

Real Options, Learning Measures, and Bernoulli Revelation Processes

First Version: March 20, 2003. Current Version: February 21st, 2005

PRELIMINARY: INCOMPLETE PAPER VERSION

By: Marco Antonio Guimarães Dias (*)

Abstract

This paper proposes a theory for technical uncertainty with main focus on learning real options and real option games applications. It discusses information revelation as processes of reduction of uncertainty and proposes the *expected percentage of variance reduction* as learning metric to simplify the complex analysis of investment decisions under uncertainty in which the technical uncertainty about a real asset is relevant. This learning measure has a direct link with the *distribution of conditional expectations* of a variable of interest, where the conditioning is the set of new information that can be revealed by a cost (learning option) or by waiting as free rider (option game application). In addition, it is shown that this learning measure has many favorable mathematical and practical properties. A set of axioms for (probabilistic) learning measures is presented. This paper also analyzes with some detail the simplest revelation process, namely the *sequential bivariate Bernoulli process*, including the analysis of Fréchet-Hoeffding limits and *interchangeable* bivariate Bernoulli process. Examples in exploration of petroleum and in portfolio of real assets, illustrate this methodology.

JEL classification: G31; G12;

Keywords: value of information, real options, learning measures, information revelation processes, exploratory process, bivariate Bernoulli process, correlation ratio, investment under uncertainty.

(*) D.Sc. and Senior Consultant at Petrobras. E-mail: marcoagd@pobox.com. Address: Petrobras/E&P-ENGP/RR/DPR. Av. Chile 65, sala 1702 – Rio de Janeiro, RJ, Brazil, 20035-900. Phone: +55 (21) 32242164. Fax: +55 (21) 32241579.

1 – Introduction and Literature Review

The modern theory of stochastic processes has been successfully used in real options applications in which the future value of a project or real asset is uncertain due to market uncertainty in prices, demand of a product, costs, etc. This theory is well developed in real options textbooks like Dixit & Pindyck (1994), Trigeorgis (1996), Copeland & Antikarov (2001), etc. However, in many applications (e.g. petroleum exploration, R&D) is very relevant the *technical uncertainty* about the true values of specific parameters of a project. In petroleum exploration, the oil company faces uncertainty on the *existence, volume and quality* of petroleum prospect when deciding about the wildcat (pioneer) well drilling investment. In R&D, there is uncertainty about the technical success of a specific project, e.g., a pharmaceutical firm faces uncertainty about the success of a drug test phase.

Some papers (e.g., Cortazar & Schwartz & Casassus, 2001) use stochastic processes like the geometric Brownian motion (GBM) to model technical uncertainty, which is not adequate for many reasons. One reason is that the variable changes by the simple passage of time under GBM hypothesis, whereas in most cases the variable with technical uncertainty changes only with the exercise of a learning option, so that the *filtration* for technical uncertainty is indexed by *events*¹, not by the *time*, as in almost all stochastic processes. Other reason is that the probabilistic laws for known stochastic processes are not adequate for learning processes for variables with technical uncertainty, e.g., in GBM models the variance is unbounded, whereas for learning processes this variance is bounded by the prior distribution variance (see proposition 1, section 3).

First, we must highlight that the concepts of information and uncertainty are connected. The paper focus is much in the spirit of Arrow (1973, p.138): “*When there is uncertainty, there is usually the possibility of reducing it by acquisition of information. Indeed, information is merely the negative measure of uncertainty*”. This is true mainly for technical uncertainty, e.g., generally we cannot reduce the uncertainty in oil prices by investing in information, but we can reduce the uncertainty in oil reserve volume by investing in appraisal wells, i.e., by exercising learning options.

This paper proposes a theory for the process generated by the exercise of learning options – called *revelation process*, as an adequate approach to consider the technical uncertainty issue in real options applications. We want to solve investment under uncertainty problems with opportunities to invest in

¹ The time between events can take months or years. In most cases the events are endogenous in the model (are *options*).

information or to receive new relevant information from an external source (e.g., externalities from other firm investing in information) in order to learn about technical parameter values of a real asset. In many cases, we have a compound option, the learning option(s) and the project development option, i.e., we have alternatives of investment in information in order to know better the true value of a project (e.g., a pilot project to test a technology, a market test to know better the demand function, an appraisal well) and the option to develop the project (e.g., a full scale production plant).

In order to study learning options, this section consider the classical literature on economics of information focused on *value of information* (VOI) issue. The classical theory of VOI is well summarized in Lawrence (1999), which uses the general concept of *informativeness* (θ) to handle VOI problems, showing the necessity of a *learning measure* for VOI problems. In some statistical applications, like *experimental design*, θ is the *sample size*; in classical VOI applications has been used the concept of *likelihood* of information. In this paper (section 3) is developed a very general theory for probabilistic learning measures and is proposed the concept of expected percentage of variance reduction as the best θ measure for most VOI applications. However, it is almost impossible to find out a specific learning measure applicable to *all* VOI problems. In this aspect Arrow (1984, Preface) wrote: “*it has proved difficult to frame a general theory of information as an economic commodity, because different kinds of information have no common unit that has yet been identified*”.

In order to illustrate the Arrow’s statement, let us compare two measures of uncertainty; namely the concept of *entropy* – from the *information theory* literature, and the traditional concept of *variance*. Many important researchers (like Arrow, Marschak, etc., in the 50’s) considered the entropy concept in VOI problems. Consider a univariate discrete distribution with prior $p(x)$. In Shannon sense (see Shannon & Weaver, 1949), entropy (H) is a measure of uncertainty² defined by:

$$H(X) = - \sum_i p(x_i) \log[p(x_i)]$$

Note in the above entropy definition that does not matter the scenario values, only the probabilities of different scenarios are considered. This feature has advantages, such as the simplicity to represent the

² Assume that $0 \cdot \log(0) = 0$. Note that for the full revelation case the entropy goes to zero because $\log(1) = 0$. The logarithm base in the entropy equation is arbitrary. Many texts use base 2, so that entropy is measure in “bits”.

uncertainty. But in *valuation* problems where the *magnitude* of losses or gains matters³, this feature can be a drawback or will need an additional variable to conjugate probabilities and scenario values.

Parties of the entropy measure will argue that the *uniform distribution* represents the highest degree of uncertainty over a bounded interval in the real line and it is consistent with the entropy concept, but not with the variance one. This argument can be convincing in many applications, given the popularity of the entropy measure in telecommunications, some branches of statistics, Komolgorov complexity theory, etc. However, this seducer argument is not coherent with applications in both corporate finance⁴ and in most cases from the economic value of information literature.

Consider a portfolio with uncertain returns represented by only three discrete scenarios denoted by $\Pi_R = \{-10\%; 10\%; 30\%\}$. The uniform distribution – which has the highest possible entropy, assigns probabilities $H = \{33.3\%; 33.3\%; 33.3\%\}$ for these scenarios, whereas one non-uniform distribution with higher variance could assign probabilities $V = \{45\%; 10\%; 45\%\}$. With the same expected return, risk-averse investors using expected utility theory in general and mean-variance optimizers in particular, will indicate the lower entropy (but higher variance) set V as “riskier” than H. Recall that mean-variance is a key concept in the modern *portfolio theory*, and used in the popular CAPM (capital asset pricing model) to set appropriate risk-premium.

Now, consider a real asset case. What if we consider that the value of this asset has technical uncertainty with zero correlation with the market portfolio so that it does not demand *risk-premium* from diversified investors? *Even* in this case we will prefer the set H than the set V. In this case the relevant issue is not the risk-premium demanded (zero in both H and V) but the results from the *economic optimization* under (technical) uncertainty. To see this, consider the same probabilities sets H and V, but the scenarios are the possible reserve volumes of an oilfield with $\Pi_R = \{200; 300; 400\}$, in million barrels. The optimal investment in processing plant capacity, pipeline diameter, number of wells, etc., depends on the true volume scenario of this reserve. If we must develop the oilfield project without additional information, for capacity design purposes in general we will minimize the error by choosing the expected value of this volume, equal to 300 for both distributions (H and V). However, is expected a lower economic value for this real asset with the distribution V because the

³ In communication channel design this is not problem because the scenarios are messages, not numerical amounts.

⁴ There are some adepts of the so-called *log-optimal* portfolio approach, which uses the concept of entropy. But the next example and mainly the Paul Samuelson criticism in the 70's make difficult to consider this alternative portfolio theory.

probability of either *over investment* (if the true scenario is 200) or *under investment* (if the true scenario is 400) is much higher with V than with H. For the other side of coin – the side of the opportunity to invest in information, the set V in general will provide a higher value of information than the set H, again due to the higher probability with the former of either over or under investment without knowledge of the true scenario.

So, the concept of variance as measure of uncertainty is in general more useful than the concept of entropy by the point of view of corporate finance and economics of information. Based in the concept of variance, sections 2 and 3 advocate that a learning measure based in the variance concept, namely the (relative) *expected variance reduction* (denoted by η^2 , we will see later the reason for this nomenclature), has nice properties and it is difficult to get a better measure for the classes of applications we are interested.

