A Study of Business Decisions under Uncertainty: The Probability of the Improbable

With Examples from the Oil and Gas Exploration Industry

Andreas Stark

DISSERTATION.COM

Boca Raton

A Study of Business Decisions under Uncertainty: The Probability of the Improbable
- With Examples from the Oil and Gas Exploration Industry

Dissertation.com
Boca Raton, Florida
USA • 2010

ISBN-10: 1-59942-349-9
ISBN-13: 978-1-59942-349-4

This dissertation is dedicated to my wonderful and amazing wife Regina.

ACKNOWLEDGEMENTS

First of all I would like to thank my wife Regina who has been my inspiration throughout our life together. Without her encouragement I would never have finished this second dissertation and I would not have written this book. She has been my emotional support and my best friend, reviewer and critic through all my efforts in trying to create this book. I thank her for her unwavering support and for putting up with me over all these years.

I would also like to thank my dissertation advisor at Rushmore University, Professor Donald Mitchell, for all his enthusiasm, direction and helpful advice and comments for improvement. I trust that his suggestions for additions and changes have made this a better book that will now have appeal to a much broader business audience. I would also like to thank my Rushmore University editor Ms. Laurel Barley for her efforts and dedication to help me write a better book.

I have used commercially available software to display and compute the examples. The packages used are:

- Excel® for all simple examples
- Crystal Ball® mainly for Monte Carlo simulations
- Mathematica® to develop and write the fractal and neural network procedures and more complicated mathematical and statistical examples, as well as most of the figures.

All data sets and examples shown have been fully worked and computed from scratch.

Mathematica® is one system that is integrated to perform higher mathematical computations, simulations, visualizations and development. It is easier to use for demonstrations and testing theories than programming languages such as C++. It is a standard development and testing tool for many engineering, financial and science applications.

The large list of references given on the last few pages all had some part in the final development of this book and should be used for more in-depth information.

The responsibility for any errors in this book resides solely with the author. The reader is encouraged to report any errors of fact or of typographical nature to**:** **info@clasinaterra.com**

CONTENTS

LIST OF FIGURES

LIST OF TABLES

UNCERTAINTY

"The more precise the knowledge of a particle's position the less precise the knowledge of that particle's momentum." - Werner Heisenberg's Principle of uncertainty

"That God chooses to play with dice I cannot believe" - Albert Einstein

"All events, even those which on account of their insignificance do not seem to follow the great laws of nature, are a result of it just as necessarily as the revolutions of the sun… Present events are connected with preceding events by a bond based upon the evident principle that something cannot occur without a source which produces it." – Marquis de Laplace Pierre-Simon

Introduction

To lay a foundation for the understanding of the concepts developed in this dissertation, I will define uncertainty by presenting an overview of probability and statistics as used in our day-to-day activities. I will explore the basic principles of probability, univariate and multivariate statistics, data clustering and mapping, as well as time sequence and spectral analysis

I will use examples from the oil and gas exploration industry not only because that is what I know best, but also because the risks taken in this industry are normally quite large and are ideal for showing the application of the various techniques for minimizing risk.

I will treat basic risk analysis, spatial and time variations of risk, geotechnical risk analysis, risk aversion and how it is affected by personal biases, how to use portfolios to hedge risk, as well as the application of real options. Subsequently, I will treat fractal analysis and its application to economics and risk analysis and compute some examples showing the change in the Value at Risk under Fractal Brownian Motions. Finally, I will discuss how to combine some of these risks and risk factors into a single neural network application to simultaneously predict the best forecast possible given a certain knowledge base and multiple data sets.

The chapters will discuss:

- Basic probability techniques and uncertainty principles
- Analysis and diversification for exploration projects
- The value and risk of information in the decision process
- Simulation techniques and modeling of uncertainty
- Project valuation and project risk return
- Modeling risk propensity or preference analysis of exploration projects
- Application of fractals to risk analysis
- Simultaneous prediction of strategic risk and decision attributes using multivariate statistics and neural networks

I will use commercially available software to demonstrate with examples. The packages used will be:

- Excel® for all simple examples
- Crystal Ball® mainly for Monte Carlo simulations
- Mathematica® to develop and write the fractal and neural network procedures and more complicated mathematical and statistical examples, as well as most of the figures.

All data sets and examples shown have been fully worked and computed from scratch.

Mathematica® is one system that is integrated to perform higher mathematical computations, simulations, visualizations and development. It is easier to use for demonstrations and testing theories than programming languages such as C++. It is a standard development and testing tool for many engineering, financial and science applications.

