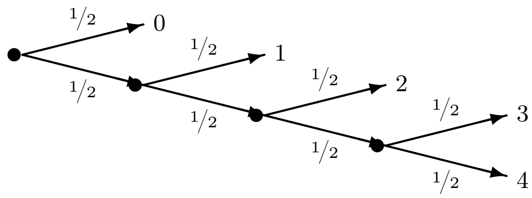
Notice that this is only the case because you are not forced to invest all your money in the lottery. If you were, all four bits would supply you with exactly one bit of relevant information, giving them a relevance rate of 1.

Code-breaking with Irrelevant Side-Information Suppose we are in the exact same code-guessing scenario as above, but that you now receive your side-information about my credit card code from an unreliable source which may abort the communication at any time. In this case, you have uncertainty about two independent variables, my actual credit card code (C), and how many characters you will receive (L).

Since receiving a character removes uncertainty not only about C, but also about L, you will, paradoxically, receive more than one bit of raw information per transmitted character under these assumptions. The amount of relevant information you receive, however, will remain the same.

For instance, suppose that just before transmitting each bit, your source flips a coin to decide whether to continue or abort. This means that L takes the five values 0, 1, 2, 3, and 4 with probabilities $1/2$, $1/4$, $1/8$, $1/16$, and $1/16$, respectively (cf. Fig. 2). Excluding one of those possible outcomes at a time, beginning from the left, gives the entropy values in Table 2.

Computing the differences between the five entropy levels, we find that the four characters you receive contain $9/8$, $5/4$, $3/2$, and 2 bits of information, respectively. However, the optimality of a strategy only depends on your chance of guessing the code in a single try, the amount of relevant information contained in the messages remains as in the previous example. This example thus illus-



i	0	1	2	3	4
$H(C)$	4	3	2	1	0
$H(L)$	$^{15}/_{8}$	$^{7}/_{4}$	$^{3}/_{2}$	1	0
Sum	$^{47}/_{8}$	$^{19}/_{4}$	$^{7}/_{2}$	2	0

Fig. 2. Left, a probability distribution on L, the number of digits you receive from the unreliable source in example 3. Right, the decreasing uncertainty about the code (C) and transmission length (L) after i digits of my credit card code have been revealed.

trates how the addition of marginally relevant variables to a model can change the information-theoretic analysis of a situation, but not the analysis in terms of relevant information.

Randomization Suppose the two of us put down \$1 for a game of Rock-Paper-Scissors, and that the winner gets both dollar bills. If we label our strategies $p = (p_1, p_2, p_3)$ and $q = (q_1, q_2, q_3)$, then your payoff is

$$u_1(p, q) = q_1(p_1 + 2p_2) + q_2(p_2 + 2p_3) + q_3(p_3 + 2p_1).$$

My payoff is $u_2 = 2 - u_1$.

Whatever your strategy is, one of my three pure strategies will be a best response. Consequently, if I adapt my strategy q^* to yours, your payoff is

$$u_1(p, q^*) = \min\{p_1 + 2p_2,\ p_2 + 2p_3,\ p_3 + 2p_1\}.$$

Your doubling rate is thus the logarithm of one the smallest of these three expressions, and it is optimal when you use a uniform distribution on $\{R, P, S\}$.

This is essentially all there is to say about the situation from a game-theoretical perspective. However, as Claude Shannon noted in the early 1950s [13], real people are in fact curiously bad at producing random numbers, and they invariably introduce computable structure into the the "random" sequences they produce. This means that a computer (or a statistician) with even a simple pattern recognition algorithm often outperforms humans vastly in randomization games such as Matching Pennies or Rock-Paper-Scissors.

The purpose of this example is to present a model of this limitation, and to measure how much it would change the situation if people had access to external resources they could use to introduce additional randomness into their choices. I therefore analyze the most extreme case of this situation, namely that of a purely deterministic device that plays Rock-Paper-Scissors using a finite number of calls to a randomization oracle.

Suppose therefore that you have to play Rock-Paper-Scissors by submitting a publicly accessible program for a Turing machine. Since the program is completely deterministic, your strategy is going to be completely predictable, and

your opponent can adapt perfectly to your strategy. This leads to a doubling rate of $\log \min\{0, 1, 2\} = -\infty$.

However, suppose now that your Turing machine has a module which can request a fixed number of fair coin flips per game. You can then encode these coin flips into the strategy in order to make it less predictable. The optimal way to do this is to feed the coin flips into an arithmetic decoder [9,4] which translates them into a distribution on $\{R, P, S\}$. The more coin flips you have, the closer this distribution can get to the uniform distribution.

i	0	1	2	3	4	5	6	7	$\cdots$	∞
p_1	$1/1$	$1/2$	$2/4$	$3/8$	$6/16$	$11/32$	$22/64$	$43/128$	$\cdots$	$1/3$
p_2	–	$1/2$	$1/4$	$3/8$	$5/16$	$11/32$	$21/64$	$43/128$	$\cdots$	$1/3$
p_3	–	–	$1/4$	$2/8$	$5/16$	$10/32$	$21/64$	$42/128$	$\cdots$	$1/3$
u_1	0	$1/2$	$3/4$	$7/8$	$5/16$	$31/32$	$63/64$	$127/128$	$\cdots$	1

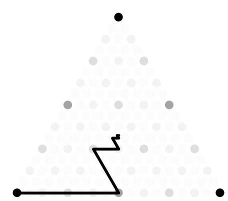


Table 1. Payoffs for increasingly randomized rock-paper-scissors strategies (left), and a graphical representation of the approximation process (right).