This paper is divided as follow. Section 2 presents a simple example in petroleum exploration economics that highlights the necessity of learning measures for both real options and option games applications. Section 3 presents the main results with propositions on revelation distributions, the properties of η^2 (the proposed learning measure), the learning decomposition theorem, axioms for learning measures, the concept of flexible information structures, and the potential applications for these concepts. Section 4 shows the case of Bernoulli revelation processes, very important in practical applications (e.g., petroleum exploration and R&D) and in theoretical studies (it is the simplest revelation process), including the Fréchet-Hoeffding bounds for a learning process, the measure η^2 in this Bernoulli context, and the simplification with interchangeable Bernoulli variables. Section 5 presents an example of a portfolio with dependent real assets (exploratory prospects) and how the Bernoulli revelation process can be used to evaluate this portfolio. Section 6 presents the concluding remarks and some suggestions for further research.

2 – A Simple Motivating Example

Here and in most of this paper, consider only technical uncertainty and investment *optionality*, for expositional clarity. However, in section 5 is presented a case in which technical uncertainty interacts with market uncertainty. In this section, to motivate to the necessity of a learning measure in both real options and option games applications, consider the following example (based in Dias, 2004). An

oil company owns the rights over a tract with two exploratory prospects. For each prospect, the value of the drilling option exercise is the *expected monetary value* (EMV)⁵, given by:

$$\text{EMV} = -I_w + [\text{CF} \cdot \text{NPV}] \quad (1)$$

Where I_w is the drilling investment in the wildcat well (option exercise price), CF is the *chance factor* about the existence of an oilfield (detailed below), and NPV is (conditional to exploratory success) the net present value of the *oilfield development*⁶. The chance factor is the parameter with technical uncertainty and is the simplest probability distribution: a Bernoulli distribution, which has two scenarios (1 = success and 0 = failure) and one parameter (p) named success probability. So, we use $\text{CF} \sim \text{Be}(p)$ to denote this Bernoulli distribution. Consider that the two prospects are symmetric, i.e., they have the same parameters and so the same EMV. Assume the numerical values $I_w = 30$ million \$, $E[\text{CF}] = p = 30\%$ and $\text{NPV} = 95$ million \$ for both prospects. So, the EMV is negative:

$$\text{EMV}_1 = \text{EMV}_2 = -30 + [0.3 \times 95] = -1.5 \text{ million \$}$$

Apparently this two real assets portfolio is worthless. However, suppose that these two exploratory prospects are in the same geologic play⁷, so that the prospects are dependent with positive correlation. If these prospects have positive correlation, in case of success in one prospect, the chance factor for the second prospect (CF_2) *must* be revised upward (to CF_2^+) and in case of failure CF_2 must be revised downward (to CF_2^-). **Figure 1** illustrates this learning updating with the information revelation generated by the first option exercise.

⁵ EMV is used in exploration economics and it is a concept analog to NPV (net present value).

⁶ In a more general case (with market uncertainty), instead the NPV we have the value of the development option.

⁷ The prospects share common geological hypotheses, e.g., existence (or not) of oil migration from the source rock to that area with presence of reservoir rock and synchronism for the sequential geologic events.

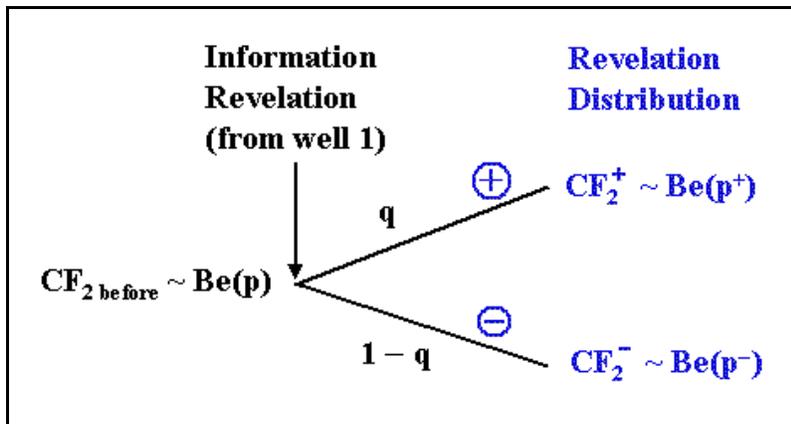


Figure 1 – Effect of the Well 1 Signal on the Chance Factor CF_2

The intensity of this update is function of the degree of dependence between the prospects and will be discussed with details mainly in the section 4. The probability of a positive information revelation (q) is the success probability for the well 1. In this symmetrical example, both prospects have the same unconditional success probability (p), so that $p = q$. In this case these random variables are called *exchangeable* (see section 4). In this example consider that the dependence degree makes $CF_2^+ = 50\%$ in case of success for the well 1. Probabilistic consistency, or more precisely the *law of iterated expectations* (see section 3), demands that $CF_2^- = 21.43\%$. In case of bad news (i.e., using CF_2^- in the eq. 1), the EMV_2 is even worse than the -1.5 million obtained with CF_2 . But it is an option so that the prospect 2 will not be drilled in case of bad news and the value of the prospect in this scenario is zero. However, in case of good news the prospect 2 becomes attractive ($EMV_2^+ > 0$) so that the drilling option is exercised in case of good news. So, the portfolio value is:

$$EMV_1 + E[\text{option}(EMV_2)] = -1.5 + [(0.7 \times \text{zero}) + (0.3 \times 17.5)] = + 3.75 \text{ million } \$$$

A very different value when compared with the case of independent prospects. Note that the positive result is due to both the optional nature of investment drilling and the information revelation generated by the first drilling.

Using this example, suppose now that firm 1 owns prospect 1 and firm 2 owns prospect 2. In this case, one alternative is to play a non-cooperative game named *war of attrition* in which one firm wait for the other firm option exercise in order to use this information as a free rider. The game prize is the information revelation value that depends on the learning intensity given by degree of dependence between the prospects. Other strategic alternative is to play a game of bargain where again the surplus to be divided is given by the effect of the information revelation on the prospects, which

depends on degree of dependence (or learning degree) between the prospects. So, in order to quantify interesting real options and option games problems is necessary to study the learning degree (or revelation power) of a signal over a variable of interest. This is done in the next two sections.

3 – Learning Measures

In this section is presented a series of definitions, lemmas, propositions, theorem and a list of axioms.

Definition. Prior distribution: it is a probability distribution that represents all prior knowledge that the decision maker knows about a random variable (based in Lawrence, 1999, p.5). The prior distribution support⁸ includes all possible values (scenarios) that this variable can assume, while the probability density represents the best estimate of the occurrence probability for these scenarios, using the current knowledge (prior information).

Definition. Information structure: comprises the space of messages (signals) plus the joint measure of states and messages (Lawrence, 1999, p.16)⁹. So, the information structure \mathcal{I} is defined by:

$$\mathcal{I} = \{ \mathbf{S}, \mathbf{p}(\mathbf{x}, \mathbf{s}) \} \quad ()$$

The joint measure is the *joint probability distribution* of two random variables $p(\mathbf{x}, \mathbf{s})$. This suggests that is necessary to study bivariate and multivariate probability distributions in order to analyze VOI problems. The above definition suggests also a *comparison of information structures* to determine if one structure is more informative than other. This was the theme of two classical papers of Blackwell (1951) and (1953).

⁸ The support of a distribution $p(\mathbf{x})$ is the set of values where $p(\mathbf{x}) > 0$.

⁹ Radner & Stiglitz (1984, p.34) definem estrutura de informação como a matriz de Markov de funções (distribuições) verossimilhança $g(\mathbf{s} | \mathbf{x})$. No contexto da tese, onde a função verossimilhança será desnecessária, ficará claro que a definição de Lawrence é mais adequada.

The term “revelation” enfatiza a mudança de expectativas com o novo cenário revelado pela informação, i. é, ressalta o *processo de aprendizagem* ou *processo de descoberta* em direção ao verdadeiro valor da variável. Assim, será usada a seguinte equação para definir a função $R_X(S)$:

$$R_X(S) = E[X | S] \quad (0)$$

O conceito de “revelação” aqui usado tem similaridades e diferenças com o famoso *princípio da revelação* (“revelation principle”) da literatura de *assimetria de informação*, mais especificamente da *teoria de desenho de mecanismos*¹⁰ (ex., Salanié, 1994) e de *jogos Bayesianos* (ex.: a vasta literatura de *jogos de sinalização*). Nessa tese, a revelação é do verdadeiro valor de um parâmetro técnico X (verdadeiro estado da natureza de X), enquanto que o conceito usado em desenho de mecanismos é relacionado ao verdadeiro tipo de um agente.