Keywords:
Business, Decision Theory, Economics, Exploration, Fractal Analysis, Linear Programming, Mathematical Finance, Neural Networks, Portfolio Analysis, Operations Research, Project Valuation, Real Options, Risk Analysis, Applied Statistics, Value at Risk, Uncertainty

Chapter 1

A review of elementary statistics

"Chance is the pseudonym God uses when He'd rather not sign His own name." **- Anatole France (1844-1924)**

"Statistics: The only science that enables different experts using the same figures to draw different conclusions." **— Evan Esar (1899-1995), American Humorist.**

"The probability of the bread falling buttered side down is directly proportional to the price of the carpet." **Anonymous**

"Probability is relative in part to our ignorance and in part to our knowledge." **– Marquis de Laplace**

The Scope of Statistics

The original name statistics meant information useful to the state. Today it is used more to mean quantitative data, which tend to fluctuate in a more or less unpredictable manner, as talked about in the newspapers, on the TV news, in business and in economics. In the broader sense, statistics is the science and also an art concerned with the collection, presentation and analysis of quantitative data so that we may form an intelligent judgment upon these data. Statistics is of great help in the social and biological sciences but for quite awhile there did not appear to be a need to use statistics in the more exact sciences. However, statistics have now become an integral part of both the physical sciences and engineering.

Frequency Distributions

From the standpoint of a mathematical analysis of statistics, the most important form of data arrangement is the frequency distribution. Data needs to be presented in a condensed and organized manner to convey a clear understanding. In order to show the frequency distribution we divide the data into classes of appropriate sizes to show the frequency or rate of occurrence of the variates in each class. These are also called absolute frequencies.

A ***variate*** is a variable in the domain of real numbers, which can be denoted by letters such as x and y. A ***variable*** is a quantity that may take on any value from a given domain; the set of values. A variable whose domain includes only one value is called a ***constant***.

An ***open interval*** denoted by (1,5) is the interval between the number one and the number five and does not include the end points. A ***closed interval*** denoted by [1,5] does include the end points. The new ISO notation used is as

follows: $$\begin{cases} \left]a,b\right[= \{x \mid a < x < b\} \\ \left[a,b\right[= \{x \mid a \le x < b\} \\ \left]a,b\right] = \{x \mid a < x \le b\} \\ \left[a,b\right] = \{x \mid a \le x \le b\} \end{cases}$$

Normally an interval is called ***left open*** if it has no minimum such as (0,3] and ***right open*** if it has no maximum such as [0,3). Note the type of bracket.

Class intervals are a matter of judgment. The choice of interval sizes that will be used for any particular set of variates depends totally upon the nature and character of the data and the purpose for their use.

Cumulative frequency distributions are useful when we would like to see if a number of observations are *"less than or equal to" or "more than or equal to"* a given value. The cumulative frequency that corresponds to the upper boundary of any class interval is the total ***absolute frequency*** of all values less than that boundary. The inverse of the cumulating process is called ***differencing***.

In order to be able to graph variables we need to introduce the concept of function. A ***function*** is denoted by $y = f(x)$ and implies that the value of y is determined when the variable x is assigned a value. Therefore x is called the ***independent variable*** and y is called the ***dependent variable***.

The definition of function is as follows:

Let there be a set of values that are assumed by the variable x. Then if to each x in the set there corresponds one or more values of y, then y is said to be a function x in its domain.

If there is only one value of y to each corresponding value of x then the function is called single valued, otherwise y is called multi-valued.

Probabilities

What is probability? There are many descriptions and definitions but the one best suited for our use would be separating the probable outcome from the possible outcome.

A Look at Combinations

Let's flip a fair coin and ask what the probability is of getting *only* one head in five tosses. This can be restated as requiring five things to happen one at a time, or *n* items *r* at a time. This is written as the following binomial $\binom{n}{r}$

The number of possible combinations of *n* items *r* at a time can be computed by:

$$C_r^n = C(n,r) = \frac{n!}{r!(n-r)!}$$

The notation 5! means 5*4*3*2*1, thus n! = n* (n-1) * …*1. Therefore the answer to the coin flipping example would be $\binom{5}{1} = \frac{5!}{1!*(5-1)!} = \frac{120}{24} = 5$. This then says that there are five possible combinations in which we will get *only* one head.

