Real Options, Learning Measures, and Bernoulli Revelation Processes

First Version: March 20, 2003. Current Version: June 24th, 2005

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 *exchangeable* bivariate Bernoulli process. Examples in petroleum exploration 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 (internal) Senior Consultant at Petrobras. Adjunct Professor (part-time) at PUC-Rio. 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. Technical uncertainty is related with the project specific characteristics, i.e., the profit function parameters and usually can be reduced by investing in information (exercising *learning options*). For instance, in a R&D of a new machine, there is uncertainty about the MTBF (mean time between fails) and the prototype phase and/or a pilot test phase can reduce this uncertainty.

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, under GBM hypothesis, the variable changes by the simple passage of time, whereas in most real applications the variable with technical uncertainty changes only with the exercise of a learning option. So, 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 on 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 Theorem 1(a) below.

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 sentence holds 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.

[1] The time between events can take months or years. In most cases the events are endogenous in the model (are *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 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 project true value (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 uncertainty[2] defined by:

$$\mathbf{H(X)} = -\sum_{i} \mathbf{p(x_i)\log[p(x_i)]}$$

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 . log(0) = 0. In case of no uncertainty (here named *full revelation*) 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 measured in "bits".

uncertainty. But in *valuation* problems where the *magnitude* of losses or gains matters[3], 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 finance[4] 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 Π_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? Technical uncertainty has 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 Π_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 probability of either *over investment* (if the true scenario is 200) or *under*

[3] This is not a problem in communication channel design 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.

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 very convenient mathematical properties for learning measure applications. The measure η^2 obeys a set of reasonable *axioms* for probabilistic *learning measures*, whereas the popular *information likelihood* doesn't obey these axioms. Likelihood is useful in many statistical cases, but it is not so useful in VOI applications, contrary to conventional wisdom. This paper defends the measure η^2 replacing likelihood and others concepts in VOI applications, especially in the dynamic real options setting.

Motivated by oil exploration applications, this paper studies a special *technical uncertainty process* generated by a *sequence of learning options exercises*: the *Bernoulli revelation process*, which is a sequence of bivariate Bernoulli distributions. The chance factor (CF) of oil existence in a prospect is a variable with technical uncertainty that is very important for exploratory investment decisions. CF is a Bernoulli random variable, which has two scenarios: 1 (success, existence of oil with probability p) and 0 (failure, with probability 1 – p). The exercise of a learning option, such as the drilling of a correlated prospect, is the *signal* S, which is also a Bernoulli random variable. The dependence degree between the these two Bernoulli variables (CF and S) must be studied in the context of bivariate Bernoulli distribution and this paper shows that the (positive) square root of η^2 (i.e., η) is an adequate dependence measure for learning purposes. In this specific case of bivariate Bernoulli, η is equal to the correlation coefficient. Using this measure, the paper studies some *Bernoulli revelation process*, highlighting the exchangeable bivariate Bernoulli distributions with recombining scenarios, which resemble the *binomial processes* used in discrete option pricing.

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 *portfolio* 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 the concluding remarks and some suggestions for further research.

2 – <u>Portfolio of Real Assets: A Simple Motivating Example</u>

Here and in most of this paper, consider <u>only technical uncertainty</u> and investment *optionality*, for expositional clarity. However, the paper will also present 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:

$$\mathbf{EMV = -I_W + [CF\,.\,NPV]} \qquad \textbf{(1)}$$

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 with the simplest probability distribution – the Bernoulli distribution, which has two scenarios (1 = success and 0 = failure) and one parameter (p) named success probability. So, we use CF ~ Be(p) to denote this Bernoulli distribution. The expected value of a Bernoulli distribution is the success probability, i.e., E[CF] = p. 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[CF] = p = 30% and NPV = 95 million \$ for both prospects. So, the EMV is negative:

$$EMV_1 = EMV_2 = -30 + [0.3 \text{ x } 95] = -1.5 \text{ million \$}$$

Apparently this two real assets portfolio is worthless. Indeed, <u>if</u> the prospects in this portfolio were *independents*, the two-prospects portfolio value would be zero. However, we will see that the

[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. Here we can also imagine that the drilling option is expiring, so that the option to wait values zero.

portfolio value can be strictly positive if they are dependent. Suppose that these two exploratory prospects are in the same geologic play[7], so that the prospects are *dependent* with positive correlation. If these prospects have positive correlation, in case of success in one prospect, the success probability p from the second prospect chance factor (CF_2) *must be* revised upward (to CF_2^+) and in case of failure p must be revised downward (to CF_2^-). **Figure 1** illustrates this learning updating with the information revelation generated by the first option exercise.

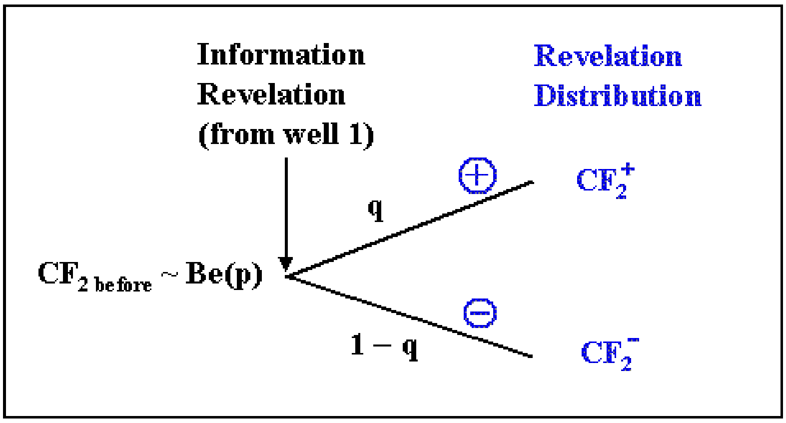

Figure 1 – Effect of the Well 1 Signal on the Chance Factor CF_2

After the signal S_1 (information revelation with the well number 1 drilling), Figure 1 shows two updated scenarios for the variable of interest CF_2: the good news case, $CF_2^+ = E[CF_2 \mid S_1 = CF_1 = 1]$, and the bad news case, $CF_2^- = E[CF_2 \mid S_1 = CF_1 = 0]$, so we have a simple two-scenario distribution of *conditional expectations*, where the conditioning is the information revelation. The distributions of conditional expectations are here named *revelation distributions*, and a set of properties for these distributions will be presented soon.

The intensity of this CF_2 updating process 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 (r.v.) 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, given by 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

[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.

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:

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

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. Because of the assets optional nature, the value of this portfolio is higher as higher is the dependence between the prospects. Hence, the real option value a portfolio of assets with technical uncertainty is an increasing function of the dependence degree between these assets, and the study of learning measures based in the dependence degree between r.v. is demanded.

Note that in this paper the role of dependence is very different of the traditional portfolio theory for financial assets (Markowitz): here there is information revelation by *sequential exercise of learning options*, an active exploitation of dependence, whereas in financial portfolio theory the role of dependence is only for *diversification* purposes. Note that, for learning purposes, is not particularly relevant the dependence *direction* (if positive or negative), whereas for diversification purposes low (or even negative) dependence is much better than high/positive dependence. In order to see this, imagine in Figure 1 that the signal S_1 has negative dependence with CF_2, but with the same intensity so that the revealed scenarios CF_2^+ and CF_2^- are *only permuted*[8]. The learning with this signal is exactly the same because applying the option rule Max(EMV, 0) results in the same portfolio value. Hence, insights from Markowitz' theory is not applicable in portfolio of real options with technical uncertainties, even more for the role played by the measure of dependence between the uncertainties.

Valuation of real assets portfolio *per se* justifies the study of learning measures. But, there are other relevant applications like option games, as in Dias & Teixeira (2004). In order to illustrate this point, instead a portfolio owned by a single firm, suppose in the Figure 1 example that firm 1 owns the prospect 1 and firm 2 owns the 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 *bargain game* where again the surplus to be divided is given by the effect of

[8] Assuming that the probability of good news for CF_2 (here $CF_1 = 0$) remains the same, i.e, $1 - q = 30\%$ in this case.

the information revelation on the prospects, which depends on degree of dependence (or learning degree) between the prospects. In the first case, the free rider works with *public* information, whereas in the bargain game the players work with *private* information. Public information is only a subset of private information and is necessary to differentiate these games with a good learning measure, as done in Dias & Teixeira (2004).

So, in order to quantify interesting real options and option games problems is necessary to study the learning degree of a signal over a variable of interest. This is done in the next two sections.

3 – Revelation Distribution and Learning Measures

In this section are presented a series of definitions, lemmas, propositions, theorem and a list of axioms for probabilistic learning measures.

Definition. **Prior distribution**: it is a probability distribution that represents all prior knowledge that the decision maker knows about one r.v. (based in Lawrence, 1999, p.5). The prior distribution support[9] includes all possible values (scenarios) that this variable can assume, while the probability density represents the best estimative of the occurrence probability for these scenarios, using the current knowledge (prior information). Notation: p(x) is the prior distribution for the r.v. X.

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

$$\mathcal{I} = \{ \mathbf{S}, \mathbf{p(x, s)} \} \tag{2}$$

The joint measure is the *joint probability distribution* of two r.v., p(x, 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 is a classical statistics theme from the *comparison of experiments* literature (Blackwell, 1951). Arrow (1992, p.169) notes that could be useful even a partial ordering for the signals independently of the specific decision problem. In our setting, we will include the proposed learning measure η^2 into the information structure $\mathcal{I}$ plus one additional condition, in order to replace the joint probability distribution input. The additional condition will depend on the kind of problem (flexible information structure), but with our learning measure η^2 into

[9] The support of a distribution p(x) is the set of values where p(x) > 0.

the information structure for all applications of our interest, we can perform the comparison of information structures: for the same prior distribution, higher η^2 means more informative structure.

The proposed learning measure η^2 is related with the concept of *revelation distribution* presented in Dias (2002). Revelation distribution is a distribution of conditional expectations where the conditioning is the information (signal) revealed by the exercise of a learning option. The term "revelation" emphasizes a process towards the true value of variable with technical uncertainty, and it has been used in related literature (eg., Grenadier, 1999; Childs et al., 2001) and before in the classic economics of information literature (eg., Wilson, 1975, p. 186). This term suggest a learning process to find out the true state of nature. Denote the r.v. associated with the revelation distribution by $\mathbf{R_X(S) = E[\,X \mid S\,]}$, where X is the *variable of interest* with technical uncertainty (e.g., chance factor of an oil prospect; reserve volume of a new oilfield) and S is the *signal* (e.g., the drilling outcome from a correlated oil prospect; the information generated by an appraisal well in a new oilfield).

Definition. **Revelation process**: is the sequence of r.v. $\{R_{X,1}, R_{X,2}, R_{X,3}, \ldots\}$[10] generated by a sequence of signals S_1, S_2, S_3, … about an interest variable X, which its main characteristic is the expected reduction of uncertainty provided by these signals. Revelation process is a *probabilistic learning process.* In the mathematical literature is sometimes referenced as "*accumulating data about a r.v.*" (Williams, 1991, p.96) or as *Doob-type martingale* (see Ross, 1996, p.297).

Revelation processes can be considered as stochastic processes, but in general indexed by events and not by time, as in most stochastic processes. This paper is interested mainly in processes with events being exercise of learning options, in order to model the technical uncertainty evolution (expected reduction of uncertainty) with the investment in information process. One example of revelation process indexed by *events* is the sequential drilling from the appraisal phase after an oilfield discovery, which reduces the technical uncertainty about the reserve volume of this field. An example of revelation process indexed by *time* is the one that generally occur with the stock return of new firms in the market, the IPO (Initial Public Offering). The stock volatility in this case is generally very high in the beginning, but with the passage of time this volatility is reduced (but never to zero) as the investors learn about the firm capability to generate return to their stockholders. In this case, we have *revelation process only during a temporal transient* with diffusion of new information to the market players. After this transient, the market can be considered efficient to price this stock,

[10] We could define this process as a sequence of probabilistic moments from posterior distributions, with the conditional expectation distribution being a particular moment. But hardly this definition could be so useful as here proposed.

ceasing the revelation (reduction of uncertainty) process. Example of process that is not revelation process is the Itô diffusion process (like geometric Brownian or mean reversion processes).