Definição. Processo de revelação: é uma *seqüência* de variáveis aleatórias $\{R_{X,1}, R_{X,2}, R_{X,3}, \dots\}$ ¹¹ geradas por uma seqüência de informações ou sinais S_1, S_2, S_3, \dots sobre uma variável de interesse X, que tem como principal característica a redução esperada de incerteza advinda da chegada de nova informação. Processo de revelação é um *processo de aprendizagem* probabilístico. Na literatura matemática, é às vezes chamado de “*acumulando dados sobre uma variável aleatória*” (Williams, 1991, p.96).

Processos de revelação podem ser vistos como processos estocásticos, mas geralmente os processos de revelação devem ser indexados por eventos e não pelo *tempo* como ocorre com a maioria dos processos estocásticos estudados. Essa tese está interessada especialmente nos processos com eventos sendo exercícios de opções de aprendizagem, a fim de modelar a evolução (redução esperada) da incerteza técnica com o processo de investimento em informação. Um exemplo de processo de revelação indexado por eventos é a perfuração seqüencial de poços de delimitação em uma reserva de petróleo com volume B incerto. Um exemplo de processo de revelação indexado pelo tempo é o que ocorre geralmente com o retorno de ações de *novas* empresas que são lançadas no mercado (IPO, *Initial Public Offering*). A volatilidade dessas ações é geralmente alta in the beginning, but with the

¹⁰ Usa-se o princípio de revelação para desenhar um mecanismo direto (exs.: contratos, regras de um leilão) de forma a ser ótimo para um jogador revelar toda a verdade sobre o seu tipo.

¹¹ Poderia se definir esse processo como uma seqüência de momentos das distribuições posteriores, com a distribuição de expectativas condicionais sendo um deles. Ou até pensar numa seqüência de distribuições de distribuições posteriores.

passage of time this volatility is reduced (but never to zero) as the investors learn about the firm capacity to generate return to their stockholders. Nesse caso se tem um *processo de revelação apenas num transiente temporal* em que houve uma difusão de informação sobre a capacidade de essa ação gerar retorno e sobre o seu valor justo (ou de equilíbrio dinâmico no mercado). Após esse transiente, o processo de revelação deixa de existir e o mercado tem informação completa. Exemplos de processo que não são de revelação são o de preços de ações seguindo um MGB (novas informações não reduzem a incerteza esperada) e o modelo de “ruído acumulativo” descrito em Childs & Ott & Riddiough (2001). Não há redução esperada de incerteza (volatilidade) nesses casos.

Definição. **Revelação total da variável X:** significa a revelação de um cenário c tal que $\Pr(X = c) = 1$, onde c é uma constante pertencente ao suporte de $p(x)$. Em termos gerais, se a informação disponível é dada pela sub-sigma-álgebra Ψ , revelação total de X significa que X é Ψ -mensurável e, logo, é válido escrever $E[X | \Psi] = X$ quase certamente¹². Intuitivamente, significa que existe informação perfeita sobre o verdadeiro estado da natureza da variável X. Em termos matemáticos, “*uma variável aleatória é Ψ -mensurável se e somente se ela assume valores constantes sobre os átomos de Ψ* ” (Shiryaev, 1996, p.80).

Será visto que todo processo de revelação converge¹³ para uma variável aleatória integrável denotada por X_∞ quando $n \rightarrow \infty$. Mas nem sempre converge para uma revelação total de X, i. é, nem sempre $X_\infty = X$. Matematicamente um processo de revelação é igual ao chamado *processo de Doob* (Karlin & Taylor, 1975, p.246 e 295). Nem toda seqüência de v.a. converge no limite para uma variável integrável X_∞ . Todo processo de revelação converge porque o processo $\{R_{X,1}, R_{X,2}, \dots, R_{X,n}\}$ é *uniformemente integrável* (definido a seguir). Será visto in the Lemma 1 que, como o processo de revelação é uniformemente integrável, então ele converge para X_∞ (que sob certas condições pode ser X) quando $n \rightarrow \infty$. Como esse processo sempre converge para algum X_∞ , seria redundante o qualificativo “convergente” para esse processo, de forma que se usará o adjetivo “convergente” apenas no caso limite de revelação total, i. é, quando $X_\infty = X$.

Mas dificilmente essa teoria teria a utilidade prática e simplicidade que será obtida trabalhando focado em expectativas condicionais.

¹² Além disso, operações algébricas ordinárias tais como soma, multiplicação e divisão, não destroem a mensurabilidade (Gallant, 1997, p.47).

¹³ Converte quase certamente (com probabilidade 1), o que implica que converge em probabilidade, o que também implica convergência em distribuição (Karlin & Taylor, 1974, p.18).

Lemma 1: Seja $\{R_{X,1}, R_{X,2}, \dots, R_{X,n}\}$ um processo de revelação, i. é, $R_{X,k} = E[X | \mathfrak{S}_k]$ são definidos no mesmo espaço de probabilidade (Ω, Σ, P) , sendo X integrável e onde \mathfrak{S}_k é uma *filtração*¹⁴ $\{\mathfrak{S}_k: k \geq 0\}$, com \mathfrak{S}_k sendo gerada pela seqüência de sinais $\{S_k\}$. Então o processo de revelação é uniformemente integrável e assim quando $n \rightarrow \infty$, existe q.c. um limite de $R_{X,n}$ em L^1 que é uma v.a. integrável, denotada por X_∞ , que é também uma expectativa condicional, i. é:

$$\lim_{n \rightarrow \infty} R_{X,n} = X_\infty = E[X | S_1, S_2, \dots] = E[X | \mathfrak{S}_\infty] \quad (0)$$

Prova: Primeiro tem de ser provado que *qualquer* processo de revelação é um martingale. Isso será mostrado no Teorema 1(d) abaixo. A prova de que integrabilidade uniforme é *suficiente* para a convergência de um martingale em L^1 é dada pelo famoso *Teorema da Convergência de Martingale de Doob*¹⁵ (ver, por ex., Brzezniak & Zastawniak, 1999, theorem 4.2, p.71-73). A prova de que o processo de revelação (um “martingale de Doob”) é uniformemente integrável é dada por ex. em Ross (1996, p.319) ou Karlin & Taylor (1975, p.295-296) e então existe um limite dado por uma v.a. integrável X_∞ . A prova que esse limite é uma expectativa condicional $E[X | \mathfrak{S}_\infty]$ é dada por Karlin & Taylor (1975, p.310)¹⁶. □

Na verdade essa proposição poderia ser ainda mais forte. Pode ser provado que *todo* martingale uniformemente integrável pode ser escrito como um processo de revelação (i. é, um processo de Doob), ver Karlin & Taylor (1975, p.311-312, em especial o Lemma 7.3) ou Brzezniak & Zastawniak (1999, theorem 4.4, p.77) ou até mesmo em Doob (1953, cap.VII).

O Teorema 1 a seguir descreve as 4 principais propriedades da distribuição de revelações (distribuição da variável R_X). As 4 propriedades são: a média de R_X , a variância de R_X , R_X no caso limite de revelação total e a propriedade de martingale dos processos de revelação (seqüências de R_X). Ele será formulado, provado e depois será discutida a intuição por trás e suas conseqüências.

¹⁴ Pode-se interpretar a filtração \mathfrak{S}_n gerada pela informação seqüencial $\{S_1, S_2, \dots, S_n\}$ como um conjunto contendo toda a informação disponível no estágio n . Em termos técnicos, é uma crescente família de sub-sigma-álgebras gerada pelas informações. Ex.: $R_{X,2} = E[X | S_1, S_2]$.

¹⁵ Esse teorema tem sido mais usado para provar teoremas tais como a lei 0-1 de Kolmogorov e do filtro de Kalman. A aplicação no contexto da tese não tem sido explorada.

¹⁶ Uma prova simples é ver que $R_{X,n}$ é uma função de S_1, S_2, \dots, S_n , de forma que o limite X_∞ é uma função mensurável dessa seqüência de sinais e portanto mensurável com respeito a \mathfrak{S}_∞ .

Proposição 1 (Distribuição de Revelações): Sejam as variáveis aleatórias X e S com média e variância finitas¹⁷ definidas no espaço de probabilidade $(\Omega, \Sigma, \mathbb{P})$. X é a variável de interesse e tem probabilidade a priori $p(x)$. S é chamado de sinal ou nova informação e gera a sigma-álgebra Ψ , onde Ψ é uma sub-sigma-álgebra de Σ , i.é, $\Psi \subseteq \Sigma$. Seja $p(R_X)$ a distribuição¹⁸ de probabilidades da variável $R_X = E[X | S]$, denominada *distribuição de revelações* de X . Então, a distribuição de revelações tem as seguintes propriedades que *quase*¹⁹ a definem:

- (a) No caso limite de revelação total da variável X , a variância de qualquer distribuição posterior é zero e a distribuição de revelações $p(R_X)$ é igual à distribuição a priori $p(x)$.
- (b) A média da distribuição de revelações é igual à média de X , i. é:

$$E[R_X] = E[X] \quad ()$$

- (c) A variância da distribuição de revelações é dada simplesmente pela redução esperada da variância de X induzida pelo sinal S , i. é:

$$\text{Var}[R_X] = \text{Var}[X] - E[\text{Var}[X | S]] \quad ()$$

- (d) Seja um processo seqüencial de informação S_1, S_2, S_3, \dots e as variáveis aleatórias $R_{X,n} = E[X | S_1, S_2, \dots, S_n]$, $n = 1, 2, \dots$. Então, o *processo de revelação* $\{R_{X,1}, R_{X,2}, R_{X,3}, \dots\}$ é um martingale.