An Oil & Gas Exploration Example

Suppose that we have a three-well drilling program and we consider the possible outcomes, either a dry hole or a successful well, assuming both are equally probable. With only three wells to drill there are $\boldsymbol{2^3}$ or eight possible sequences. This would lead us to the chance of one successful well in eight, or a probability of 0.125. These outcomes can be predicted by the use of Pascal's triangle, which is derived as follows:

$$\begin{aligned}
&(a+b)^0 = 1 \\
&(a+b)^1 = a+b \\
&(a+b)^2 = a^2 + 2ab + b^2 \\
&(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3 \\
&etc
\end{aligned}
\quad \Rightarrow Pascal's Triangle \left\{ \begin{array}{c}
1 \\
1 \quad 1 \\
1 \; 2 \; 1 \\
1 \; 3 \; 3 \; 1 \\
1 \; 4 \; 6 \; 4 \; 1 \\
1 \; 5 \;\; 10 \;\; 5 \; 1 \\
\text{etc}
\end{array} \right.$$

The combination of three things one at a time is written as: $\binom{3}{1} = \frac{3!}{1!(3-1)!} = \frac{6}{2} = 3$

The sum of possibilities of all outcomes must equal one, consequently if we look at a dry hole versus a successful well in our three-well program, we get (*success* + *dry*) =1. The probability of getting one successful well was 0.125, as stated earlier. The probability of getting three successful wells with both success and dry being equally probable, i.e. P(success) = P(dry) = 0.5, would be computed as follows:

$$\text{Probability of success} = P_r = C_r^n P^r\left(success\right)\left(1-P\right)^{(n-r)}$$

$$P\left(3 \text{ successes}\right) = C_3^3 P^3\left(success\right) P^{(3-3)}\left(failure\right) = \frac{3!}{3!(3-3)!}(0.5)^3\left(1-0.5\right)^{(3-3)} = 1*0.125*1 = 0.125$$

$$P\left(2 \text{ successes}\right) = C_2^3 P^2\left(success\right) P^{(3-2)}\left(failure\right) = \frac{3!}{2!(3-2)!}(0.5)^2\left(1-0.5\right)^{(3-2)} = 3*0.25*0.5 = 0.375$$

$$P\left(1 \text{ success}\right) = C_1^3 P^1\left(success\right) P^{(3-1)}\left(failure\right) = \frac{3!}{1!(3-1)!}(0.5)^1\left(1-0.5\right)^{(3-1)} = 3*0.5*0.25 = 0.375$$

$$P\left(3 \text{ Dry}\right) = C_3^3 P^0\left(success\right) P^{(3-0)}\left(failure\right) = \frac{3!}{3!(3-3)!}(0.5)^0\left(1-0.5\right)^{(3-0)} = 1*1*0.125 = 0.125$$

and knowing that the outcome is $DDD + (DDS + DSD + SDD) + (SSD + SDS + DSS) + SSS\left(=1+3+3+1\right)$

$$\therefore \begin{cases} \text{one success} = \frac{3}{8} = 0.375 \\ \text{Two successes} = \frac{3}{8} = 0.375 \\ \text{Three successes} = \frac{1}{8} = 0.125 \\ \text{Three Dry} = \frac{1}{8} = 0.125 \end{cases} = 1.0$$

Thus this is the same result as shown before, since by using Pascal's triangle we would have $\frac{1}{1+3+3+1} = 0.125$ for the possibility of three drilling successes.

The binomial probability distribution for 5 successes out of 7 wells with a probability of 40 % each can be written as follows.

$$P^B = \left(7,5,0.4\right) = C_5^7 S^5 F^{(7-5)} = \frac{7!}{5!(7-5)!}*0.4^5*0.6^{(7-5)} =$$

$$= 21*0.01024*0.36 = 0.0774144$$

If we now consider :

s = number of successes
N = number of trials
p = probability of success

We can then write the following binomial probability distribution: $P^B = \left(s,n,p\right)$

Thus our three well and three successes example will look like this:

$$P^B = \left(3,3,0.5\right) = C_3^3 S^3 F^{(3-3)} = 1*0.5^3*1 = 0.125$$

If we change the chance of success from 50% to only 25%, we then have: $P^B = (3,3,0.25) = C_3^3 S^3 F^{(3-3)} = \frac{3!}{3!(3-3)!} * 0.25^3 * 0.75^{(3-3)} == 1 * 0.25^3 * 1 = 0.015625$ Our chance for having three successful wells out of three wells drilled is now 1 in 64. or 0.015625=1.56%.

