

**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*

Copyright © 2010 Andreas Stark

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the publisher.

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

ACKNOWLEDGEMENTS	3
CONTENTS	4
LIST OF FIGURES	8
LIST OF TABLES	15
Introduction	17
Chapter 1	19
<i>A review of elementary statistics</i>	19
The Scope of Statistics	20
Frequency Distributions	20
Probabilities	21
A Look at Combinations	21
An Oil & Gas Exploration Example	21
Binomial Distributions	23
Range of Values	24
Mean Deviation	25
Variance	25
Standard Deviation	25
Normal Curve	25
Variance	29
Distributions	29
Standard Deviation	29
Covariance	31
Correlation Coefficient	32
Weighted Averages	33
Useful computations and formulas	34
Significance of Observed Proportions	34
Cumulative Distribution Shapes	36
Student's t-Distribution	36
Testing Normal Populations	37
Degrees of Freedom	39
Confidence Limits	39
The t Test with a Hypothetical Population Mean of 18%	39
The F Test	40
The Chi Squared Test	41
Review of Matrix Algebra	41
Simple vector notation	41
Simultaneous equations	42
Matrix multiplication	43
Matrix Inversion	44
Transpose of a Matrix	45
Determinants	46
Vector Space	47
Cramer's Rule	48
Eigenvalues and Eigenvectors	48
Principal Component Analysis	52
Markov Chains and Matrices	56
Sequences of the same state	58
CHAPTER 2	60
<i>A Review Of Spatial Point Distributions, Gridding And Mapping</i>	60
Introduction	61

Basic mapping concepts	61
Uncertainty and Mapping Accuracy	66
Distribution of Points	68
Uniform Distribution	69
Random Distribution	69
Gridding using Nearest Neighbor Method	71
Gridding using Inverse Distance Method	74
Gridding using Kriging and Co-Kriging	76
Gridding using Minimum Curvature	85
Gridding using Polynomial Regression	86
Gridding using the Radial Basis Function	88
Gridding using Shepard's Method	90
Gridding using Triangulation	91
Voronoi Diagrams	91
Contouring	92
Structural Noses	93
Regional Dip	94
Trend Surface Analysis	97
Warnings and Comparisons	97
Contour Maps as Surfaces	98
Regional Trends	98
Pitfalls in Trend Surfaces	105
CHAPTER 3	106
A Review Of Time Series	106
Introduction to Time Series Analysis	107
Taylor Series and MacLaurin Series	114
Linear Systems	115
Systems Obeying the Superposition Principle	116
Fourier Analysis	116
Dirichlet Conditions	116
Fourier Series	117
Euler Formula	118
Spectral Decomposition	119
Principle of Superposition	120
Convolution	122
Correlations	126
Filtering of time series	127
Linear Interpolation of points	129
Double Fourier and time series	132
Multivariate extensions of elementary statistics	138
Hotelling T Square Distribution	139
Discriminant Functions	140
Cluster Analysis	148
CHAPTER 4	151
Risk And Uncertainty	151
Introduction to Risk and Uncertainty	152
Dependence and Independence	153
Decision Trees	154
Conditional Probability or Bayes' Theorem	155
Stochastic Processes	157
Biases and Opinions	161
Chance of Success	163

Triangular Distribution	167
Log normality	169
Uncertainty and Risk	174
The Magnitude of uncertainty	176
Technical, Market, Environmental and Political Risks	178
Independent multiple risk estimates	179
Reserve estimation	180
Production areas	184
Play assessment	185
Play Risk	186
Play and Prospect Resource Estimates	190
Detailed Play Risk Analysis	199
Recovery factors	201
Economic profitability	202
Multiple prospective zones	213
Optimum Working Interest	220
CHAPTER 5	221
<i>Economic Analysis Under Uncertainty</i>	221
Introduction	222
Price, Utility and the Saint Petersburg Paradox	222
von Neumann and Morgenstern	223
Small versus Large Gambles	226
Prospect Theory	226
Ellsberg paradox	229
Process Utilities and Regret Theory	230
Risk Aversion	231
Utility Functions and Statistics	236
Portfolios	242
Portfolio Analysis	246
Determining Working Interest	258
Incomplete probability information	270
The Law of Large Numbers	271
Decision Theory	273
Value of Information	274
Bayes' Theorem and Terminal Action Cost	277
Minimization and Maximization	283
Risk Adjusted Value and Price Sensitivity	286
CHAPTER 5 - APPENDIX	303
NOTE	303
Future and Present Value	303
The Value of Money, Discount Rates and Opportunity Costs	308
Discounted cash flow models	311
CHAPTER 6	314
<i>Adventures And Misadventures With Real Options</i>	314
Introduction	315
Old and New Paradigms	316
Real Options	319
Discrete simulation of uncertainty using the binary lattice approach	322
Option to Acquire Additional Information	330
Financial versus Real Options	336
Option to Delay or Timing Option	339
Option to Expand	342