Prova:

(a) A Proposição 5 acima garante que *um* limite *sempre* existe com probabilidade 1. Além disso, sabe-se que $\text{Prob}(X = c) = 1 \Leftrightarrow \text{Var}[X] = 0$ (prova: DeGroot & Schervish, 2002, theorem 4.3.1, p.198). Então, para cada possível cenário c do suporte da distribuição a priori que possa ser revelado, estará associada uma distribuição posterior com variância zero. O restante da prova decorre diretamente da definição de distribuição a priori. Para ver isso, como a distribuição de revelações é a distribuição das médias das distribuições posteriores e essas médias colapsam para os cenários c (pois $E[c | S] = c$, se c é constante) do suporte da distribuição a priori, então o suporte da distribuição de revelações é simplesmente o mesmo suporte da distribuição a priori. Que as probabilidades de

¹⁷ Um matemático poderia preferir dizer apenas que X e S são duplamente integráveis, i. é, $E[|X^2|] < \infty$ e $E[|S^2|] < \infty$, ou de forma equivalente que X e $S \in L^2$, pois isso automaticamente implica que X e $S \in L^1$ (Williams, 2001, p.65). Mas aqui isso não é restritivo pois as aplicações práticas do teorema são focadas apenas em problemas em que as médias e variâncias são finitas.

¹⁸ Quando se menciona apenas “distribuição” e/ou se usa letra minúscula, significa a função densidade de probabilidades (caso contínuo), etc. Para a função distribuição acumulada, se usará letra maiúscula e em geral o texto será mais explícito.

¹⁹ Definição: distribuição quase-definida é aquela em que se conhece pelo menos a *média*, a *variância* e que pertence a um processo seqüencial de distribuições com *origem* definida e *convergente* para uma distribuição totalmente definida. Então o Teorema 1 quase-define $R_X(S)$. Uma distribuição *totalmente* definida é simulável e tem função distribuição acumulada conhecida.

ocorrência (massa ou densidade) desses cenários são iguais às descritas pela distribuição a priori decorre diretamente da definição de distribuição a priori (ver item 3.1.2). Assim, em caso de revelação total, a distribuição de revelações é igual à distribuição a priori. \square

(b) Essa propriedade é conhecida na literatura por *lei das expectativas iteradas*. Esse item pode ser formulado de forma mais geral, com a sub-sigma-álgebra Ψ (em vez da v.a. S) e da seguinte forma: se R_X é qualquer *versão*²⁰ de $E[X | \Psi]$ então $E[R_X] = E[X]$.

Definição: Sejam duas variáveis aleatórias X e S com médias e variâncias finitas, definidas no espaço de probabilidade $(\Omega, \Sigma, \mathbb{P})$. Define-se a *redução percentual esperada de variância de X dado S* como:

$$\eta^2(\mathbf{X} | \mathbf{S}) = \frac{\text{Var}[\mathbf{X}] - \mathbf{E}[\text{Var}[\mathbf{X} | \mathbf{S}]]}{\text{Var}[\mathbf{X}]} \quad ()$$

A notação η^2 é adotada por duas razões: (a) para facilitar a conexão com sua interpretação estatística, a *razão de correlação* (“correlation ratio”), também chamada de razão de *correlações* e também conhecida por “*eta-squared*” em alguns livros de estatística; e (b) em algumas situações (ex.: processos de Bernoulli) é mais intuitivo (para efeito da interpretação) usar a raiz positiva de η^2 , ou seja, simplesmente η .

Aplicando o item (c) do Teorema 1 sobre a variância da distribuição de revelações na equação que define $\eta^2(\mathbf{X} | \mathbf{S})$, se obtém:

$$\eta^2(\mathbf{X} | \mathbf{S}) = \frac{\text{Var}[\mathbf{E}[\mathbf{X} | \mathbf{S}]]}{\text{Var}[\mathbf{X}]} = \frac{\text{Var}[\mathbf{R}_X]}{\text{Var}[\mathbf{X}]} \quad ()$$

Ou seja, a medida de aprendizagem proposta é a variância normalizada da distribuição de revelações, sendo normalizada pela variância inicial, i. é, pela variância da distribuição a priori.

The following example illustrates this paper claim that a good learning measure must be asymmetric for the general case. Again we use an example from information theory. The *expected entropy* (also

known as *conditional entropy* of X given S or *equivocation*, see McEliece, 2002, p.20) is the average entropy of the posterior distributions $p(x|s)$ from all possible outcomes of the signal S , i.e., for the discrete case the concept of expected entropy is defined as:

$$H(X | S) = - \sum_x p(x | y) \log[p(x | y)]$$

In words, expected entropy measures the expected remaining uncertainty about X after S has been observed. The difference between the (unconditional) entropy $H(X)$ and the expected entropy after the signal $H(X | S)$ is a kind of expected reduction of uncertainty (entropy) with the information revealed by S . This amount is called *mutual information*, also known as *information transmitted* or *uncertainty removed* (Lawrence, 1999, p.62), is defined by:

This example also shows that we can be interested in asymmetric effects from the interaction of two random variables. For many statistical applications, like “distance of a joint distribution from the independence case”, a symmetric measure for the dependence of X and S looks natural, because distance is a symmetrical concept. But symmetry has nothing of “natural placing” in value of information applications. In particular we are interested in how S reduce the variance of X or how valuable is S for X . It doesn't matter how the variable X reduces the variance of S . So, there is an *asymmetric economic interest* to evaluate only one direction for the relations between the random variables X and S . So, a measure of learning that captures this asymmetric interest must be asymmetric in general and symmetric only in particular cases.

²⁰ Se R_X^* é uma versão de R_X , então $R_X^* = R_X$ quase certamente (Williams, 1991, p.84).

It is easy to prove that $\eta^2 \in [0, 1]$. This means that as measure η^2 doesn't concern if either the dependence is positive or negative²¹. This is an advantage²² in our context because we are interested in learning, in improving our knowledge over X by using the information from the signal S . In order to see this, consider the following simple example.

Example: It is known that an unfair coin has either two heads or two tails and with 50% for each scenario. Therefore, the prior probability for the outcome head is a Bernoulli distribution with parameter $p = 50\%$. There are two experts who know all the truth about the coins. The expert S_1 is known because always says the truth. The expert S_2 is known because always lies. Of course we gain the same knowledge by buying either S_1 or S_2 advises. The positive dependence between X and S_1 and the negative dependence between X and S_2 provide the same knowledge. So, by the VoI point of view they are indistinguishable.

Because it is always positive, somebody could suggest a metric like *the reliability of the information*, i.e., the *likelihood* of the information $L(S)$. However, the previous example shows that the likelihood metric assumes two values for the same provided knowledge (full revelation or the true value of X), i.e., $L(S_1) = 100\%$ and $L(S_2) = 0\%$. Hence, a relative advantage of the metric $EVR(\cdot)$ over $L(\cdot)$ is that the former assumes a single value for full learning situation, i.e., $EVR(X | S_1) = EVR(X | S_2) = 1$.

Proposição 6: Sejam duas variáveis aleatórias não triviais²³ X e S com médias e variâncias finitas, definidas no espaço de probabilidade $(\Omega, \Sigma, \mathbb{P})$. Seja a medida de aprendizagem $\eta^2(X | S)$ definida pela equação (106). Então, essa medida de aprendizagem tem seguintes propriedades:

²¹ Metrics that allow negative values, like the correlation coefficient, are not measures.

²² This is a disadvantage in other applications like portfolio theory. In order to reduce the portfolio variance we are interested in shares with negative correlation. So this positive/negative distinction is relevant in some applications.

²³ Por “v.a. não-triviais”, se assume que as variâncias são estritamente positivas. O caso trivial só interessa na análise de limites de processos. Assim, quando se refere a v.a. em geral elas são não-triviais. Como sempre, a proposição é válida com probabilidade 1 (i. é, quase certamente).