This situation is depicted in Table 1, which shows the arithmetic payoffs that you can achieve with i calls to the coin flipping module. As the table shows, the first coin flip will contain infinitely much relevant information, since it increases your arithmetic payoff from 0 to 1/2. The second contains

$$\log 3/4 - \log 1/2 = 0.59 \text{ bits of relevant information.}$$

The third, fourth, and fifth contain 0.22, 0.10, and 0.05 bits, respectively.

Readers familiar with Kelly gambling should note the difference between a horse race and the present model. In the horse race, a gambler chooses bets but does not control the probabilities, while in the present model, the reverse is true. Both models implicitly define a distance measure on the space of probability distributions on $\{R, P, S\}$.

Non-cooperative Pragmatics Following loosely the ideas from [11] and [6], suppose you regularly hire new staff from a pool of people that have taken two qualifying exams. Suppose further that the grades on these two exams, X and Y, are distributed uniformly on the set $\{1, 2, 3, \ldots, 10\}$. We can define the productivity of a hired person as units of profit per unit of salary, and we may assume that this profit rate depends on the two qualifying grades as

$$R = \frac{X + Y}{10}.$$

Hiring a person will thus in general affect your doubling rate $W = \log R$ either negatively or positively, depending on whether that person is qualified above or below a threshold of $X + Y = 10$ (cf. Fig. 3a and Table 2).

10	0.14	0.26	0.38	0.49	0.58	0.68	0.77	0.85	0.93	1.00
9	0.00	0.14	0.26	0.38	0.49	0.58	0.68	0.77	0.85	0.93
8	−0.15	0.00	0.14	0.26	0.38	0.49	0.58	0.68	0.77	0.85
7	−0.32	−0.15	0.00	0.14	0.26	0.38	0.49	0.58	0.68	0.77
6	−0.51	−0.32	−0.15	0.00	0.14	0.26	0.38	0.49	0.58	0.68
5	−0.74	−0.51	−0.32	−0.15	0.00	0.14	0.26	0.38	0.49	0.58
4	−1.00	−0.74	−0.51	−0.32	−0.15	0.00	0.14	0.26	0.38	0.49
3	−1.32	−1.00	−0.74	−0.51	−0.32	−0.15	0.00	0.14	0.26	0.38
2	−1.74	−1.32	−1.00	−0.74	−0.51	−0.32	−0.15	0.00	0.14	0.26
1	−2.32	−1.74	−1.32	−1.00	−0.74	−0.51	−0.32	−0.15	0.00	0.14
	1	2	3	4	5	6	7	8	9	10

Table 2. The doubling rates associated with different combinations of grades.

However, under the distribution assumed here, it is in fact rational to hire a person in the absence of any information about that person's skill level. This holds because your average doubling rate across the whole pool of applicants, $E\left[W\right]$, is slightly larger than 0:

$$\begin{aligned} E\left[\log \frac{X+Y}{10}\right] &= \sum_{i=1}^{10}\sum_{j=1}^{10} \Pr(X=i, Y=j) \times \left(\log \frac{i+j}{10}\right) \\ &= 15.23 \text{ millibits.} \end{aligned}$$

Hiring a randomly plucked person will thus give you an expected productivity of $E\left[2^W\right] = 2^{E[W]} = 2^{0.01523} = 1.01$ units of profit per unit of salary.

However, suppose you take a person into an interview, and that person shows you one of his or her grades. Assuming that you were shown the largest of the two grades, how much relevant information does this piece of data give you? At which grade level should you hire the applicant?

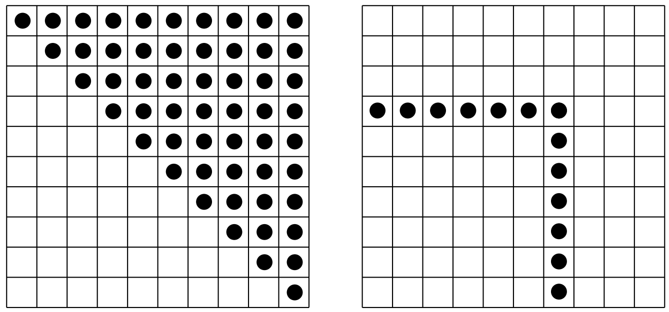
To answer this question, let $M = \max\{X, Y\}$ be the grade you were shown. The doubling rate you can expect from hiring an applicant with $M = m$ is then

$$E\left[W \mid M = m\right] = \sum_{i=1}^{10}\sum_{j=1}^{10} \Pr(X = i,\, Y = j \mid M = m) \log \frac{i+j}{10}.$$

By fixing m and summing up over all pairs (i, j) for which $i, j \leq m$ and $i = m$ or $j = m$ (cf. Fig. 3b), this doubling rate turns out to be negative for $m < 7$ and positive for $m \geq 7$. In other words, hiring a person whose largest grade is smaller than 7 will, on average, lead to a loss. The optimal decision in that case is thus to keep the salary in your pocket, remaining at an expected doubling rate of 0.