Definition. **Full revelation of X**: is the revelation of a scenario c so that Pr(X = c) = 1, where c is a constant belonging to p(x) support. In general terms, if the available information is given by the sub-sigma-algebra Ψ, full revelation of X means that X is Ψ–measurable and, hence, we can write $E[X \mid \Psi] = X$ almost surely (a.s.)[11]. Intuitively, it means that there is *perfect information* about the true state of nature for the variable X.

We will see that *any* revelation process converges[12] to an integrable r.v. denoted by X_∞ for $n \to \infty$, where n is the number of (relevant) signals. But not always converges to the full revelation limit, that is, the convergence is not always $X_\infty = X$. Mathematically, a revelation process is the *Doob´s process* (Karlin & Taylor, 1975, p.246 and 295). Note that *not* all r.v. sequence converges to an *integrable* r.v. X_∞. All revelation processes converges because the process $\{R_{X,1}, R_{X,2}, \ldots R_{X,n}\}$ is *uniformly integrable* (see Appendix A for the definition). Using this, Lemma 1 below shows that this implies that the $\{R_{X,n}\}$ sequence converges to X_∞ when $n \to \infty$.

Lemma 1: Let $\{R_{X,1}, R_{X,2}, \ldots R_{X,n}\}$ be a revelation process, i.e., $R_{X,k} = E[X \mid \Im_k]$ are defined in the same probability space (Ω, Σ, P), being X integrable. $\Im_k$ is a *filtration*[13] $\{\Im_k: k \geq 0\}$, with $\Im_k$ being generated by the signals sequence $\{S_k\}$. Then, the revelation process is uniformly integrable and, hence, when $n \to \infty$, there exists a.s. a limit of $R_{X,n}$ in $\boldsymbol{L}^1$ (i.e., in mean) that is an integrable r.v. denoted by X_∞, that is also a conditional expectation, i.e.:

$$\mathbf{lim}_{n \to \infty} \mathbf{R}_{X,n} = \mathbf{X}_\infty = \mathbf{E}[\mathbf{X} \mid \mathbf{S}_1, \mathbf{S}_2, \ldots] = \mathbf{E}[\mathbf{X} \mid \Im_\infty] \qquad \mathbf{(3)}$$

Proof: First must be proved that *any* revelation process is a martingale. Theorem 1(d) below shows this. The proof that uniform integrability is *sufficient* for the martingale convergence in $\boldsymbol{L}^1$ is given by the famous *Doob´s Martingale Convergence Theorem*[14] (see, e.g., Brzezniak & Zastawniak, 1999, theorem 4.2, p.71-73). The proof that the revelation process is uniformly integrable is done, e.g., in Ross (1996, p.319) or in Karlin & Taylor (1975, p.295-296) for Doob-type martingale (and so for

---

[11] In addition, algebraic operations like sum, product and division, don't destroy measurability (Gallant, 1997, p.47).

[12] It converges almost surely (with probability 1), which implies that converges in probability, which also implies that converges in distribution (Karlin & Taylor, 1975, p.18).

[13] We can interpret the filtration $\Im_n$ generated by a sequential information process $\{S_1, S_2, \ldots S_n\}$ as a set with all available information in the stage n. In technical terms, it is a increasing family of sub-sigma-algebras generated by the information revelation, e.g., $R_{X,2} = E[X \mid S_1, S_2]$.

[14] This theorem has been used to prove other theorems like the Komolgorov 0-1 law and Kalman filter equations. Here we exploit other application for this famous theorem.

revelation process). Hence, there is a limit given by an integrable r.v. X_∞. The proof that this limit is a conditional expectation $E[X \mid \Im_\infty]$ is given by Karlin & Taylor (1975, p.310)[15]. □

As the revelation process *always* converges to some X_∞, is redundant the qualifier "convergent" to this process, and we will use this qualifier only for the *full revelation* convergence, i. e., when $X_\infty = X$ (or $Var[X \mid S_1, S_2, \dots, S_n] \rightarrow 0$ when $n \rightarrow \infty$). One practical example of *convergent* revelation process is the drilling of appraisal wells in order to reduce the uncertainty about the oil-in-place volume from one oilfield. It is convergent because if we drill a very large (infinite) number of wells we get the true value of this volume. An example of non-convergent process is the return of IPOs.

Theorem 1 describes the 4 main revelation distribution properties (distribution of the r.v. R_X). These properties are: the R_X mean, the R_X variance, R_X in the limit case of full revelation and the martingale property for revelation processes (R_X sequences).

Theorem 1 (Revelation Distributions): Let the r.v. X and S two-times integrable (i.e, finite mean and finite variance) defined in the probability space $(\Omega, \Sigma, \mathbb{P})$. The interest variable X has prior distribution p(x). The signal S generates the sigma-algebra Ψ, a sub-sigma-algebra of Σ, i.e., $\Psi \subseteq \Sigma$. Let $p(R_X)$ be the probability density of $R_X = E[X \mid S]$, i.e., the revelation distribution of X given S. Then, the revelation distribution is almost defined[16] by the following properties:

(a) In the limit case of full revelation, the variance of any posterior distribution is zero and the revelation distribution $p(R_X)$ is equal to prior distribution p(x).

(b) The revelation distribution mean is equal to the prior mean of X, i. e.:

$$\mathbf{E[R_X] = E[X]} \quad \mathbf{(4)}$$

(c) The revelation distribution variance is simply the *expected variance reduction* of X caused by the signal S, i.e., the prior variance less the expected posterior variance:

$$\mathbf{Var[R_X] \;=\; Var[X] \;-\; E[\, Var[X \mid S\,]\,]} \quad \mathbf{(5)}$$

(d) Consider a sequential exercise of learning options generating the signals $S_1, S_2, S_3, \dots$ and the r.v. $\{R_{X,n}\} = \{E[X \mid S_1, S_2, \dots S_n]\}$, n = 1, 2, … Then, the *revelation process* $\{R_{X,1}, R_{X,2}, R_{X,3}, \dots\}$ is a martingale.

[15] A simple proof: note that $R_{X,n}$ is a function of $S_1, S_2, \dots S_n$, so that the limit X_∞ is a measurable function of this sequence of signals. Hence, it is measurable with respect to $\Im_\infty$.

[16] Definition: almost defined distribution is a distribution that we know at least the *mean*, the *varianc*e and that belongs to a *sequential process* of distributions with *known initial distribution* and *convergent to a known distribution*.

Proof: See Dias (2002).

Some highlights: (a) Note that Lemma 1 guarantees that always exists a limit with probability 1 and note that **$\Pr(X = c) = 1 \Leftrightarrow \text{Var}[X] = 0$** (see DeGroot & Schervish, 2002, theorem 4.3.1, p.198). The remaining of the proof is given by the prior distribution definition itself. This property also claims that revelation process variance is bounded by the prior distribution variance. (b) This property is known as *law of iterated expectations*. It can be formulated in a more general fashion by using the sub-sigma-algebra Ψ (instead the r.v. S): if R_X is any *version*[17] of $E[X \mid \Psi]$ then $E[R_X] = E[X]$, a.s. (c) This property is the heart of Theorem 1 and we will see that it is linked with the proposed learning measure η^2. Variance of R_X is very practical in the context of technical uncertainty because the revelation distribution variance is analogous to the role played by volatility in the classical (market uncertainty) real option problem: as higher is the revelation distribution variance as higher is the learning option value. Note also the consistency between (c) and (a): in the full revelation limit, $E[\text{Var}[X \mid S]] = 0 \Rightarrow \text{Var}[R_X] = \text{Var}[X]$. (d) This property is useful to study revelation processes and it points that the expected values are the same for all the revelation distributions in this sequence.

Note that revelation distribution does not require risk-adjustment to use the risk-neutral approach (as in case of market uncertainty distributions) because technical uncertainty does not demand risk-premium from diversified investors. So, *revelation distributions are naturally risk-neutral*. With the Theorem 1 properties, Dias (2002) combines revelation distributions with risk-neutral (or adjusted) stochastic process into a Monte Carlo framework in order to solve the following real option problem on oilfield development decision. An oil company has an undeveloped oilfield with still relevant remaining technical uncertainty about the oil reserve volume (B) and quality (q). In addition, we have market uncertainty on oil prices (P) and on development investment (D). There are many alternatives of investment in information (vertical appraisal well, horizontal appraisal well, long-term production test, pilot test, etc.) with different learning costs, different times to learn, and different learning intensities (given by the revelation distribution *variances*). For each alternative of investment in information, we can run a Monte Carlo simulation as illustrated in Figure 2.

[17] If R^*_X is a version of R_X, then $R^*_X = R_X$ almost surely (Williams, 1991, p.84).

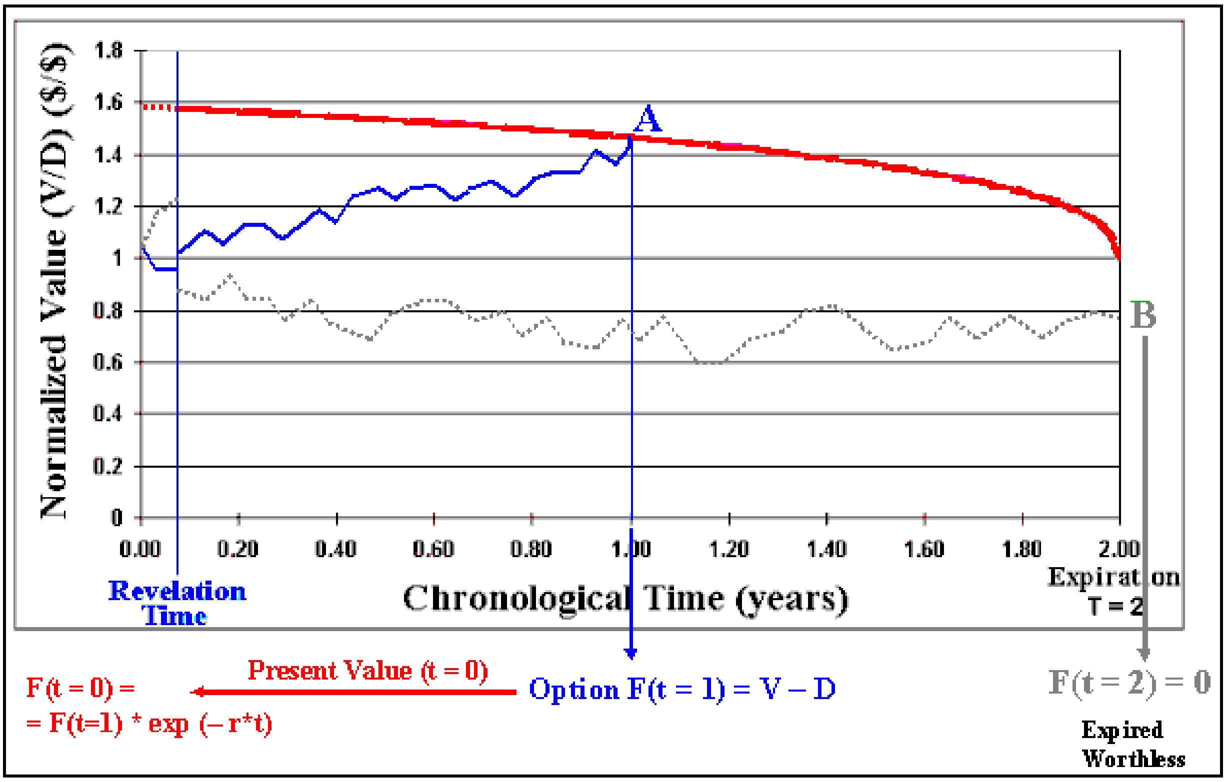

Figure 2 – Practical Use of Revelation Distributions in Dias (2002)

In case of *development* option exercise, we get the net present value: NPV = V – D = q B P – D. Figure 2 shows the (normalized) development option threshold (red line) where is optimal the exercise of development option. It also shows two sample-paths generate by Monte Carlo simulation. Between t = 0 and t = "revelation time", the normalized project value V/D oscillates due to both the oil price P and the investment D market uncertainties. At t = "revelation time", we get new information about q and B, updating the expectations E[q] and E[B], so that the jump size in V/D is caused by jumps in these expectations, which are drawn from the revelation distributions $p(R_q)$ and $p(R_B)$. After running many interactions in this Monte Carlo framework, we get the option value F. By subtracting the learning cost, we get the real option value for this oilfield using one specific learning alternative. We repeat the simulation (but with different time to learn, different revelation distributions, etc.) for the other learning alternatives. The alternative with the higher value from these simulations is the best learning option alternative. This apparent complex problem (compound options: learning and development options; 5 state variables: P, D, q, B and t) can be solved easily with a simple Excel spreadsheet plus a Monte Carlo simulation device in seconds or few minutes.