(a) A medida $\eta^2(X | S)$ existe sempre que a variância de X for estritamente positiva (i. é, o problema for não-trivial) e finita;

(b) A medida η^2 é em geral assimétrica, i. é, $\eta^2(X | S) \neq \eta^2(S | X)$

(c) A medida η^2 é normalizada no intervalo unitário, ou seja²⁴,

$$0 \leq \eta^2 \leq 1 \quad (0)$$

(d) Se X e S são v.a. independentes, então a medida η^2 é zero:

$$\mathbf{X \text{ e } S \text{ independentes}} \Rightarrow \eta^2(X | S) = \eta^2(S | X) = 0 \quad (0)$$

Além disso, η^2 é zero se e somente se a variância da distribuição de revelações for zero também:

$$\eta^2(X | S) = 0 \Leftrightarrow \text{Var}[\mathbf{R}_X(S)] = 0 \quad (0)$$

(e) Se $\eta^2(X | S) = 1 \Leftrightarrow$ existe uma função real, a v.a. $g(S)$, em que $X = g(S)$;

(f) A medida $\eta^2(X | S)$ é invariante sob transformações lineares de X, i. é, para quaisquer números reais a e b, com $a \neq 0$, tem-se:

$$\eta^2(a X + b | S) = \eta^2(X | S) \quad (0)$$

(g) A medida $\eta^2(X | S)$ é invariante sob transformações lineares e não-lineares de S se a transformação $g(S)$ for uma função 1-1 (injetiva ou com função inversa). Em geral, se $g(S)$ é uma função mensurável pela sigma-álgebra gerada por S, então vale a desigualdade:

$$\eta^2(X | g(S)) \leq \eta^2(X | S), \text{ com igualdade se } g(s) \text{ for uma função 1-1} \quad (0)$$

(h) Se as v.a. Z_1, Z_2, \dots são independentes e identicamente distribuídas (iid) e se $S = Z_1 + Z_2 + \dots + Z_j$ e $X = Z_1 + Z_2 + \dots + Z_{j+k}$ para quaisquer inteiros não-negativos j e k, com $j + k > 0$, a medida $\eta^2(X | S)$ proposta é dada diretamente por:

$$\eta^2(X | S) = \frac{j}{j + k} \quad (0)$$

Prova:

(a) Essa propriedade é trivial, considerando a premissa de $\text{Var}[X] > 0$ mas finita e a premissa de $\text{Var}[S]$ finita. \square

(b) A medida η^2 é *em geral* assimétrica se pelo menos um exemplo mostrar assimetria. Isso foi mostrado em um exemplo nesse sub-item (Figura 38). \square

(c) As eqs. (106) e (107) mostram isso: pela eq. (107) η^2 é uma razão de variâncias e assim não pode ser menor que zero devido a definição de variância. A eq. (106) mostra que o valor máximo é igual a 1, pois η^2 é maximizado minimizando $E[\text{Var}[X | S]]$, cujo mínimo ocorre para $\text{Var}[X | S] = 0$ para todo $s \in S$, i. é, para $E[\text{Var}[X | S]] = 0 \Rightarrow \eta^2(X | S) = 1$. \square

(d) Se X e S são v.a. independentes, então $E[X | S] = E[X]$ e $E[S | X] = E[S]$ (Lema 2(d), eq. (86), propriedade elementar da independência, ver, por ex., Williams, 1991, p.88)²⁵. Além disso, a variância dum variável Y qualquer é *definida* por $\text{Var}[Y] = E[(Y - E[Y])^2]$. Logo, as variâncias das distribuições de revelações $R_X(S)$ e $R_S(X)$ são ambas iguais a zero, pois:

$$\text{Var}[R_X(S)] = E[(E[X | S] - E[E[X | S]])^2] = E[(E[X | S] - E[X])^2] = 0$$

$$\text{Var}[R_S(X)] = E[(E[S | X] - E[E[S | X]])^2] = E[(E[S | X] - E[S])^2] = 0$$

Onde foram usados o item (b) do Teorema 1 e a propriedade elementar da independência mencionada acima. Se as variâncias das distribuições de revelações são iguais a zero, então também o serão as medidas $\eta^2(X | S)$ e $\eta^2(S | X)$, pois essas medidas de aprendizagem são variâncias normalizadas das distribuições de revelações (ver eq. 107). Assim fica provado a eq. (128) e a volta (\Leftarrow) da eq. (129). Para provar a ida (\Rightarrow) da eq. (129), note na eq. (107) que $\eta^2(X | S) = 0 \Rightarrow \text{Var}[R_X(S)] = \eta^2(X | S) \text{Var}[X] = 0$ se $\text{Var}[X] > 0$. \square

²⁴ Poderia se acrescentar que η^2 é *realmente* uma medida, pois $\eta^2 \geq 0$.

²⁵ Mas se $E[X | S] = E[X]$ e $E[S | X] = E[S]$ não implica que X e S são independentes (embora sejam na grande maioria dos casos). Por ex., se X e Y tem distribuição uniforme sobre um círculo centrado na origem, claramente são dependentes, mas $E[X | Y] = E[X]$ e $E[Y | X] = E[Y]$.

(e) Se $\eta^2(X | S) = 1$, a eq.(106) $\Rightarrow E[\text{Var}[X | S]] = 0 \Rightarrow E[(X - E[X | S])^2] = 0$ e assim com probabilidade 1 $\Rightarrow X = E[X | S] \Rightarrow X$ é mensurável pela sigma-álgebra de S e portanto pode ser escrito $X = g(S)$.

Em Hall (1970, p.342), usa-se o axioma que se uma medida de dependência é igual a 1 $\Rightarrow X = g(S)$, mas sem a volta (\Leftarrow). Em Hall, a volta $X = g(S) \Rightarrow$ medida de dependência igual a 1, não é considerada necessária. A explicação é que nesses estudos mais teóricos se quer incluir o caso de variância de X igual a infinito (o que não é o caso dessa tese). Aqui, assumindo que $\text{Var}[X]$ é finito, então vale a volta para a medida η^2 . A prova da volta (\Leftarrow) é ainda mais simples: se S é revelado então, pela definição de função, $X = f(S)$ é *unicamente* determinado. Assim $E[\text{Var}[X | S]] = 0$ e logo $\eta^2(X | S) = 1$ pela eq. (106). Conforme Hall (1970, p.342), no caso de variância infinita de X, quando o conhecimento de S *reduz a variância de infinito para um valor finito*, isso poderia ser considerado um *completo* estado de dependência de maneira análoga ao caso de variância de X finita, em que S reduz a variância de X de um valor finito para um valor igual a zero. Nas aplicações da tese em que $\text{Var}[X]$ é sempre finito, escrever a propriedade mais forte (\Leftrightarrow) é muito mais conveniente. \square

(f) A prova é simples. Aplicando a eq. (106) para a variável $Y = aX + b$:

$$\eta^2(aX + b | S) = \frac{\text{Var}[aX + b] - E[\text{Var}[aX + b | S]]}{\text{Var}[aX + b]} \Rightarrow$$

$$\eta^2(aX + b | S) = \frac{a^2 \text{Var}[X] - a^2 E[\text{Var}[X | S]]}{a^2 \text{Var}[X]}$$

Como $a \neq 0$, pode-se simplificar (cortar a^2) e obter $\eta^2(X | S)$. \square

(g) O caso de igualdade se a função $Y = g(S)$ for 1-1 é porque a inversa $S = g^{-1}(Y)$ existe, ou seja, g^{-1} é uma *função* que por definição determina S de forma única se for conhecido $Y = g(S)$. Assim, conhecer $g(S)$ equivale a conhecer S.

Se $g(S)$ não fosse 1-1, o conhecimento de $g(S)$ seria menor do que o conhecimento de S, por ex., se $Y = S^2$ então o valor $Y = 1$ poderia ser tanto devido a $S = 1$ como a $S = -1$. Por isso, a intuição diz que $\eta^2(X | g(S))$ deve ser menor que $\eta^2(X | S)$. Para provar formalmente essa desigualdade, considere a eq. (117) do Lema 4, onde $\eta^2(X | S)$ é o supremo de $\rho^2(X, f(S))$ para *qualquer* função real $f(S)$. Isso implica que $\eta^2(X | S) \geq \rho^2(X, f(S))$ para *qualquer* função real $f(S)$. Se essa função é qualquer, então isso inclui a função $f(S) = E[X | g(S)]$. Logo:

$$\eta^2(X | S) \geq \rho^2(X, E[X | g(S)])$$

Mas o Lema 4(c), eq. (119), diz que $\eta^2(X | Y) = \rho^2(X, E[X | Y])$. Fazendo $Y = g(S)$ obtém-se $\eta^2(X | g(S)) = \rho^2(X, E[X | g(S)])$. Substituindo na desigualdade anterior, vem:

$$\eta^2(X | S) \geq \eta^2(X, g(S)) \quad \square$$

Essa propriedade vale também para o caso de X e S serem v.a. complexas, conforme mostrou Hall (1970, p.353-354 e 349), numa demonstração similar²⁶.