So, observing $m < 7$ leads to a doubling rate of 0. On average the expected doubling rate resulting from learning the value of M will thus be

$$\sum_{m=7}^{10} \Pr(M = m) \times E\left[W \mid M = m\right].$$



(a) (b)

Fig. 3. (a) The applicants with positive productivity. (b) The applicants whose largest grade is $M = 7$ (see also Table 2).

The probabilities in this sum are of the form $\Pr(M = m) = \frac{2m-1}{100}$, and the expected doubling rates are 0.08, 0.27, 0.44, 0.59 for $m = 7, 8, 9, 10$. A bit of computation then gives the result $E[W'] = 0.24$ bits. Since the prior doubling rate was $E[W] = 0.01523 \approx 0.02$, M contains on average

$$K(M) = 0.24 - 0.02 = 0.22 \text{ bits of relevant information.}$$

It follows that if you can observe the applicant's maximal grade before hiring him or her, your expected capital will, on average, grow by a factor of

$$R = 2^{E[W']} = 2^{0.24} = 1.18.$$

Further, you should thus be willing to trade up to $1 - 2^{-0.22} = 14.1\%$ of your future profits in return for this piece of information.

Finally, let us compute the amount of raw information contained in M. Observing the event $M = m$ narrows down the space of possible values for $X \times Y$ so that it has $2m - 1$ possible values instead of 100. Since these values are equally probable, the amount of information contained in the message $M = m$ is

$$H(X \times Y) - H(X \times Y \mid M = m) = \log 100 - \log(2m - 1).$$

To compute the average value of this quantity, we note that $M = m$ has point probabilities $\frac{2m-1}{100}$. On average, the information gain resulting from learning the value of M is thus

$$G(M) = \sum_{m=1}^{10} \left(\frac{2m-1}{100}\right) \times (\log 100 - \log(2m - 1)).$$

Computing this sum, we find that $M = \max\{X, Y\}$ contains 3.05 bits of raw information. However, as we have seen, learning its value only buys you an increase of $0.24 - 0.02 = 0.22$ bits in doubling rate on average. M thus has a relevance rate of $K(M)/G(M) = \frac{0.22}{3.05} = 0.07$ bits of relevant information per bit of raw information.

4 Conclusion

In this paper, I have proposed a logarithmic of value-of-information measure as a quantification of the concept of relevance.

This leads to an agent-oriented measure of relevance, as opposed to a system-oriented one [2]: It takes relevance to be a relation between events and agents rather than events and events. As a consequence, it forms a natural bond with decision theory, Bayesian statistics, and Shannon information theory. The proposed concept consequently represents a fairly conservative extension of the canonical calculus of reasoning which is already used in the behavioral sciences.

This new concept of relevance may also shed some new light on the ways in which dynamics of information can interact with problems of resource allocation. The examples I have given can, I believe, only be fully understood if we see the microeconomic and the information-theoretic aspects of the situation as two sides of a single coin. The concept of relevant information I have proposed here might be one out of several paths into such a style of analysis.

References

1. M. Avriel and A. C. Williams. The value of information and stochastic programming. *Operations Research*, 18(5):947–954, 1970.
2. Pia Borlund. The concept of relevance in IR. *Journal of the American Society for information Science and Technology*, 54(10):913–925, 2003.
3. William S. Cooper. A definition of relevance for information retrieval. *Information storage and retrieval*, 7(1):19–37, 1971.
4. Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-interscience, 2006.
5. Luciano Floridi. Understanding epistemic relevance. *Erkenntnis*, 69(1):69–92, 2008.
6. Jacob Glazer and Ariel Rubinstein. A study in the pragmatics of persuasion: a game theoretical approach. *Theoretical Economics*, 1(4):395–410, 2006.
7. John L. Kelly. A new interpretation of information rate. *IRE Transactions on Information Theory*, 2(3):185–189, 1956.
8. Solomon Kullback and Richard A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
9. David MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.
10. Claude E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27:379–423 and 623–656, 1948.
11. Michael Spence. Job market signaling. *The quarterly journal of Economics*, 87(3):355–374, 1973.
12. Dan Sperber and Deirdre Wilson. *Relevance: Communication and Cognition*. Blackwell Publishing, 1995.
13. Jean-Daniel Zucker and Christophe Meyer. Apprentissage pour l'anticipation de comportements de joueurs humains dans les jeux à information complète et imparfaite: les «mind-reading machines». *Revue d'intelligence artificielle*, 14(3-4):313–338, 2000.