In this problem, for each learning alternative k (that generates a signal S_k), Dias (2002) used the following flexible information structure that simplified the computational task:

$$\mathcal{I}_k = \{\eta^2(q \mid S_k), \eta^2(B \mid S_k), A\} \qquad (6)$$

Where A is the assumption "the revelation distributions of *q and B are approximately*[18] *of the* *same type (shape) of the limit case of full revelation*", e.g., if the prior distribution of B is lognormal, we assume that the revelation distributions for B are also lognormal. This is totally true only for the full revelation limit, where the revelation distribution is equal to the prior distribution. With this information structure plus the prior distributions of q and B, we have all the input required for probabilistic modeling of learning options. Note that, in this problem, we don't need the distribution of S nor the joint distribution p(x, s) to solve our real option problem. However, for the discrete Bernoulli revelation processes we'll use other kind of information structure, specifying η^2 and the distribution of S (replacing the assumption "A" in eq. 6). In all cases, the learning measure η^2 is used because it defines the learning potential and it has nice/convenient properties, as we will see below. We study η^2 now, starting with some definitions and the main properties.

Definition. **Learning measure η^2**: Consider two r.v. X and S with finite variances, defined in the same probability space $(\Omega, \Sigma, \mathbb{P})$. The expected percentage of variance reduction of X given S is:

$$\eta^2(X \mid S) = \frac{\text{Var}[X] - E[\,\text{Var}[\,X \mid S\,]\,]}{\text{Var}[X]} \qquad (7)$$

The notation η^2 is adopted due to two reasons: (a) it eases the connection with the statistical interpretation of η^2, namely the *correlation ratio*[19], also known by "*eta-squared*" in some statistical books; and (b) in some situations (e.g.: Bernoulli processes) is more intuitive the positive *root* of η^2, i.e., simply η, as we will see. By applying Theorem 1(c) we get:

$$\eta^2(X \mid S) = \frac{\text{Var}[\,E[\,X \mid S]\,]}{\text{Var}[X]} = \frac{\text{Var}[R_X]}{\text{Var}[X]} \qquad (8)$$

That is, the proposed learning measure is the normalized revelation distribution variance, being normalized by the initial variance (i.e., the prior distribution variance). Because variance is a non-negative number, eq. (8) shows that η^2 is always positive or zero. Eq. (5) (or eqs. 7 and 8) shows that, in average, the posterior distribution variance never grows, i.e.,

[18] With the exact mean and variance, the error caused for a non-exact shape is generally of second order. But this approximation is good for continuous distributions, but not for discrete distributions (we'll use other structure).
[19] The famous statistician Karl Pearson introduced the correlation ratio in 1903. Kolmogorov, 1933, p.60, linked this concept with the conditional expectations concept.

$$\mathbf{Var[R_X] \geq 0 \;\Rightarrow\; E[\,Var[X \mid S\,]\,] \leq Var[X]} \qquad \textbf{(9)}$$

This learning measure is asymmetric, i.e., $\boldsymbol{\eta^2(X \mid S) \neq \eta^2(S \mid X)}$. This is an *advantage* of η^2. 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. For the discrete case the concept of expected entropy is defined as:

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

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 (here measured by 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), and is defined for the discrete case by:

$$\mathbf{I(X; S) \;=\; H(X) - H(X \mid S)} \qquad \textbf{(10)}$$

This is perhaps the most important concept from information theory. Note that eq. (10) resembles our Theorem 1(c) (eq. 5), but with entropy replacing variance. Let us see a classic numerical example[20]. Consider two r.v. A and B, with A assuming one value from the set {– 1, +1, – 2, + 2}, each scenario with probability ¼, whereas B = A^2. Figure 3 shows this example.

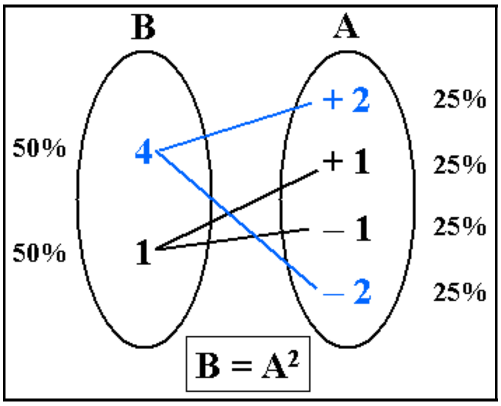

Figure 3 – Example of Asymmetric Learning

[20] This example appeared in Feller (1968, p.236) to show the inadequacy of correlation coefficient to express dependence in the general case (nonlinear relations between r.v.). McEliece (2002, p.23-24; 45) uses the same example to show the superiority of mutual information over the correlation coefficient. This paper uses this same example, but to show the asymmetric measure η^2 superiority over symmetric measures like mutual information!

There is an obvious dependence between the r.v. A and B (there is a function linking them!), but the correlation coefficient is zero. McEliece (2002) pointed the superiority of the metric based in entropy, because the mutual information I(A, B) is different of zero: I(A, B) = 1 bit. However, this metric is symmetric, I(A, B) = I(B, A). For learning purposes, if we know the value of A, this information reveals all the truth about the value of B. Our metric shows that: $\eta^2(B|A) = 100$ % (full revelation case). However, if we know the value of B (e.g. B = 4), we still don't know the value of A (e.g., either A = + 2 or A = – 2, this uncertainty is even worse than the initial case for mean-variance optimizers in the classic finance). In this case is necessary to use some measure that considers this asymmetry. Our metric captures this learning asymmetry: $\eta^2(A|B) = 0\ \% \neq \eta^2(B|A) = 100\ \%$.

So, this example showed that we could be interested in asymmetric effects from the interaction of two r.v. But, for some 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 see how S reduce the variance of X or how valuable is S for X. It doesn't matter the opposite, i.e., 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 r.v. X and S. A learning measure that captures this asymmetric interest must be asymmetric in general and symmetric only in particular cases. In this way, we can penalize many dependence measures as candidate to learning measures. One example is the increasing popular "copula" (see a good explanation at Nelsen, 1999), which is not adequate candidate for learning measure because is always symmetric and it is not directly applicable to discrete distributions.

We'll prove that $\eta^2 \in [0, 1]$. This means that as measure η^2 doesn't concern with the learning *direction*, i.e., if either the dependence is positive or negative[21]. This is an *advantage*[22] in our context because we are interested in learning, in improving our knowledge over X by using the information from the signal S. Remember Figure 1 example discussion, a negative signal but with the same intensity (same η^2) just permutes CF_2^+ and CF_2^- with the portfolio value remaining the same (same learning effect). The example below also shows this point and permits an intuitive discussion of the

[21] Metrics that allow negative values, like the correlation coefficient, are not measures.
[22] This is a disadvantage in some applications like *finance* portfolio theory: in order to reduce the portfolio variance is better negative than positive correlations. So this positive/negative distinction is relevant in same applications.

inadequacy, for learning purposes, of metric based in *likelihood* of information S about the variable X, i.e., a metric related with the inverse probability p(s | x).

<u>Example</u>: Two experts who own "infallible crystal balls", know all the truth about the next day performance of stock X (if will go up or down) in the stock exchange market. An investor want to buy the expert advise and prior this advice there are 50% chances for each scenario. 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. By the point of view of *reliability of the information*, i.e., the *likelihood* of the information L(S), this metric assumes <u>two different values for the same knowledge</u> (full revelation of X), i.e., $L(S_1) = p(S_1 = a \mid X = a) = 100\%$ and $L(S_2) = p(S_2 = a \mid X = a) = 0\%$, a = up or down. In addition, L(S) metric set the value 0 for the full revelation case in this example! So, despite its wide use in VOI/economics of information literature, metrics based in likelihood don't look adequate for learning purposes. In this example, it is easy to see that our proposed measure attributes the same value for signals that result in the same learning, i.e., $\eta^2(X \mid S_1) = \eta^2(X \mid S_2) = 1$, because the posterior variance (after either S_1 or S_2) goes to zero.

In Dias (2002) example, we could find by simulation the VOI for different learning intensities, i.e, different η^2 values. Figure 4 shows these simulations results (VOI is before the cost of learning subtraction), which shows a nonconcavity for low values of η^2 and a rough linear behavior.

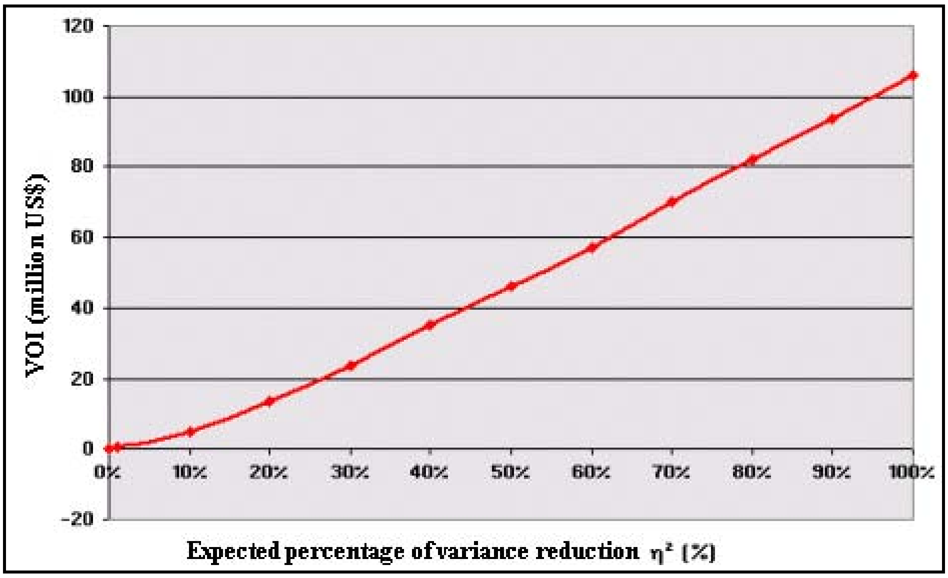

Figure 4 – Value of Information x η^2 in Dias (2002)

The nonconcavity behavior for low learning intensities is consistent with a classic paper of Radner & Stiglitz (1984). Because it is much easier to work with perfect information (full revelation, η^2 = 100%) in VOI problems and for no learning ($\eta^2 = 0$) the VOI is zero, in some practical cases the linear approximation VOI(η^2) can be useful for fast calculation. Now, we set some key η^2 properties.