(h) Essa propriedade foi sugerida por Hall (1970) como um dos axiomas que uma medida de dependência deveria ter para se relacionar com o coeficiente de correlação, indicando que a razão de correlação η^2 atende esse axioma²⁷, mas sem apresentar a prova. Assim, é conveniente apresentar uma demonstração, a qual é bem simples se usar a eq. (107), i. é, η^2 como a variância normalizada da distribuição de revelações. Como Z_1, Z_2, \dots tem as mesmas distribuições, suas médias e variâncias são iguais e conhecidas com a informação corrente, ou seja:

$$E[Z_1] = E[Z_2] = \dots = E[Z]$$

$$\text{Var}[Z_1] = \text{Var}[Z_2] = \dots = \text{Var}[Z]$$

Como as v.a. Z_i são independentes, a variância da soma é igual à soma das variâncias. Assim, a variância de X (soma de $j + k$ v.a. Z_i) é:

$$\mathbf{Var}[X] = \text{Var}[Z_1] + \text{Var}[Z_2] + \dots + \text{Var}[Z_{j+k}] = (\mathbf{j} + \mathbf{k}) \mathbf{Var}[Z] \quad (0)$$

Além disso, o Lema 2(a) (eq. 82) permite decompor $E[X | S]$ como:

$$E[X | S] = E[Z_1 + Z_2 + \dots + Z_j | S] + E[Z_{j+1} + \dots + Z_{j+k} | S]$$

Como as v.a. Z_{j+1}, \dots, Z_{j+k} são independentes das v.a. $Z_{i < j+1}$, então pelo Lema 2(d) (eq. 86), pode-se escrever:

²⁶ Existe uma propriedade para v.a. complexas que é parecida com a mostrada no Lema 4.

²⁷ Na verdade tem uma pequena imprecisão em Hall (1970) que menciona que η (e não η^2) atende a esse axioma. Como é mostrado aqui, η^2 é que atende ao seu axioma. O curioso é que Kruskal (1958, p.818) comete a mesma imprecisão para ρ (o correto é que a eq. 132 vale para ρ^2).

$$E[X | S] = E[Z_1 + Z_2 + \dots + Z_j | S] + E[Z_{j+1} + \dots + Z_{j+k}]$$

Além disso, $Z_1 + Z_2 + \dots + Z_j$ é S -mensurável (pois $S = Z_1 + Z_2 + \dots + Z_j$) e assim pelo Lema 2(b) (eq. 83), pode-se tirar para fora do operador $E[\cdot | S]$, i. é:

$$E[X | S] = Z_1 + Z_2 + \dots + Z_j + k E[Z] \quad ()$$

Logo, a variância da distribuição de revelações é dada por:

$$\text{Var}[E[X | S]] = \text{Var}[Z_1 + Z_2 + \dots + Z_j] + k^2 \text{Var}[E[Z]]$$

Mas como as v.a. Z_i são independentes e $E[Z]$ é uma constante (variância igual a zero), pode-se escrever a variância da distribuição de revelações como:

$$\text{Var}[E[X | S]] = j \text{Var}[Z] \quad ()$$

Dividindo a eq. (135) pela eq.(133) se obtém a medida de aprendizagem η^2 :

$$\eta^2(X | S) = \frac{j \text{Var}[Z]}{(j+k) \text{Var}[Z]} = \frac{j}{j+k} \quad \square$$

In order to motivate to the next theorem about the connection between independent signals and full revelation, let us first present a simple example. Consider an already discovered oilfield with uncertainty about the true volume of this reserve. Denote B the random variable number of barrels (or volume) of this reserve. There is one area (one reservoir) with B_0 million barrels of proved reserves, and two independent reservoirs (different geologic ages) with reserves of B_1 and B_2 if these reservoirs are filled with petroleum, see Figure 3. But there is uncertainty if the reservoir is filled with water or with petroleum. The first reservoir has probability q of petroleum and $(1 - q)$ of water, whereas the second reservoir has probabilities p and $(1 - p)$ for petroleum and water, respectively. By drilling one appraisal well in each area with uncertainty we reveal all the truth about the variable B , that is, the reserve volume B is a solely function of two independent random variables S_1 and S_2 :

$$B(S_1, S_2) = B_0 + (B_1 \times S_1) + (B_2 \times S_2)$$

Where the signals are independent Bernoulli random variables, $S_1 \sim \text{Be}(q)$ and $S_2 \sim \text{Be}(p)$. For Bernoulli variables independence means (iff) that both $\eta^2(S_2 | S_1)$ and $\eta^2(S_1 | S_2)$ are equal zero. Here we are interested in $\eta^2(B | S_1)$ and $\eta^2(B | S_2)$, the signals S_1 and S_2 relevancy to predict B .

Let us work with numbers: $B_0 = 100$; $B_1 = 50$; $B_2 = 50$; $q = p = 50\%$. The expected value equation for B is: $E[B] = B_0 + (B_1 \times q) + (B_2 \times p) = 150$. The unconditional variance is $\text{Var}[B] = 1250$. It is easy to calculate that the revelation of the signal S_1 (by drilling the first well) reduces $\text{Var}[B]$ by the half, i.e., $\eta^2(B | S_1) = 50\%$. Similarly $\eta^2(B | S_2) = 50\%$. The example most interesting feature is:

$$\eta^2(B | S_1) + \eta^2(B | S_2) = \eta^2[f(S_1, S_2) | S_1] + \eta^2[f(S_1, S_2) | S_2] = 1 \quad (3.1)$$

We will see that this relation is much more general, valid for *any function*, and for $n > 0$ independent variables. This η^2 property is not verified by the “competitors” correlation $(0.71 + 0.71 > 1)^{28}$ and mutual information $(1 + 1 > 1)^{29}$. In addition, due to its scenario insensitivity, mutual information does not change if the volume B_1 is the double of B_2 , when clearly the signal S_1 becomes more relevant than the signal S_2 in terms of reducing uncertainty about the volume B in problems related to value of information. If $B_1 = 100$ (remaining $B_2 = 50$, etc.), the η^2 property (equation 3.1) remains valid but with higher weight to signal S_1 , i.e., $\eta^2(B | S_1) + \eta^2(B | S_2) = 0.8 + 0.2 = 1$.

The key concept used in the equation 3.1 was *independence*. Following Breiman (1969, chapter 4), independence of random variables is a strong condition; independence is a *family* or *hereditary* property. It is illustrated by the following known results (e.g., Breiman, 1969), which will be next used in the full revelation theorem for independent signals.

Lemma: Let the signals S_1, S_2, \dots, S_n , be independent random variables. Then:

- (a) Any smaller group of these variables is also independent;

²⁸ In this simple example, because the function $B(S_1, S_2)$ is *linear*, the correlation-square (named coefficient of determination) is equal to EVR. This is a practical advantage of EVR.

²⁹ Considering base 2 in the logarithms from entropy equation. Other base will not make things better in the general case.

- (b) For *any* functions f, g, \dots, h , the variables $f(S_1), g(S_2), \dots, h(S_n)$, are independent; and
- (c) Functions of disjoint groups of these variables are independent (e.g., $f(S_1)$ and $g(S_2, S_3)$ are independent).

The full revelation theorem for independent signals is presented below. It can be used directly to know the participation of each variable in the full revelation process over the variable of interest X .

Theorem 1: Let the signals S_1, S_2, \dots, S_n , be independent random variables. We want to learn about X , a random variable with $\text{Var}[X] > 0$. Assume as finite all the relevant expectations and variances³⁰. Let $X = f(S_1) + g(S_2) + \dots + h(S_n)$, where f, g, \dots, h , are any real valued functions. Then:

$$\eta^2(\mathbf{X} | S_1) + \eta^2(\mathbf{X} | S_2) + \dots + \eta^2(\mathbf{X} | S_n) = \mathbf{1} \quad (3.2)$$

Proof: Recall the definition of $\eta^2(\mathbf{X} | S_i) = \text{Var}[E(\mathbf{X} | S_i)] / \text{Var}[X]$. So, the left side of (3.2) is:

$$\begin{aligned} & \eta^2(\mathbf{X} | S_1) + \eta^2(\mathbf{X} | S_2) + \dots + \eta^2(\mathbf{X} | S_n) = \\ &= \frac{\text{Var}[E(\mathbf{X} | S_1)]}{\text{Var}[X]} + \frac{\text{Var}[E(\mathbf{X} | S_2)]}{\text{Var}[X]} + \dots + \frac{\text{Var}[E(\mathbf{X} | S_n)]}{\text{Var}[X]} = \\ &= \frac{\text{Var}[E\{f(S_1) + g(S_2) + \dots + h(S_n) | S_1\}] + \dots + \text{Var}[E\{f(S_1) + g(S_2) + \dots + h(S_n) | S_n\}]}{\text{Var}[f(S_1) + g(S_2) + \dots + h(S_n)]} = \end{aligned}$$

But $E\{f(S_i) | S_i\} = f(S_i)$. Due to independence between S_i and S_j , we have $E\{f(S_i) | S_j\} = E\{f(S_i)\}$ and $\text{Var}[f(S_i) + g(S_j)] = \text{Var}[f(S_i)] + \text{Var}[g(S_j)]$. Hence:

$$= \frac{\text{Var}[f(S_1) + E\{g(S_2)\} + \dots + E\{h(S_n)\}] + \dots + \text{Var}[h(S_n) + E\{f(S_1)\} + E\{g(S_2)\} + \dots + E\{h(S_{n-1})\}]}{\text{Var}[f(S_1)] + \text{Var}[g(S_2)] + \dots + \text{Var}[h(S_n)]} =$$

³⁰ With the reasonable assumption that the variance of the prior distribution $\text{Var}[X]$ is finite, the variance of the revelation distribution is also finite and bounded by $\text{Var}[X]$, even if $n \rightarrow \infty$. So, we don't need to work with infinite variance in the limit. It contrasts with the Brownian motion, which the variance is unbounded: tends to infinite when $t \rightarrow \infty$.