A more detailed approach would show the following computation:

$$\text{Probability of success} = P_r \ = C_r^n P^r\left(success\right)\left(1-P\right)^{(n-r)}\left(failure\right)$$

$$P\left(3 \text{ successes}\right) = C_3^3 P^3\left(success\right) P^{(3-3)}\left(failure\right) = \frac{3!}{3!(3-3)!}(0.25)^3\left(1-0.25\right)^{(3-3)} = 1*0.015625*1 = 0.015625$$

$$P\left(2 \text{ successes}\right) = C_2^3 P^2\left(success\right) P^{(3-2)}\left(failure\right) = \frac{3!}{2!(3-2)!}(0.25)^2\left(1-0.25\right)^{(3-2)} = 3*0.0625*0.75 = 0.140625$$

$$P\left(1 \text{ success}\right) = C_1^3 P^1\left(success\right) P^{(3-1)}\left(failure\right) = \frac{3!}{1!(3-1)!}(0.25)^1\left(1-0.25\right)^{(3-1)} = 3*0.25*0.5625 = 0.421875$$

As shown above, our chance for a success is now 1 in 64 or 0.015625=1.56%.

Let's take a look at a ten-well program with a success rate of 30%. What are our chances of getting exactly five successful wells?

$$P^B = (10,5,0.3) = C_5^{10} S^5 F^{(10-5)} = \frac{10!}{5!(10-5)!} * 0.3^5 * 0.7^{(10-5)} = 252 * 0.00243 * 0.16807 = 0.1029$$

thus the probability of getting five successful wells out of a ten-well drilling program when the chance for success is 30% is only 0.103=10.3%.

Let's reduce the number of wells in the program to seven and increase the chance for success to 40%. We then have:

$$P^B = (7,5,0.4) = C_5^7 S^5 F^{(7-5)} = \frac{7!}{5!(7-5)!} * 0.4^5 * 0.6^{(7-5)} = 21 * 0.01024 * 0.36 = 0.0774144$$

Thus the probability of getting five successful wells in a seven-well program with a chance for success of 40 % is only 0.077 or 7.7%.

Binomial Distributions

The distributions described above have only two outcomes, either success or failure, and will follow a certain pattern called the binomial distribution. By letting the number of events or trials increase to a very large or to an infinite amount and then creating a continuous distribution, we can define the normal curve.

It should be noted here that there are many different types of continuous distributions and the symmetrical normal curve is only one example of this. In chapter 4 I will treat the different types of distributions that show different shapes and lopsidedness or skewness such as log normal, Beta, Gaussian, Weibull, etc. and their applications, but for now we will concentrate on the basic concepts of the standard normal curve.

The ***binomial distribution*** is a ***Bernoulli distribution,*** which is equal to one (1) for the success probability *p* and equals zero (0) for the failure probability *q=1-p*. This then leads to the ***probability mass or density function*** defined as:

$$f(k;p)=\begin{cases} p & if\ k=1 \\ 1-p & if\ k=0 \\ 0 & otherwise \end{cases} \Rightarrow f(k;p)=p^k(1-p)^{1-k}$$

The expected value of the variate or random variable x is $E(x)=p$ and the variance is given by $Var(x)=p(1-p)$. Each term in the binomial distribution is represented by a height $f(x)$ of the rectangle with a unit base. The area under the curve is the sum of all the terms in the distribution from $x=a$ to $x=b$ inclusive, so in the interval $\left[a-\frac{1}{2},b+\frac{1}{2}\right]$, see figure 1-1 below, and with the mean of the normal distribution being $\mu=s*\theta$ and the variance of the normal distribution being $Var(x)=\sigma^2=s*\theta*(1-\theta)$, the area from $a-\frac{1}{2}$ to $b+\frac{1}{2}$ under the normal curve will be represented by the area from $[z_1,z_2]$, where $z_1=\frac{a-\frac{1}{2}-\mu}{\sigma}$ and $z_2=\frac{b+\frac{1}{2}-\mu}{\sigma}$.

The area is thus computed from $\int_{z_1}^{z_2}\phi(z)dz=\Phi(z_2)-\Phi(z_1)$. These values are obtained from the probability z tables.