Proposition 1: Let X and S be two non-trivial r.v.[23] with finite variances, defined in the same probability space (Ω, Σ, $\mathbb{P}$). Consider the learning measure $\eta^2(X \mid S)$ defined by eq. (7). Then, this measure has the following properties:

(a) The measure $\eta^2(X \mid S)$ always exists;

(b) The measure η^2 is, in general, asymmetric, i.e., $\eta^2(X \mid S) \neq \eta^2(S \mid X)$

(c) The measure η^2 is normalized in unit interval, that is[24],

$$\mathbf{0 \leq \eta^2 \leq 1} \qquad \mathbf{(11)}$$

(d) If X and S are independent r.v., then η^2 is zero:

$$\textbf{X and S independent} \Rightarrow \eta^2(\mathbf{X} \mid \mathbf{S}) = \eta^2(\mathbf{S} \mid \mathbf{X}) = \mathbf{0} \qquad \mathbf{(12)}$$

In addition, η^2 is zero if and only if the revelation distribution variance is zero:

$$\eta^2(\mathbf{X} \mid \mathbf{S}) = \mathbf{0} \Leftrightarrow \mathbf{Var[R_X(S)] = 0} \qquad \mathbf{(13)}$$

(e) $\eta^2(X \mid S) = 1 \Leftrightarrow$ exists a real function, the r.v. g(S), so that X = g(S);

(f) The measure $\eta^2(X \mid S)$ é invariant under linear transformations of X, i.e., for any real numbers a and b, with $a \neq 0$, we have:

$$\eta^2(\mathbf{a\,X + b} \mid \mathbf{S}) = \eta^2(\mathbf{X} \mid \mathbf{S}) \qquad \mathbf{(14)}$$

(g) The measure $\eta^2(X \mid S)$ is invariant under linear and nonlinear transformation of S if the transformation g(S) is a 1-1 function (invertible function).

$$\eta^2(\mathbf{X} \mid \mathbf{g(S)}) = \eta^2(\mathbf{X} \mid \mathbf{S}), \qquad \text{g(s) is invertible} \qquad \mathbf{(15a)}$$

In general, for any g(S) measurable by the sigma-algebra generated by S, then the inequality below holds:

[23] Non-trivial means strictly positive variances. Proposition 1 is valid almost surely (with probability 1).
[24] We could also highlight that η^2 is a *truly* measure, because $\eta^2 \geq 0$.

$$\eta^2(\mathbf{X} \mid \mathbf{g(S)}) \leq \eta^2(\mathbf{X} \mid \mathbf{S}) \text{ , with equality if g(s) is invertible} \quad \textbf{(15b)}$$

(h) If the r.v. $Z_1, Z_2, \dots$ are independent and identically distributed (iid) and if $S = Z_1 + Z_2 + \dots + Z_j$ and $X = Z_1 + Z_2 + \dots + Z_{j+k}$ for any non-negative integers j and k, with $j + k > 0$, the proposed measure $\eta^2(X \mid S)$ is given directly by:

$$\eta^2(\mathbf{X} \mid \mathbf{S}) = \frac{\mathbf{j}}{\mathbf{j+k}} \quad \textbf{(16)}$$

Proof: Appendix B.

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 5.

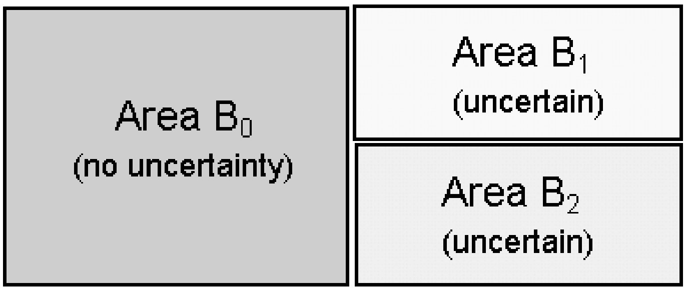

Figure 5 – Oilfield with Two Reservoirs (B_1 and B_2) with Uncertainty

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:

$$\mathbf{B(S_1, S_2) = B_0 + (B_1 \times S_1) + (B_2 \times S_2)} \quad \textbf{(17)}$$

Where the signals are independent Bernoulli random variables, $S_1 \sim Be(q)$ and $S_2 \sim Be(p)$. For Bernoulli variables we'll see that independence means (iff) that $\eta^2(S_2 \mid S_1) = \eta^2(S_1 \mid S_2) = 0$. Here we are interested in $\eta^2(B \mid S_1)$ and $\eta^2(B \mid S_2)$, i.e, 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 (prior) variance is $Var[B] = 1250$. It is easy to calculate that the revelation of the signal S_1 (by drilling the first well) reduces Var[B] by the half, i.e., $\eta^2(B \mid S_1) = 50\%$. Similarly $\eta^2(B \mid S_2) = 50\%$. The example most interesting feature is:

$$\boldsymbol{\eta^2(B \mid S_1) + \eta^2(B \mid S_2) \;=\; \eta^2[f(S_1, S_2) \mid S_1] + \eta^2[f(S_1, S_2) \mid S_2] \;=\; 1} \qquad \textbf{(18)}$$

It is not a coincidence nor because B is linear in S_1 and S_2. We'll see that this relation is much more general, valid for *any function* and for $n > 0$ independent variables. This η^2 property <u>is not verified</u> by the "competitors" coefficient of correlation $(0.71 + 0.71 > 1)$[25] or mutual information $(1 + 1 > 1)$[26]. 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 valuable than the signal S_2 in terms of reducing uncertainty about the volume B. If $B_1 = 100$ (remaining $B_2 = 50$, etc.), the η^2 property (eq. 18) remains valid but with higher weight to signal S_1, i.e., $\eta^2(B \mid S_1) + \eta^2(B \mid S_2) = 0.8 + 0.2 = 1$.

The key concept used in the eq. (18) 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 2: Let the signals $S_1, S_2, \ldots, S_n$, be independent random variables. Then:

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

(b) For *any* functions f, g, … , h, the variables $f(S_1), g(S_2), \ldots, h(S_n)$, 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.

<u>Theorem 2</u> (Learning measure decomposition): Let the signals $S_1, S_2, \ldots, S_n$, be independent random variables. We want to learn about X, a random variable with $Var[X] > 0$. Assume as finite all the relevant expectations and variances. Let $X = f(S_1) + g(S_2) + \ldots + h(S_n)$, where f, g, . . . , h, are any real valued functions. Then:

---

[25] In this simple example, because the function $B(S_1, S_2)$ is *linear*, we'll see that the correlation-square ρ^2 is equal to η^2.

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

$$\eta^2(X \mid S_1) + \eta^2(X \mid S_2) + \ldots + \eta^2(X \mid S_n) = 1 \quad \textbf{(19)}$$

Proof: Appendix C.

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(RF, GV, NTG, ϕ, Sw, Bo) = RF x [GV x NTG x ϕ x (1 − Sw)] / 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 formation factor. Another function is B(RF, A, h, ϕ, Sw, Bo) = RF x (A x h x ϕ x (1 − Sw)] / 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 2 holds for X.

In the previous example (Figure 5) with S_1 and S_2 independent and B = f(S_1) + g(S_2), this theorem is verified: $\eta^2(B \mid S_1) + \eta^2(B \mid S_2) = 0.5 + 0.5 = 1$. 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, eq. (19) 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 $\eta^2(B \mid S_1) > 50\%$. The same reasoning is valid to S_2, so that also $\eta^2(B \mid S_2) > 50\%$. Hence, we shall expect that $\eta^2(B \mid S_1) + \eta^2(B \mid S_2) > 1$. This shows that, although each *individual* signal has higher revelation power with the dependence hypothesis (signals are more valuable), when 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.

Now is presented a set of axioms (desirable properties) for probabilistic learning measures. It is inspired in the famous axiom list of Rényi (1959) for probabilistic dependence measures. Here our focus is learning measures, whereas Rényi thought in applications like distance from independence.

Axioms for Probabilistic Learning Measures: The following axiom list gives the desirable properties for a probabilistic learning measure denoted by M(X | S):

A) M(X | S) shall exist at least for all *non-trivial* r.v. X and S and with *finite* uncertainty;

B) M(X | S) shall be, in general, asymmetric;

C) M(X | S) shall be <u>normalized</u> in the unit interval in order to ease the interpretation, i.e.,

$$\mathbf{0 \leq M(X \mid S) \leq 1} \quad \mathbf{(20)}$$

D) If X and S are <u>independent</u> $\Rightarrow$ M(X | S) = M(S | X) = 0, because there is no probabilistic learning. In addition,

$$\mathbf{M(X \mid S) = 0 \;\Rightarrow\; \underline{zero\ learning}} \quad \mathbf{(21)}$$

Where "zero learning" can occur not only for the case of independence. The *learning* concept is defined in the specific measure, but its sense must be *invariable* (eg., the measure of uncertainty shall be the same for all applications);

E) In case of <u>functional dependence</u>, M(X | S) shall be maximum, i.e., for any real function f(.):

$$\mathbf{X = f(S) \;\Rightarrow\; M(X \mid S) = 1} \quad \mathbf{(22)}$$

In addition:

$$\mathbf{M(X \mid S) = 1 \;\Rightarrow\; \underline{maximum\ learning}} \quad \mathbf{(23)}$$

Where "maximum learning" means that is not possible to learn more about X, and *learning* concept is defined in the specific measure, but its sense must be *invariable*;

F) M(X | S) shall be invariant under linear transformations (changes of scale) of either X or S, i.e., for real constants a and b, a $\neq$ 0:

$$\mathbf{M(a\,X + b \mid S) = M\,(X \mid S)} \quad \mathbf{(24)}$$

$$\mathbf{M(X \mid S) = M\,(X \mid a\,S + b)} \quad \mathbf{(25)}$$

G) M(X | S) shall be <u>practical</u> in the sense of *easy interpretation* (intuitive) and *easy to quantify/estimate*.

H) M(X | S) shall be <u>additive</u> in the following sense: if the information S can be decomposed into a sum of *independent* factors $S_1 + S_2 + \ldots + S_n$, so that the knowledge of *all* these factors provides *maximum learning*, then the summation of the individual learning measures shall be equal 100%, i.e.:

$$\mathbf{M(X \mid S_1) + M(X \mid S_2) + \ldots + M(X \mid S_n) = 1} \quad \mathbf{(26)}$$

<u>**Theorem 3**</u> (<u>Learning measure η^2</u>): The proposed learning measure η^2 obeys the entire axiom list.

Proof: Note that η^2 uses the variance as measure of uncertainty. So, *learning* means that is expected the reduction of uncertainty measured by the variance. In this way, *maximum learning* means full reduction of variance (posterior variance equal to zero) and *zero learning* means zero reduction of variance. The Proposition 1 proves most of the listed axioms, in some cases in stronger way. Axioms A, B and C are proved by Proposition 1 (a), (b) and (c). Axiom D is proved by Proposition 1 (d) and by the fact that eq. (13) implies in E[Var[X | S]] = Var[X], so proving eq. (21) by the definition of "zero learning" for η^2. Axiom E is proved by Proposition 1 (e) and by the definition of "maximum learning", which implies that E[Var[X | S]] = 0, which implies that η^2(X | S) = 1. Axiom F is proved by Proposition 1 (f) and (g), but with η^2 holding for more general cases: g(S) can be *any* an invertible (or 1-1) function (not only linear functions). Axiom G is more subjective, but the measure η^2 holds widely in the sense that it has an intuitive interpretation of reduction of uncertainty, in percentage terms, and can be showed (Dias, 2005) that η^2 can be estimate with non-parametric[27] or with parametric methods (when applicable), including popular parametric statistical methods like regressions (linear or nonlinear)[28] and ANOVA. Axiom H is proved with Theorem 2, in a stronger (more general) version, because it is valid for any real function (not only linear functions). □

Theorem 3 shows the η^2 strength as learning measure. Surprising this learning measure has not been used before in VOI literature, which uses in most cases measures based in likelihood function (that doesn't obey most axioms). Now, we will see an application of this learning measure in order to build the Bernoulli revelation processes, which is useful in applications like oil/gas exploration and R&D.