However, the *unconditional* $E\{f(S_i)\}$ is not a random variable. It is a number known at the initial moment, so that $\text{Var}[E\{f(S_i)\}] = 0$. Hence, many terms from the above equation vanish:

$$\eta^2(X | S_1) + \dots + \eta^2(X | S_n) = \frac{\text{Var}[f(S_1)] + \text{Var}[g(S_2)] + \dots + \text{Var}[h(S_n)]}{\text{Var}[f(S_1)] + \text{Var}[g(S_2)] + \dots + \text{Var}[h(S_n)]} = 1$$

This kind of property can be useful also when the variable of interest is a function of a *product* and/or *quotient* of independent random variables, because we can make a logarithm transformation and apply the theorem. For example, in oil companies and in professional literature the reserve volume B is estimated considering B as a function of many independent variables. One function used is $B(\text{RF}, \text{GV}, \text{NTG}, \phi, \text{Sw}, \text{Bo}) = \text{RF} \times [\text{GV} \times \text{NTG} \times \phi \times (1 - \text{Sw})] / \text{Bo}$, where RF = recovery factor; GV = gross volume of rock; NTG = % net to gross thickness; ϕ = porosity; Sw = saturation of water; and Bo = oil volume of formation factor. Another function is $B(\text{RF}, A, h, \phi, \text{Sw}, \text{Bo}) = \text{RF} \times (A \times h \times \phi \times (1 - \text{Sw})) / \text{Bo}$, where A = area; h = net pay (thickness with oil) and the remaining variables as before. With the log-transformed $X = \ln(B)$, we can write X as a sum of functions of independent random variables, and the theorem 3.1 holds for X .

In the previous example with S_1 and S_2 independent and $B = f(S_1) + g(S_2)$, this theorem is verified: $\eta^2(B | S_1) + \eta^2(B | S_2) = 0.5 + 0.5 = 1$. Now, in the same example imagine that the Bernoulli variables S_1 and S_2 are not independent. Even B being a solely function of S_1 and S_2 , equation 3.2 will not be valid with the dependence between S_1 and S_2 . The intuition is that – *in addition* to reveal the existence of oil in its area (direct revelation on B), the signal S_1 provides also relevant information about S_2 . So, we must expect that $\eta^2(B | S_1) > 50\%$. The same reasoning is valid to S_2 , so that also $\eta^2(B | S_2) > 50\%$. Hence, we shall expect that $\eta^2(B | S_1) + \eta^2(B | S_2) > 1$. This shows that, although each *individual* signal has higher revelation power with the dependence hypothesis (signals

are more valuable), by buying both signals S_1 and S_2 , we are gathering “excess of information” because there is overlapping information set with the signals S_1 and S_2 .

Let us confirm this intuition returning to the numerical example, but now assuming *dependence*. Imagine we drill first the well number 1 (due to symmetry, the same result is obtained if we start with the well 2)³¹. The outcome of the first well will change the chance factor for the second 2, CF_2 . Consider again the Figure 1 dependence structure with $S_2 = CF_2 \sim \text{Be}(p = 50\%)$, and suppose that in case of positive revelation from the first well, we update to $CF_2^+ \sim \text{Be}(p^+ = 75\%)$, whereas in case of bad news we update $CF_2^- \sim \text{Be}(p^- = 25\%)$. It is easy to check³² that this dependence between S_1 and S_2 corresponds to $\eta^2(S_1 | S_2) = 0.25$. Note that the introduction of dependence between S_1 and S_2 , changes the prior distribution of B (increasing the variance of B). However, we are interested in the values of $\eta^2(B | S_1)$ and $\eta^2(B | S_2)$ to check equation 3.1. It is easy to calculate that $\eta^2(B | S_1) = 0.75$. By symmetry, $\eta^2(B | S_2)$ is also 0.75, so that $\eta^2(B | S_1) + \eta^2(B | S_2) = 1.5 > 1$. This result is more general, $\eta^2(B | S_1) + \eta^2(B | S_2) > 1$ if $\eta^2(S_2 | S_1) > 0$ and/or $\eta^2(S_1 | S_2) > 0$.

Remark:

Agora serão apresentados os axiomas ou propriedades desejáveis em uma medida de aprendizagem probabilística. Elas são adequadas em especial para problemas de análise econômica, mas possivelmente não apenas (medicina?).

Axiomas para Medidas de Aprendizagem Probabilística: As seguintes propriedades são desejáveis para uma medida de aprendizagem $M(X | S)$:

- A) $M(X | S)$ existe pelo menos para todas as v.a. X e S com incerteza não-trivial (estritamente positiva) e todas as v.a. com incerteza finita;

³¹ Note that, because B is a solely function of S_1 and S_2 , *after* getting the signal S_1 , $\text{EVR}([B | S_1] | S_2) = 1$. But we are interested in the ex-ante (before any information revelation) contribution of each signal i in terms of $\text{EVR}(B | S_i)$.

³² See the spreadsheet with this example http://www.puc-rio.br/marco.ind/xls/full-revelation-theorem_example.xls

- B) $M(X | S)$ deve em geral ser assimétrica para poder medir os casos em que se aprende mais sobre X conhecendo S , do que se aprende sobre S conhecendo X , e vice versa;
- C) $M(X | S)$ deve ser normalizada no intervalo unitário para facilitar a interpretação, i. é,

$$0 \leq M(X | S) \leq 1 \quad ()$$

- D) Se X e S forem independentes $\Rightarrow M(X | S) = M(S | X) = 0$, pois não há aprendizagem probabilística. Além disso,

$$M(X | S) = 0 \Rightarrow \text{não haver aprendizagem} \quad ()$$

Onde não aprender pode ocorrer não apenas por casos de independência. O conceito *aprender* é definido na medida, mas o seu sentido tem de ser invariável (ex.: medida de incerteza igual a zero para todas as aplicações);

- E) Em caso de dependência funcional a medida é máxima, i. é, para toda função real $f(\cdot)$:

$$X = f(S) \Rightarrow M(X | S) = 1 \quad ()$$

Além disso, se a medida é máxima, $M(X | S) = 1$, então:

$$M(X | S) = 1 \Rightarrow \text{aprendizagem é máxima} \quad ()$$

Onde aprendizagem máxima significa não ser possível aprender mais sobre X e *aprender* é definido na medida, desde que o seu sentido seja invariável;

- F) $M(X | S)$ deve ser invariante a transformações lineares (mudança de escala) da v.a. X ou da v.a. S , i. é, para a e b constantes reais, $a \neq 0$:

$$M(aX + b | S) = M(X | S) \quad ()$$

$$M(X | S) = M(X | aS + b) \quad ()$$

- G) $M(X | S)$ deve ser prática no sentido de ser de *fácil interpretação* (intuitivo) e *fácil de ser quantificada e estimada*.

- H) $M(X | S)$ deve ser aditiva no sentido de, caso a informação S possa ser decomposta numa soma de n fatores independentes $S_1 + S_2 + \dots + S_n$, de forma que o conhecimento de todos esses fatores proporcione uma aprendizagem máxima (aprender definida na medida, mas invariável), então a soma das medidas de aprendizagem deve ser unitária, i. é:

$$M(X | S_1) + M(X | S_2) + \dots + M(X | S_n) = 1 \quad (142)$$

Teorema 2 (Medida de Aprendizagem η^2): A medida η^2 atende a todos os axiomas de medidas de aprendizagem acima.