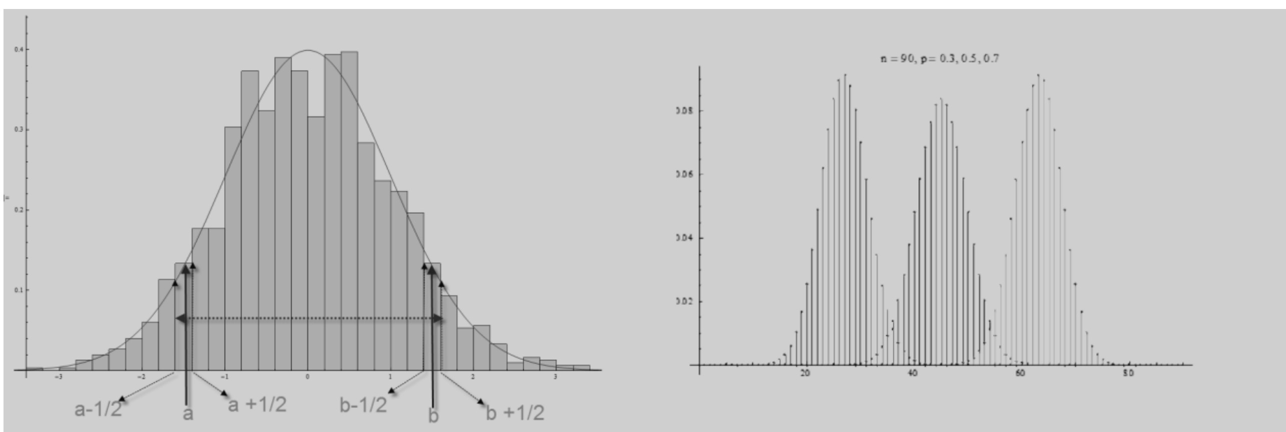

Figure 1-1 Left: interval $\left[a-\frac{1}{2},b+\frac{1}{2}\right]$ Right: Binomial distributions where n=90 and p=0.3, 0.5 and 0.7 respectively

At this point we will define some terms that are quite frequently used.

Range of Values

The range of values is defined as the variation in the data. It describes the minimum and maximum and all values in between.

Mean Deviation

The mean deviation is defined as a measure of dispersion from the mean of the data. It is computed by summing the differences of the individual values and the mean value and dividing this by the number of data points.

Thus:

$$\text{Mean Deviation} = MD = \frac{\text{Differences between the individual values and the mean value}}{\text{number of data points}}$$

Variance

The variance is a statistical term and not a descriptive term and is defined as the sum of the squares of the differences of the individual values and the mean value, divided by the number of data points.

Thus:

$$\text{Variance} = Var(x) = \frac{\text{The sum of the squares of the differences between the individual values and the mean value}}{\text{number of data points}}$$

$$\therefore \text{Variance} = Var(x) = \frac{\sum (x_i - \overline{x})^2}{N}$$

Standard Deviation

The standard deviation is defined as the square root of the variance.

Thus:

$$\text{Standard Deviation} = SD(x) = \sqrt{\frac{\sum (x_i - \overline{x})^2}{N}}$$

Normal Curve

If we take repeated measurements on very large samples of a natural population, we create a characteristic distribution pattern known as the ***bell curve*** , or a ***normal distribution***, or a ***Gaussian distribution***, mathematically defined as $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$, where $\mu, \sigma > 0$ and μ, σ are arbitrary and represent the population mean and the standard deviation respectively. Note the population (*see definition below*) mean is not the same as the sample mean!

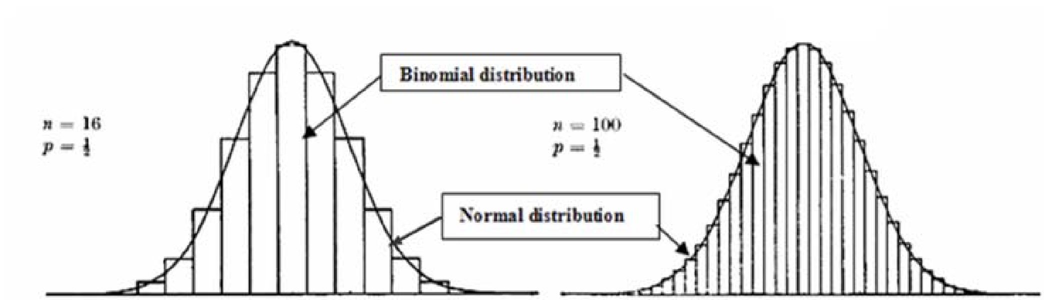

Figure 1-1A Normal and binomial distributions