4 – Bernoulli Revelation Processes

In order to evaluate the effect of a *binary signal*[29] S over another binary r.v. X, we must study the dependence relation between two Bernoulli distributions, i.e., the *joint distribution* between X and S, the bivariate Bernoulli distribution.

Bivariate Bernoulli distribution is defined with three parameters: the two parameters that define the marginal distributions (success probabilities p and q) and a third parameter that establishes the dependence between the Bernoulli marginal distributions. The later can be, e.g., the joint success

---

[27] Note that η^2 is non-parametric because it is related only with *variances*, not assume a specific type of distribution.
[28] If the correct regression, in the sense of minimization of the *mean square error*, is linear (e.g., X and S are normal r.v.), η^2 is equal the squared-correlation coefficient ρ^2; if a non-linear regression is correct, then η^2 is R^2, the regression fitting factor.
[29] Examples: a neighboring well can reveal success or failure (see example Figure 1); a 3D seismic record can indicate success (indicative of favorable oil reservoir structure) or failure; a R&D project phase can reveal success or failure, etc.

probability p_{11} = Pr(X = 1 **and** S = 1). However, later we'll replace p_{11} by η^2. Table 1 presents the bivariate Bernoulli distribution as well the univariate marginal distributions.

Table 1 – Bivariate Bernoulli Distribution and Marginal Distributions

		Signal S (eg.: seismic)		Marginal distribution of X (CF)
		S = 1	S = 0	
Variable X (eg.: chance factor)	X = 1	$\mathbf{p_{11}}$	$\mathbf{p_{10}}$	p
	X = 0	$\mathbf{p_{01}}$	$\mathbf{p_{00}}$	1 – p
Marginal Distribution of S		q	1 – q	

Without loss of generality, we'll set X = CF, the exploratory chance factor, but theory can be used for other applications (like R&D). By notational convenience for Bernoulli revelation processes, instead the success probability notation p will be used the notation CF_0, i.e., $\mathbf{p = CF_0}$.

The <u>revelation distribution</u> has <u>two scenarios</u> in this case, conveniently denoted by CF^+ and CF^-:

$$\mathbf{CF^+ = Pr[CF = 1 \mid S = 1] = E[CF \mid S = 1]} \qquad \textbf{(27)}$$

$$\mathbf{CF^- = Pr[CF = 1 \mid S = 0] = E[CF \mid S = 0]} \qquad \textbf{(28)}$$

So, CF_0 *evolves* to CF^+ or CF^-, depending on the signal S. These revelation distribution scenarios have occurrence probability of q and (1 – q), respectively. Elementary probability theory permits write the following equations for marginal and joint distributions (to be used in proofs):

$$\mathbf{CF_0 = p = p_{11} + p_{10}} \qquad \textbf{(29)}$$

$$\mathbf{q = p_{11} + p_{01}} \qquad \textbf{(30)}$$

$$\mathbf{p_{11} + p_{10} + p_{01} + p_{00} = 1} \qquad \textbf{(31)}$$

Traditional probability theory defines conditional probability as P(A | B) = P(A ∩ B) / P(B). Hence:

$$\mathbf{CF^+ = \frac{p_{11}}{q}} \qquad \textbf{(32)}$$

$$\mathbf{CF^- = \frac{p_{10}}{1 - q} = \frac{CF_0 - p_{11}}{1 - q}} \qquad \textbf{(33)}$$

Where was used eq. (29) to set eq. (33). By combining eqs. (29) and (30) in eq. (31), we get the probability p_{00} in terms of more basic variables CF_0, q and p_{11}:

$$\mathbf{p_{00} = 1 + p_{11} - CF_0 - q} \qquad \textbf{(34)}$$

It is easy to prove (or see Kocherlakota & Kocherlakota, 1992, p.57) that CF and S are independent Bernoulli r.v. if and only if its joint success probability $Pr(CF \cap S) = p_{11}$ is equal to the product of the marginal success probability:

$$\textbf{CF and S independent} \Leftrightarrow p_{11} = CF_0\, q \qquad (35)$$

The correlation coefficient ρ(CF, S) is obtained easily by calculating the covariance, normalizing, and using the previous equations (or see Kocherlakota & Kocherlakota, 1992, p.57):

$$\rho(CF, S) = \frac{p_{11} - CF_0\, q}{\sqrt{CF_0\,(1 - CF_0)\, q\,(1 - q)}} \qquad (36)$$

The multivariate distribution literature shows that are necessary *limits of consistence* for these distributions, i.e., given the marginal distributions, it is not possible *any* dependence intensity because, e.g., eq. 31 must hold, etc. In particular, *full revelation* is not possible for *any* marginal distributions of CF and S. Any measure of dependence of distributions with given marginals must obey these limits, which are called Fréchet-Hoeffding bounds (see, e.g., Nelsen, 1999). For bivariate Bernoulli distributions, the Fréchet-Hoeffding limits for the double success probability and for the correlation coefficient are respectively (proof: Joe, 1997, p.210):

$$\text{Max}\{0, CF_0 + q - 1\} \le p_{11} \le \text{Min}\{CF_0, q\} \qquad (37)$$

$$\text{Max}\left\{-\sqrt{\frac{CF_0\, q}{(1 - CF_0)\,(1 - q)}}\;,\; -\sqrt{\frac{(1 - CF_0)\,(1 - q)}{CF_0\, q}}\right\} \le \rho \le$$

$$\le \sqrt{\frac{\text{Min}\{CF_0\,, q\}\,(1 - \text{Max}\{CF_0\,, q\})}{\text{Max}\{CF_0\,, q\}\,(1 - \text{Min}\{CF_0\,, q\})}} \qquad (38)$$

Now we are ready to set the Theorem 4, using our proposed learning measure η^2 to express dependence between the marginal distributions. In this theorem, we assume positive dependence that implies in $CF^- < CF_0 < CF^+$. The case of negative dependence is symmetric: $CF^+ < CF_0 < CF^-$, with changes in equations as indicated in the theorem.

Theorem 4 (learning measure η^2 and bivariate Bernoulli distribution): Let the (non-trivial) marginal distributions be CF ~ $Be(CF_0)$ and S ~ Be(q), linked in a bivariate Bernoulli distribution by the learning measure $\eta^2(CF \mid S)$ defined by eq. (7) or by eq.(8). Then:

(a) The revealed success probabilities CF^+ and CF^-, in case of *positive* dependence, are:

$$\mathbf{CF^+ = CF_0 + \sqrt{\frac{1-q}{q}}\ \sqrt{CF_0\ (1-CF_0)}\ \sqrt{\eta^2(CF\ |\ S)}} \quad \textbf{(39)}$$

$$\mathbf{CF^- = CF_0 - \sqrt{\frac{q}{1-q}}\ \sqrt{CF_0\ (1-CF_0)}\ \sqrt{\eta^2(CF\ |\ S)}} \quad \textbf{(40)}$$

In case of *negative dependence*, eqs. (39) and (40) hold, but inverting the signal after CF_0.

(b) The learning measure η^2 in this case is equal to the square of correlation coefficient ρ:

$$\mathbf{\eta^2(CF\ |\ S)\ =\ \rho^2(CF, S)\ =\ \frac{(p_{11} - CF_0\ q)^2}{CF_0\ (1-CF_0)\ q\ (1-q)}} \quad \textbf{(41)}$$

(c) The learning measure η^2 in this case is symmetric:

$$\mathbf{X\ e\ S \sim Bernoulli\ \Rightarrow\ \eta^2(CF\ |\ S)\ =\ \eta^2(S\ |\ CF)} \quad \textbf{(42)}$$

(d) The learning measure η^2 in this case is equal to zero if *and only if* CF and S are independent:

$$\mathbf{\eta^2(CF\ |\ S)\ =\ 0\ \Leftrightarrow\ CF\ and\ S\ independent} \quad \textbf{(43)}$$

(e) To assure the bivariate Bernoulli distribution existence, the Fréchet-Hoeffding bounds in terms of η^2, being allowed learning with positive or negative dependence, are given by:

$$\mathbf{0 \le \eta^2(CF\ |\ S) \le Max\left\{\begin{array}{l} Max\left\{\dfrac{CF_0\ q}{(1-CF_0)\ (1-q)}\ ,\ \dfrac{(1-CF_0)\ (1-q)}{CF_0\ q}\right\}, \\ ,\ \dfrac{Min\{CF_0\ ,\ q\}\ (1-Max\{CF_0\ ,\ q\})}{Max\{CF_0\ ,\ q\}\ (1-Min\{CF_0\ ,\ q\})} \end{array}\right\}} \quad \textbf{(44)}$$

Proof: Dias (2005). Provided under request, it is based in algebraic manipulation of previous equations presented in this chapter and equations from Theorem 1, in most cases.

The equations from theorem 4 solve important practical problems such as the example presented in Figure 1, a portfolio of dependent exploratory assets, and option games problems such as the case presented in Dias & Teixeira (2004): an oil exploration game with information spillover with cooperative Nash bargain with disagreement point being an equilibrium from a war of attrition game. Dias & Teixeira (2004) used different values for the case of bargain (private information is revealed) and for the war of attrition case (only public information can be revealed), with different values of η^2 in each game (bargain with higher η^2). Another example was a real business valuation to enter in a long-term oil exploration in a new international basin, considering the information revelation effect cause by the activity of wildcat drilling. The information structure used in these problems is simply:

$$\mathcal{I} = \{\mathbf{S}, \eta^2(\mathbf{CF} \mid \mathbf{S})\} \qquad \textbf{(45)}$$

With this information structure plus the prior distribution (here defined by CF_0), we solve all related problems. However, there is a way to simplify even more this class of problems, which is useful mainly for *sequence of signals* generating a Bernoulli revelation process. This simplification is the assumption that the chance factor CF and the signal S are exchangeable r.v. (see, e.g., O'Hagan, 1994, p. 112-118, 156, 290). This class of *symmetrically dependent* r.v. has been assumed intuitively in many oil exploration models of dependent prospects (e.g., Wang et al., 2000). In the case of two Bernoulli r.v., exchangeability implies that the marginal success probabilities are the same, and many equations simplify, as indicated by Proposition 2 below.

Proposition 2: Let CF and S be non-trivial Bernoulli exchangeable r.v., with the bivariate Bernoulli distribution defined by the success probabilities CF_0 and q and by the learning measure η^2. Then:

(a) The marginal distributions are equal. The converse also holds, i.e.,

$$\textbf{CF and S exchangeable} \Leftrightarrow \mathbf{CF_0 = q} \qquad \textbf{(46)}$$

(b) The full revelation ($\eta^2 = 1$) is always possible for any non-trivial values of CF_0 and q. That is, the Fréchet-Hoeffding bounds are not restrictions anymore:

$$\textbf{Fréchet-Hoeffding bounds:} \quad \mathbf{0 \le \eta^2 \le 1} \qquad \textbf{(47)}$$

(c) The revealed success probabilities CF^+ and CF^-, in case of *positive* dependence, are:

$$\mathbf{CF^+ = CF_0 + (1 - CF_0)\,\eta} \qquad \textbf{(48)}$$

$$\mathbf{CF^- = CF_0 - CF_0\,\eta} \qquad \textbf{(49)}$$

In case of negative dependence, eqs. (48) and (49) hold, but inverting the signal after CF_0.