Prova: Note que η^2 tem a variância como medida de incerteza. Assim, *aprender* significa que é esperada uma redução de incerteza medida pela variância. *Aprendizagem máxima* significa reduzir para zero a variância posterior (revelação total) e *não aprender* é não se esperar uma redução de variância. A Proposição 6 prova todos os itens do Teorema 2, sendo que em alguns casos a medida η^2 atende condições até mais fortes do que os axiomas acima. Os axiomas A, B e C são provados pela Proposição 6 (a), (b) e (c). O Axioma D é provado pela Proposição 6 (d) e pelo fato da eq. (129) implicar em $E[\text{Var}[X | S]] = \text{Var}[X]$, logo provando a eq. (137) pela definição de “não aprender” dessa medida. O Axioma E é provado pela Proposição 6 (e) e pela definição de “aprendizagem máxima”, que implica em $E[\text{Var}[X | S]] = 0$, que implica em $\eta^2(X | S) = 1$. O Axioma F é provado pela Proposição 6 (f) e (g), sendo que η^2 atende condições mais restritas: $g(S)$ pode ser uma função 1-1 qualquer (não apenas linear). O Axioma G é o mais subjetivo, mas a medida η^2 atende amplamente no sentido que tem uma interpretação intuitiva de redução de incerteza, e será visto no item 3.3.4 que ele é facilmente estimado com métodos estatísticos populares tais como a regressão e ANOVA. O Axioma H será provado numa versão mais forte (mais geral) no Teorema 3 (ver item 3.3.3 a seguir), em que não apenas a soma de sinais independentes, mas a soma de *qualquer função* (não necessariamente 1-1) real de variáveis independentes, tem a propriedade de aditividade igual a 1 em caso de aprendizagem máxima. □

4 – Bernoulli Revelation Processes

Para avaliar o efeito de um *sinal* binário³³ S correlacionado (informação relevante) no fator de chance exploratório FC, é necessário estudar a relação de dependência entre duas distribuições de Bernoulli: o sinal binário S e o FC do prospecto de interesse. A maneira geral de estudar essa dependência é através da especificação da *distribuição de probabilidade conjunta* entre sinal e fator de chance. No

³³ Exs.: poço vizinho pode revelar sucesso (encontrou petróleo) ou insucesso; sísmica 3D pode indicar sucesso (indicação de presença de rocha selante) ou insucesso.

caso, a distribuição bivariada de Bernoulli, será estudada a seguir. Essa seção será importante para entender a construção de processos de revelação de Bernoulli, uma classe *especial* de experimentos *dependentes* de Bernoulli.

Os três parâmetros usados para definir a distribuição bivariada de Bernoulli são os dois parâmetros das distribuições marginais ou incondicionais (probabilidades de sucesso p e q) e um terceiro parâmetro, que estabelecerá a correlação entre as distribuições marginais. Esse último pode ser, por ex., a probabilidade de sucesso conjunto $p_{11} = \Pr(X = 1 \text{ e } S = 1)$. Posteriormente, p_{11} será substituído por η . A Tabela 1 apresenta a distribuição bivariada de Bernoulli, assim como as suas distribuições univariadas marginais.

Tabela 1 – Distribuição Bivariada de Bernoulli e Distribuições Marginais

		Sinal S (ex.: sísmica)		Distribuição Marginal de X (FC)
		S = 1	S = 0	
Variável X (ex.: fator de chance)	X = 1	p_{11}	p_{10}	p
	X = 0	p_{01}	p_{00}	$1 - p$
Distribuição Marginal de S		q	$1 - q$	

Por conveniência de notação, especialmente no item sobre processos de revelação de Bernoulli, em vez de trabalhar com p será adotada a notação FC_0 para a probabilidade de sucesso do FC inicial (antes da informação), i. é:

$$FC_0 = p \quad ()$$

A distribuição de revelações (distribuição de expectativas condicionais) nesse caso tem dois cenários, por conveniência chamados de FC^+ e FC^- :

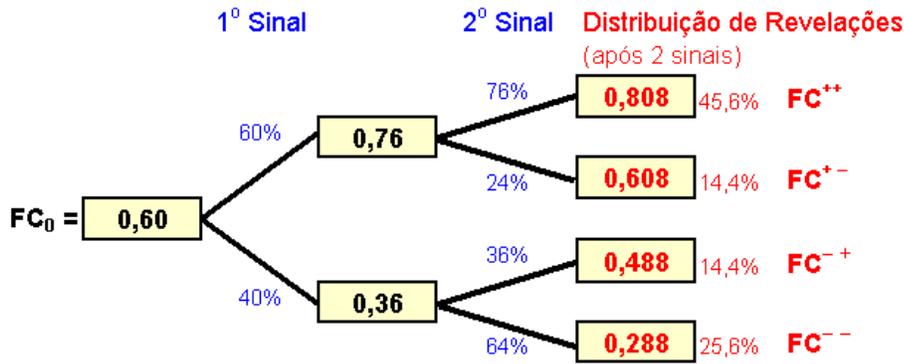
$$FC^+ = \Pr[FC = 1 | S = 1] = E[FC | S = 1] \quad ()$$

$$FC^- = \Pr[FC = 1 | S = 0] = E[FC | S = 0] \quad ()$$

					0.789
				0.754	5.3%
			0.707	7.0%	0.649
		0.636	10.0%	0.592	8.6%
Estimativa Inicial do FC	0.5214	15.6%	0.513	11.7%	0.509
	0.3000	30%	0.396	17.1%	0.430
	100%	0.2051	28.7%	0.319	16.6%
		70%	0.156	26.0%	0.268
			55.6%	0.126	23.6%
				47.0%	0.105
					41.1%
					0.091
					36.7%

probabilidade do cenário da célula acima

Assim, FC_0 evolui para FC^+ ou FC^- , a depender do sinal S. Esses cenários da distribuição de revelações têm probabilidades de ocorrência de q e (1 - q), respectivamente. A Figura 42 mostra a



Raiz positiva da redução

esperada de variância η_k : $\eta_1 = 40\%$

$\eta_2 = 20\%$

distribuição de revelações para um sinal, sendo à esquerda como um processo de evolução

dependente de um sinal e à direita a função probabilidade (massa) da distribuição de revelações de

Bernoulli.

5 – Exploratory Portfolio Example

6 - Conclusion

In this paper we presented.

APPENDIXES

A)

In this appendix we prove.

B) Proof for the

We want to prove the following result of the

Bibliographical References

Arrow, K.J. (1973). Information and Economic Behavior. In *Collected Papers of Kenneth J. Arrow – Vol.4: The Economics of Information*. Cambridge (USA): Belknap Press of Harvard University Press.

Arrow, K.J. (1984): "Collected Papers of Kenneth J. Arrow – Vol.4: The Economics of Information"
Cambridge (USA): Belknap Press of Harvard University Press.

Breiman, L. (1969): "Probability and Stochastic Processes – With a View Toward Applications"
Boston: Houghton Mifflin Co.

Copeland, T. & Antikarov, V. (2001): "Real Options – A Practitioner's Guide". New York: Texere LLC Publishing, 2001.

Cortazar, G. & Schwartz, E.S. & Casassus, J. (2001): "Optimal Exploration Investments under Price and Geological-Technical Uncertainty: A Real Options Model". R&D Management, vol.31(2), pp. 181-189.

Dias, M.A.G. (1997). The Timing of Investment in E&P: Uncertainty, Irreversibility, Learning, and Strategic Consideration. Dallas: *Proceedings of 1997 SPE Hydrocarbon Economics and Evaluation Symposium*, 135-148 (SPE paper 37949).

Dias, M.A.G. (2004): "Valuation of Exploration & Production Assets: An Overview of Real Options Models". Journal of Petroleum Science and Engineering, vol. 44(1-2), October, pp.93-114

Dixit, A.K. & Pindyck, R.S. (1994). *Investment under Uncertainty*. Princeton: Princeton University Press.

Lawrence, D.B. (1999). *The Economic Value of Information*. New York: Springer Verlag.

McEliece, R.J. (2002). *The Theory of Information and Coding*. Cambridge, UK: Cambridge University Press, 2nd Ed.

Shannon, C.E. & Weaver, W. (1949): "The Mathematical Theory of Communication". Urbana and Chicago: University of Illinois Press, 1949.

Trigeorgis, L. (1996). *Real options – managerial flexibility and strategy in resource allocation*. Cambridge, MA: MIT Press.

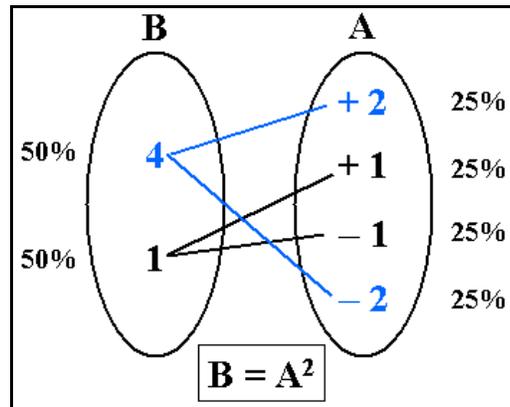


Figure 2 – Example of Asymmetric Learning

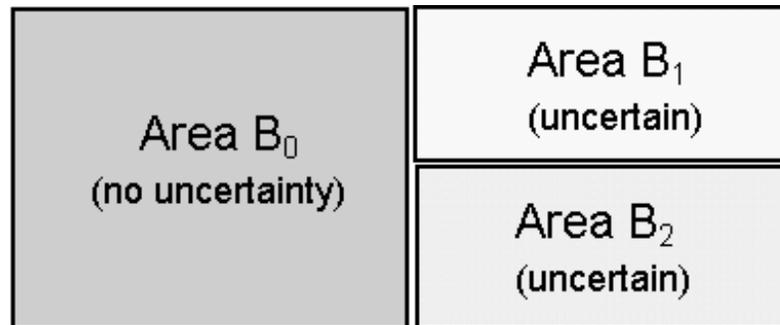


Figure 3 – Oilfield with Two Reservoirs (B₁ and B₂) with Uncertainty