Proof: Dias (2005). Provided under request.

Lemma 3: Let CF and S be non-trivial Bernoulli r.v., with the bivariate Bernoulli distribution defined by the success probabilities CF_0 and q and by the learning measure η^2. The necessary condition for full revelation of CF is that CF and S be exchangeable r.v.:

$$\mathbf{\eta^2(CF \mid S) = 1} \Rightarrow \textbf{CF and S exchangeable r.v.} \qquad \textbf{(50)}$$

Proof: Dias (2005). Provided under request.

Note that in eqs. (48) and (49) are used the positive square root of the proposed learning measure, i.e., η. In this way, the revealed success probabilities CF^+ and $CF^-$are *linear functions* of η. By using these equations, it is easy to see that the difference of revealed success probabilities is simply η:

$$\mathbf{CF^+ - CF^- = \eta} \qquad \textbf{(51)}$$

Figure 6 below shows the revealed success probabilities CF^+ and CF^-, in case of *positive* dependence for many values of η. Note that the variation is linear and the chance factors difference is η.

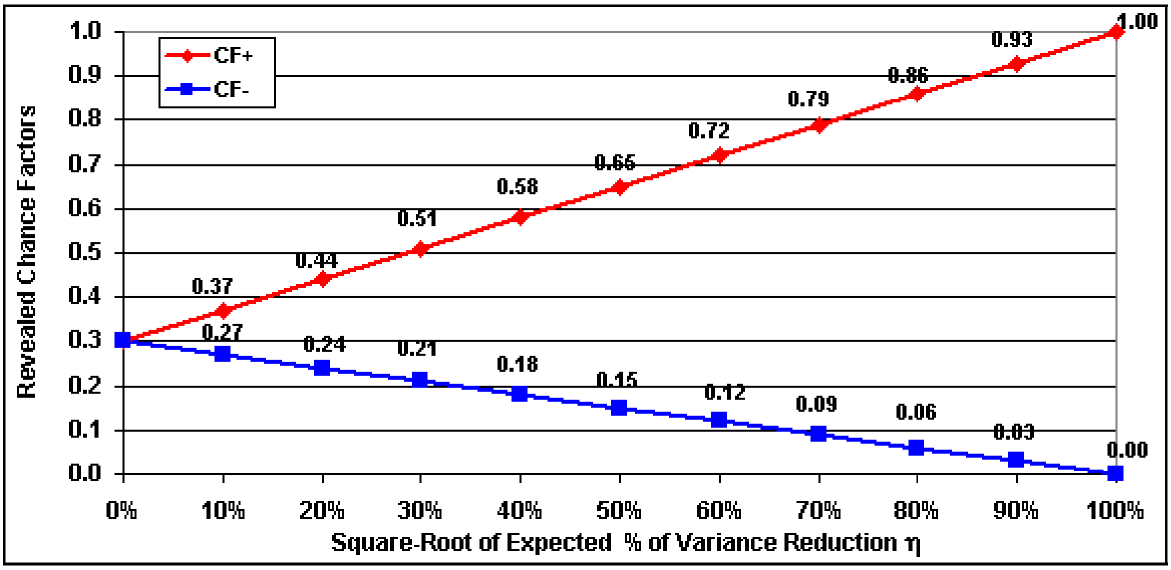

Figure 6 – Revealed Chance Factors x Square-Root of Expected % of Variance Reduction

With these equations and concepts about bivariate Bernoulli distributions, we are ready to study Bernoulli revelation processes, starting with some definitions.

Definition. **Exploratory discovery process**: is a sequence of learning options exercises that results in a discovery. In general, these learning options have different learning costs, different time to learn, and different revelation powers (that can measured by η^2). In oil exploration this sequence can be magnetic search, gravimetric search, seismic record search, drilling of one or more wildcat wells. In R&D, this sequence can be the R&D phases.

Definition. **Exploratory revelation process**: is the probabilistic effect over the interest variable caused by the exploratory discovery process. This interest variable (X) can be a chance factor (success probability), oilfield in-place volume, fluids quality in an oilfield, etc. In R&D can be a chance factor, MTBF, operational efficiency of a new machine, etc.

Definition. **Bernoulli revelation process**: is a sequence of revelation distributions generated by a sequence of bivariate Bernoulli distributions that represents the interaction between a sequence of signals S with the chance factor of interest CF.

In particular, we focus the *exchangeable* revelation Bernoulli processes. In this way, given the prior distribution (i.e., CF_0) and a sequence of η_k, k = 1, 2, …, we can construct all the revelation process because the success probabilities for the signals S_k are automatically defined with the exchangeable assumption. For example, in case of positive revelation with the first signal S_1, the success probability for the second signal also rises to CF^+ (to be exchangeable, see eq. 46). This simplifies the construction of revelation process because the sequence of information structures comprises only a sequence of η_k and the exchangeability assumption. We assume that the signal revelation power η_k is independent of the CF scenario (changes only with k), only the signal success probability changes with the scenario. Some figures below will make clear these points.

Revelation Bernoulli processes can be *non-recombining* or *recombining*. Figures 7 and 8 show, respectively, non-recombining and recombining exchangeable revelation Bernoulli processes with two signals. Inside the rectangles are the current (revealed) chance factor success probabilities. Each vertical set of rectangles is the set of scenarios from the revelation distribution after each signal.

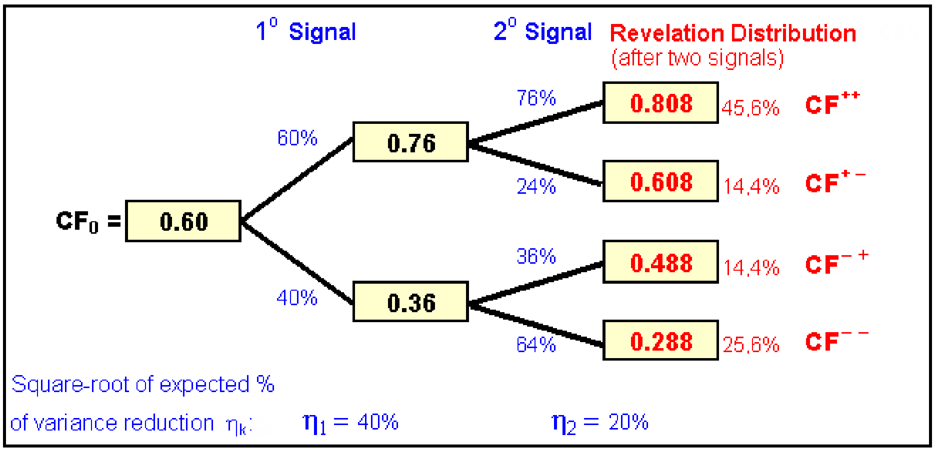

Figure 7 – Non-Recombining Exchangeable Revelation Bernoulli Process

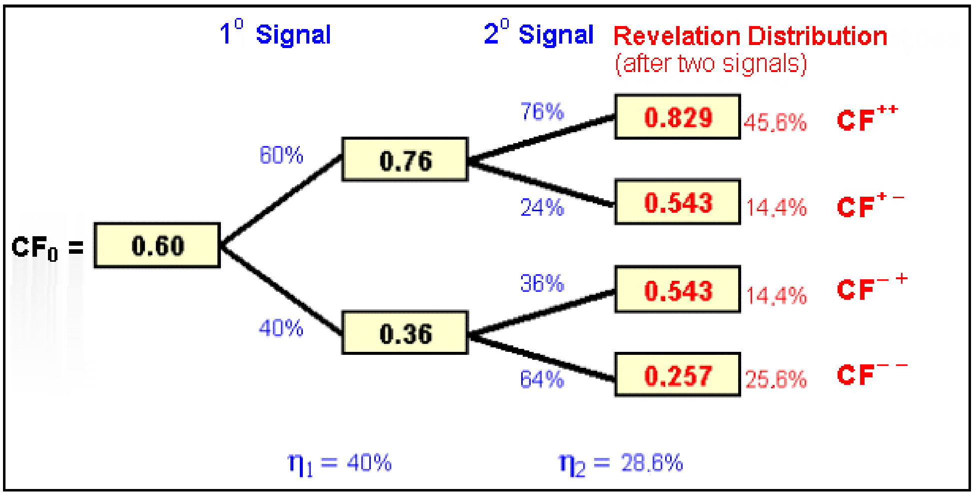

Figure 8 – Recombining Exchangeable Revelation Bernoulli Process

Figures 7 and 8 show also the probability of occurrence for each scenario (in red, outside the rectangles), after the second signal. In order to recombine, given CF_0, we must use a *specific* and *decreasing* sequence of η_k, as showed in Figure 8 (note that $\eta_2 = 28.6\% < \eta_1$). We'll return to this point. The recombining revelation process resembles the *binomial lattice used in discrete-time option pricing*. But here we have discrete-event process. Figure 9 shows this recombining process after 10 signals, drawn in a format that resembles the binomial.

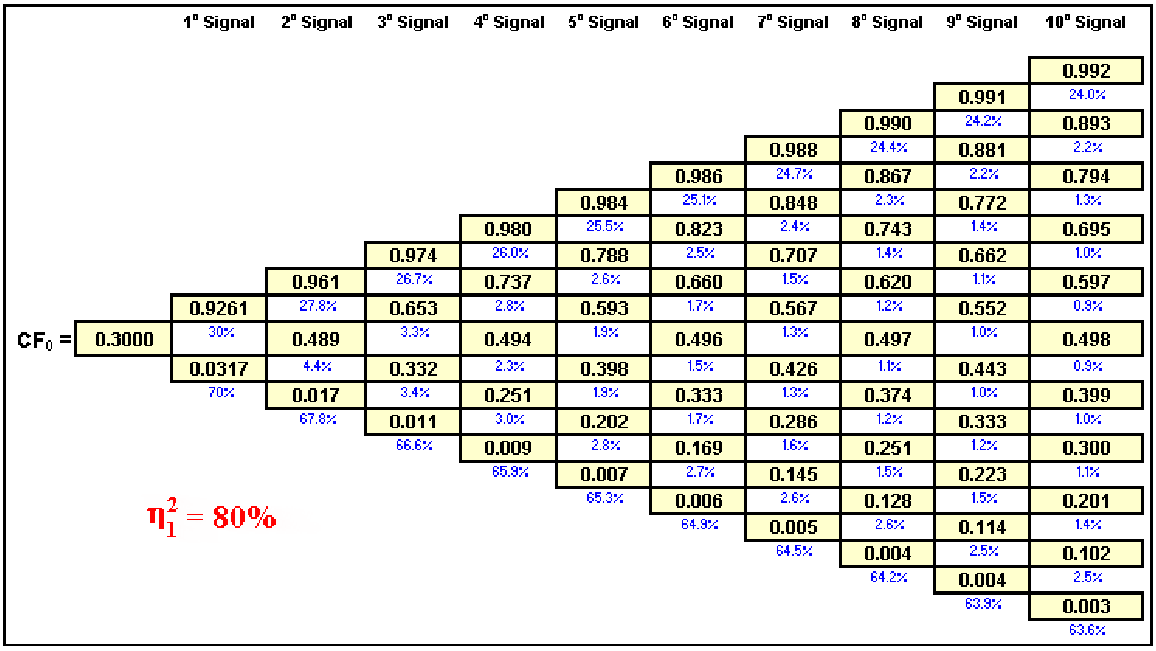

Figure 9 - Recombining Exchangeable Revelation Bernoulli Process with 10 Signals

The revelation process in Figure 9 was built with only two inputs (plus the assumptions of recombining and exchangeable process): the prior chance factor CF_0 (= 0.3) and the first learning measure η_1^2 (= 80%, that implies in η_1 = 89.4%). It is easy to show that any subsequent η_k in this process is given by the simple equation, knowing only the first one (η_1) in the sequence. This equation also shows that the η_k are decreasing in k in this exchangeable recombining process.

$$\eta_k = \frac{\eta_1}{1 + (k - 1)\,\eta_1} \tag{52}$$

We'll see that exchangeable recombining Bernoulli revelation process, despite of the decreasing (average) posterior variance, never converges to the full revelation case because the specific decreasing schema for η_k (eq.52)[30]. Figure 9 used a very high initial value for η_k in order to highlight another point: Theorem 1(a) says that in the full revelation limit the revelation distribution converges to prior distribution. In this case, the prior distribution is a Bernoulli one, which has only two scenarios (1 with probability CF_0 and 0 with probability 1 – CF_0). By looking the "diffusion" revelation process from the last three figures, the number of revelation distribution scenarios rises without limit (k + 1 scenarios for the recombining case). Is it a paradox? No! Even without converging to full revelation, the high initial value for η_k in Figure 9 makes the revelation distribution after 10 signals close of the full revelation case and we can see in Figure 10 the histogram for the revelation distribution, which will solve this paradox in an intuitive way.

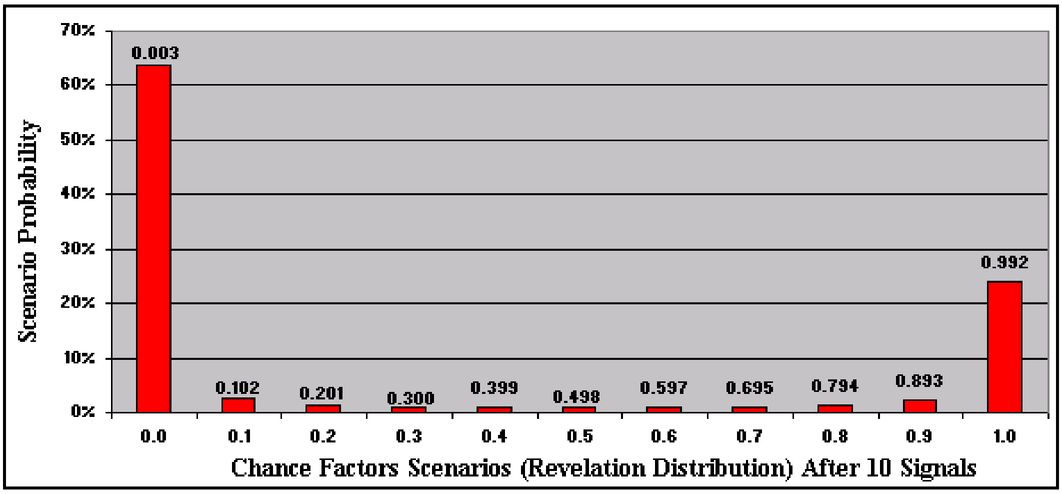

Figure 10 – Histogram for Revelation Distribution After 10 Signals

[30] But increasing (with k) or constant η_k sequences <u>always</u> converges to full revelation when $k \to \infty$. It is easy to see this: after each signal S_k, the average CF variance is reduced by the factor $(1 - \eta_k^2)$, and after infinite signals, the infinite product of these factors reduces the variance to zero if η_k is constant (e.g., $0.9^{100} \cong 0$) or if η_k increases with k.

Note in Figure 10 that the middle scenario values have small probabilities and the extreme scenarios (near 0 or 1) has most of probability mass. This is not coincidence: as the number of signals increase, the probability mass migrates from middle to the extreme scenarios. In case of convergent revelation process, when $k \to \infty$ all the probability mass in the middle scenarios go to zero and the scenarios 0 and 1 receive all the probability mass. Note in Figure 10 that the scenario near the value 0 (0.003) has probability approaching to $1 - CF_0 = 70\%$, whereas the scenario near 1 (0.992) has probability approaching to $CF_0 = 30\%$.

We said that recombining exchangeable revelation Bernoulli processes don't converges to full revelation. The revelation process presented at Figure 8 ($CF_0 = 0.6$ and $\eta_1 = 0.4$), after a large number of signals, has its mean posterior variance as % of prior variance showed at Figure 11.

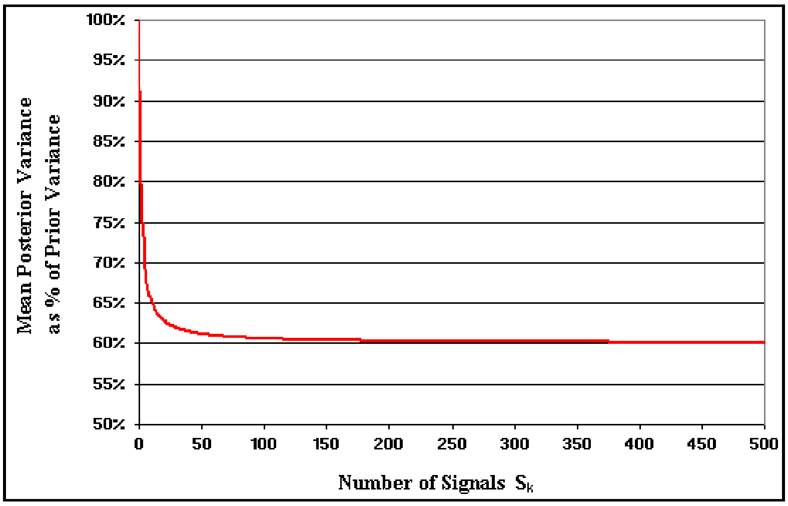

Figure 11 – Partial Convergence of a Recombining Revelation Process

It is not by chance that this mean posterior variance converges to 0.6 ($= 1 - \eta_1$) of the initial (prior) variance. Some algebraic manipulation shows the recombining exchangeable revelation Bernoulli processes converges only to partial revelation in which the mean posterior variance as % of prior variance is exactly $1 - \eta_1$. Equation 53 formalizes this partial convergence.

$$\lim_{n \to \infty} \prod_{k=1}^{n} \left(1 - \frac{\eta_1^2}{[1 + (k-1)\,\eta_1]^2} \right) = 1 - \eta_1 \qquad (53)$$

Note that the term inside the parenthesis is the factor $(1 - \eta_k^2)$, which reduces the average CF variance, combined with eq.(52). Although recombining exchangeable revelation Bernoulli processes never converges to full revelation (assuming non-trivial cases with $\eta_1 \neq 1$), the study of these processes is important because: (a) it is very simple; (b) the number of scenarios doesn't "explode" as in the non-recombining processes (= 2^k); and (c) can be more realistic in some practical cases. An example for the latter point is a petroleum basin: even after a very large number of wildcats drilled in a basin (e.g., shallow waters in Gulf of Mexico), the cumulative knowledge is not sufficient to say that one area has or not petroleum, although the chance factors are or very high or very low.

The recursive equations to calculate the revealed CF scenarios in a (not necessarily exchangeable) Bernoulli revelation process is presented below, with the Figure 12 below, for the recombining cases, making clear the used notation.

$$\mathbf{CF_k^{+-} = CF_{k-1}^{+} - \sqrt{\frac{q_k^+}{1-q_k^+}}\sqrt{CF_{k-1}^{+}\,(1-CF_{k-1}^{+})}\,\sqrt{\eta_k^2}} \qquad \textbf{(54)}$$

$$\mathbf{CF_k^{-+} = CF_{k-1}^{-} + \sqrt{\frac{1-q_k^-}{q_k^-}}\sqrt{CF_{k-1}^{-}\,(1-CF_{k-1}^{-})}\,\sqrt{\eta_k^2}} \qquad \textbf{(55)}$$

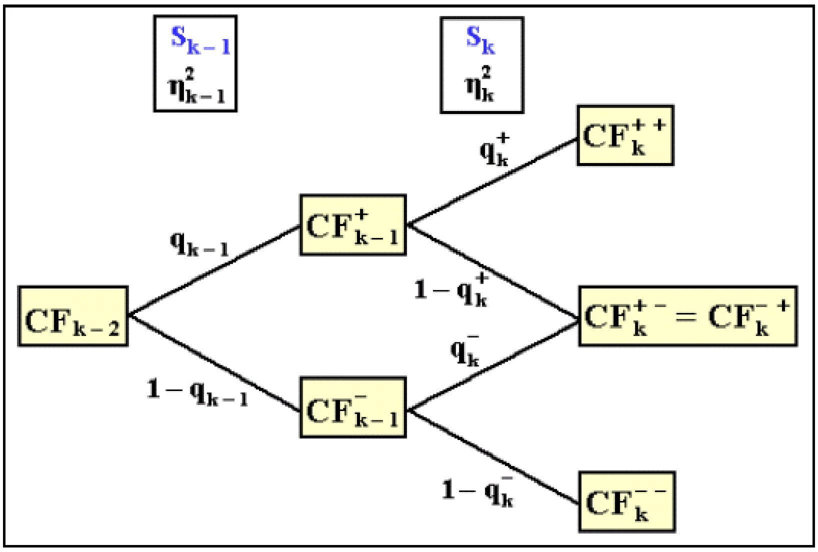

Figure 12 – General Schema for Recombining CF Scenarios

For the recombining cases, eqs. 54 and 55 provide the same value. For non-exchangeable processes, we have two variables to adjust the recombination: q_k and η_k. With the Figure 12 is easy to see that the revelation distribution scenarios probabilities $p_{i,k}$ are obtained recursively with the following equation, being the scenario i – 1 the upper one and i the bottom one. Assume the value zero for the probability p when there is no scenario (in the extreme scenarios case).

$$\mathbf{p_{i,k} = p_{i-1,k-1}\,(1 - CF_{i-1,k-1}) + p_{i,k-1}\,CF_{i,k-1}} \qquad \textbf{(56)}$$

An important point in a revelation processes is that, at least in petroleum applications, the event-indexed scale (signals) is *very* different of time-indexed scale. By observing the *event-diffusion process*, such as the one displayed in Figure 9, in case of positive revelation, increases the chance of a rapid new exercise of the learning option (signal), because many prospects in the basin can become "deep-in-the-money" with this higher CF, a cascade effect. In case of negative revelation, the opposite occur: firms will postpone the wildcat drilling (exercise of learning options generating signals) in that basin. So, time run faster in the upper branches of Figure 9 than in the bottom branches of this event-driven diffusion process. This is an additional argument to classify technical uncertainty modeling with traditional Brownian models as "too rough" or inadequate.

5 – Conclusion

In this paper was presented a theory for technical uncertainty modeling with focus on real options and option games applications. This theory is based in the concept of conditional expectations. The distribution of conditional expectations is here named revelation distribution in order to highlight the property of always to reduce the process variance in average terms (process towards the truth about a variable of interest). The revelation distribution properties are useful in a Monte Carlo framework to solve complex real options problems with both technical and market uncertainties. The paper link the revelation distribution concept with a proposed learning measure η^2, which has surprising adequate mathematical properties and an intuitive interpretation as expected variance reduction with the learning process. A list of axioms for learning measures is proposed and showed that some known measures used in VOI problems are not so adequate to model the learning power of a learning alternative. This learning measure is used to build the simplest revelation process, the Bernoulli revelation process, which is useful for real asset portfolio valuation in oil exploration problems, in option games applications and in R&D modeling. Some specific Bernoulli revelation processes were studied, such as the recombining process that has an interesting simplicity for applications. Many other revelation processes can be studied for specific applications and in order to enrich this theory on technical uncertainty and learning options exercises. This paper can be considered only a starting point to this goal.

APPENDIXES

A) Definition of Uniformly Integrable Sequence

The r.v. sequence $R_{X,1}$, $R_{X,2}$, … $R_{X,n}$ is called uniformly integrable if, for all $\varepsilon > 0$, exists a $M > 0$ so that for all n = 1, 2, … :

$$\int_{\{|R_{X,n}| > M\}} |R_{X,n}| \; dP < \varepsilon \quad \textbf{(A1)}$$

Uniform integration is a necessary condition for a sequence of integrable r.v. $\{R_{X,n}\}$ to converge in $\boldsymbol{L}^1$, i.e., to converge in mean to a integrable r.v., which in our case is $\lim_{n \to \infty} E[|R_{X,n} - X_\infty|] = 0$. The integral in the eq. (A1) is of Lebesgue-Stieltjes type, i.e., **P** is a probability measure or simply the cumulative probability distribution function $\mathbf{P}(R_{X,n})$.

B) Proposition 1 Proof

(a) This is a trivial property considering the assumption Var[X] > 0, but finite, and the assumption of finite Var[S]. □

(b) The measure η^2 is *in general* asymmetric if at least one example exhibits asymmetry. This example was presented in the text (see Figure 3). □

(c) The eqs. (7) and (8) show that: by (8) η^2 is a quotient of variances and so it cannot be lower than zero due to variance definition. The eq. (7) shows that the maximum value is equal to 1, because η^2 is maximized by minimizing E[Var[X | S]], which has a minimum value at Var[X | S] = 0 for all $s \in S$, i.e., for $E[Var[X \mid S]] = 0 \Rightarrow \eta^2(X \mid S) = 1$. □

(d) If X and S are independent r.v., then E[X | S] = E[X] and E[S | X] = E[S] (elementary independence property, see, e.g., Williams, 1991, p.88). In addition, the variance of any r.v. Y is *defined* by $Var[Y] = E[(Y - E[Y])^2]$. Hence, the revelation distribution variances from $R_X(S)$ and $R_S(X)$ are both equal to zero because:

$$Var[R_X(S)] = E[(E[X \mid S] - E[E[X \mid S]])^2] = E[(E[X \mid S] - E[X])^2] = 0$$

$$Var[R_S(X)] = E[(E[S \mid X] - E[E[S \mid X]])^2] = E[(E[S \mid X] - E[S])^2] = 0$$

Where were used the Theorem 1(b) and the mentioned elementary independence property. If the revelation distribution variances are equal to zero, the measures $\eta^2(X \mid S)$ and $\eta^2(S \mid X)$ are also equal to zero, because these learning measures are normalized revelation distribution variances. This proves eq. (12) and the return (⇐) in eq. (13). In order to prove the the direction ⇒ in eq. (13), note in eq. (8) that $\eta^2(X \mid S) = 0 \Rightarrow Var[R_X(S)] = \eta^2(X \mid S)\ Var[X] = 0$ if $Var[X] > 0$. □

(e) If $\eta^2(X|S) = 1$, eq.(7) $\Rightarrow E[Var[X|S]] = 0 \Rightarrow E[(X - E[X|S])^2] = 0$ and so, with probability 1 ⇒ $X = E[X \mid S] \Rightarrow$ X is measurable by sigma-algebra of S, and hence we can write X = g(S).

In Hall (1970, p.342), is used the axiom that, if a dependence measure is equal to 1 ⇒ X = g(S), but without the return (⇐). In Hall, the return X = g(S) ⇒ dependence measure equal to 1, is not considered necessary. In these more theoretical studies, Hall wished to study cases with infinite variance of X (without interest here). By assuming that Var[X] is finite, then holds the return for the measure $\eta^2_{\cdot}$. The proof for the return (⇐) is even simpler: if S is revealed, then by function definition, X = f(S) is *unically* determined. So, $E[Var[X \mid S]] = 0$ and hence $\eta^2(X \mid S) = 1$ by eq. (7). According Hall (1970, p.342), in case of *infinite* variance of X, when the knowledge of S *reduces the variance from infinite to a finite value*, this can be considered a *complete* state of dependence, in an analogous way of the case of finite X variance, in that S reduces the X variance from a finite value to the zero value. In applications of this paper in that Var[X] is always finite, writing the stronger property version (⇔) is more convenient. □

(f) The proof is simple. Applying eq. (7) for the variable Y = aX + b:

$$\eta^2(a\,X + b \mid S) = \frac{Var[a\,X + b] - E[Var[a\,X + b \mid S]]}{Var[a\,X + b]} \Rightarrow$$

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

Because $a \neq 0$, we can simplify (by cutting a^2) to get $\eta^2(X \mid S)$. □

(g) The equality case if the function Y = g(S) is 1-1 is because the inverse $S = g^{-1}(Y)$ exists, i.e., g^{-1} is a *function* that, by definition determines uniquely S if we know Y = g(S). So, knowing g(S) is equivalent to know S.

If g(S) was <u>not</u> 1-1, the knowledge of g(S) would be lower than the knowledge of S, e.g., if $Y = S^2$ then the value Y = 1 could be either due to S = 1 as S = – 1. Then, the intuition said that $\eta^2(X \mid g(S))$ shall be lower than $\eta^2(X \mid S)$. Formally, Rényi (1970, p.278-279) shows that $\eta^2(X \mid S)$ is the supreme of $\rho^2(X, f(S))$ for *any real function* f(S). This implies that $\eta^2(X \mid S) \geq \rho^2(X, f(S))$ for *any real function* f(S). If this function is any, this includes the function $f(S) = E[X \mid g(S)]$. So:

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

Rényi also shows that $\eta^2(X \mid Y) = \rho^2(X, E[X \mid Y])$. By making Y = g(S) we get $\eta^2(X \mid g(S)) = \rho^2(X, E[X \mid g(S)])$. By substituting in the previous inequality, we get: $\eta^2(X \mid S) \geq \eta^2(X, g(S))$ □

(h) This property was suggested by Hall (1970) as one dependence axiom in order to related with the correlation coefficient. This was proved in Dias (2005), which is provided under request.

C) Theorem 2 Proof

Recall the definition of $\eta^2(X \mid S_i) = Var[E(X \mid S_i)] / Var[X]$. So, the left side of (19) is:

$$\eta^2(X \mid S_1) + \eta^2(X \mid S_2) + \ldots + \eta^2(X \mid S_n) =$$

$$= \frac{Var[\,E(X|S_1)]}{Var[X]} + \frac{Var[\,E(X|S_2)]}{Var[X]} + \ldots + \frac{Var[\,E(X|S_n)]}{Var[X]} =$$

$$= \frac{Var[\,E\{f(S_1)+g(S_2)+\ldots+h(S_n)\ |S_1\}] + \ \ldots + Var[\,E\{f(S_1)+g(S_2)+\ldots+h(S_n)\ |S_n\}]}{Var[f(S_1)+g(S_2)+\ldots+h(S_n)]} =$$

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

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

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

$$\eta^2(X \mid S_1) + \ldots + \eta^2(X \mid S_n) \quad = \quad \frac{Var[\,f(S_1)]+Var[\,g(S_2)]+\ldots\ +Var[\,h(S_n)]}{Var[f(S_1)]+Var[g(S_2)]+\ldots+Var[h(S_n)]} = 1$$ □

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.

Arrow, K. (1992). Informational Equivalence of Signals. In Partha Dasgupta, et al. eds. *Economic Analysis of Markets and Games*. Cambridge: MIT Press, pp. 169-183

Blackwell, D. (1951). *Comparison of Experiments*. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1951, pp.93-102

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

Brzezniak, Z. & T. Zastawniak (1999). *Basic Stochastic Processes*. Springer-Verlag London Ltd., 225 p.

Childs, P.D. & S.H. Ott & T.J. Riddiough (2001). *Valuation and Information Acquisition Policy for Claims Written on Noisy Real Assets*. Financial Management, vol.30(2), Summer 2001, pp.45-75.

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.

DeGroot, M.H. & M.J. Schervish (2002). *Probability and Statistics*. Addison-Wesley, Boston, 3rd Edition, 816 p.

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. (2002). *Investment in Information in Petroleum: Real Options and Revelation*. Working Paper, Dept. of Industrial Engineering, PUC-Rio, presented at MIT seminar *Real Options in Real Life*, May 2002; and at *6th Annual International Conference on Real Options*, Cyprus, July 2002

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

Dias, M.A.G. (2005). *Opções Reais Híbridas com Aplicações em Petróleo* (*Hybrid Real Options with Petroleum Applications*). Doctoral Thesis, Dept. of Industrial Engineering, PUC-Rio, January 2005. 509 p. (in Portuguese).

Dias, M.A.G. & J.P. Teixeira (2004). *Continuous-Time Option Games Part 2: Oligopoly, War of Attrition and Bargaining under Uncertainty*. Working Paper, PUC-Rio, presented at the 8th Annual International Conference on Real Options, Montreal, June 2004.

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

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications – Vol. I*. John Wiley & Sons, Inc., 3rd ed., 509 pp.

Gallant, A.R. (1997). *An Introduction to Econometric Theory – Measure Theoretic Probability with Applications to Economics*. Princeton: Princeton University Press, 202 p.

Grenadier, S.R. (1999). *Information Revelation Through Option Exercise*. The Review of Financial Studies, Spring, vol.12, nº 1, pp.95-129

Hall, W.J. (1970). *On Characterizing Dependence in Joint Distributions*. In Bose, R.C. et al, eds., *Essays in Probability and Statistics*. University of North Carolina Press, Chapel Hill, p.339-376.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall/CRC, 399 p.

Karlin, S. & H.M. Taylor (1975). *A First Course in Stochastic Processes*. Academic Press Limited, Second Edition, 557 p.

Kocherlakota, S. & K. Kocherlakota (1992). *Bivariate Discrete Distributions*. Marcel Dekker, Inc., New York, 361 p.

Kolmogorov, A.N. (1933). *Foundations of the Theory of Probability*. American Mathematical Society, Chelsea Publishing (1956), 2nd English Ed. (original version in German, 1933), 84 p.

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.

Nelsen, R.B. (1999). *An Introduction to Copulas*. Lectures Notes in Statistics v.139, Springer Verlag New York, 216 p.

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics – Volume 2B: Bayesian Inference*. Arnold Publishers and Oxford University Press, 332 p.

Radner, R. & J.C. Stiglitz (1984). *A Nonconcavity in the Value of Information*. In M. Boyer & R. Kihlstrom, Eds., *Bayesian Models in Economic Theory*, Elsevier Science Pub., 1984, p.33-52

Rényi, A. (1959). *On Measures of Dependence*. Acta Mathematica Academiae Scientiarum Hungaricae, vol.10, p. 441-451

Rényi, A. (1970). *Probability Theory*. Amsterdam: North-Holland Pub. Co., 670 p.

Ross, Sheldon M. (1996). *Stochastic Processes*. John Wiley & Sons, Inc., 2nd Edition, 1996, 510 p.

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.

Wang, B. & Kokolis, G.P. & Rapp, W.J. & Litvak, B.L. (2000). *Dependent Risk Calculations in Multiple-Prospect Exploration Evaluations*. SPE paper 63198, presented at the 2000 SPE Annual Technical Conference and Exhibition held in Dallas, Texas, 1–4 October, 14 p.

Williams, D. (1991). *Probability with Martingales*. Cambridge University Press, 1991, 251 p.

Wilson, R. (1975). *Informational Economies of Scale*. Bell Journal of Economics, vol.6, 1975, pp.184-195