Option to Abandon	343
Option uses	345
Option to Choose	346
The Black-Scholes Model	347
Monte Carlo Simulations	349
CHAPTER 7	355
<i>Fractals And Neural Networks In Applied Risk Analysis</i>	355
Introduction	356
Fractals	356
Fractal Dimension	357
Hurst Exponent and Rescaling	359
R/S and Financial Markets	360
Fractal Statistics	362
Fractal Analysis	365
Dynamical Systems	366
Hénon map	367
Fractal Application Example	370
Oil and Gas Example of Fractal Application	373
Artificial Neural-Network Paradigm	377
The Single Neuron as a Classifier	378
Basic Neural Network Concepts	379
Types of Networks	382
Network Learning	384
Hopfield Networks	387
Genetic Algorithm	389
Conclusion	392
References and Bibliography	394
Index	401

LIST OF FIGURES

<i>Figure 1-1 Left: interval $\left[a - \frac{1}{2}, b + \frac{1}{2}\right]$</i>	<i>Right: Binomial distributions where n=90 and p=0.3, 0.5 and 0.7 respectively</i>	24
<i>Figure 1-1A Normal and binomial distributions</i>		25
<i>Figure 1-2: Standard normal form and the standard curve.</i>		27
<i>Figure 1-3 Standard deviation and probabilities</i>		27
<i>Figure 1-4 Mode, median and mean</i>		28
<i>Figure 1-5 Binomial normal curve for samples s=5 with p=0.5 Left and s=10, s=25, s=100 and s=1000 with p=0.5 Right</i>		35
<i>Figure 1-6 cumulative more than and less than curves</i>		36
<i>Figure 1-7 3D plot of the three eigenvectors</i>		51
<i>Figure 1-8 Principal component analysis</i>		55
<i>Figure 1-9 plot of the two covariance matrix row vectors and eigenvectors</i>		55
<i>Figure 2-1 example of a topographic map sheet</i>		62
<i>Figure 2-2 Blank map sheet with Township and Range grid drawn on it</i>		63
<i>Figure 2-2A Blank map sheet showing Township and Range grid and LSD subdivisions</i>		63
<i>Figure 2-3 Fully posted map sheet with Township grid, LSD's, culture, wells and time value postings</i>		64
<i>Figure 2-4 grid data, grid nodes with contours and a grid search pattern</i>		65
<i>Figure 2-5 Top - original data; Middle - original data in black and gridded data in the middle; Bottom - contours with grid and original data in black and grid nodes in color</i>		66
<i>Figure 2-6 Certainty and map accuracy</i>		67
<i>Figure 2-7 Point distributions</i>		69
<i>Figure 2-8 Left - Random Point distribution and Right - clustered distributions created from left image</i>		70
<i>Figure 2-9 nearest neighbor search</i>		72
<i>Figure 2-10 Left - clustered distributions of the same class and Right - clustered distributions of different classes</i>		73
<i>Figure 2-11 Left - a random distribution and Right - same distribution with the 6 nearest neighbors of the point (0,0) in Red</i>		74
<i>Figure 2-12 A data set gridded by nearest neighbor method, grid spacing 9 by 9 left, and grid spacing 12 by 12 right; note the differences.</i>		74
<i>Figure 2-13 An example of gridding with inverse distance to a power, grid spacing 3 by 2.</i>		75
<i>Figure 2-14 Example of a Variogram</i>		77
<i>Figure 2-15 Different types of variograms</i>		78
<i>Figure 2-16 Example of gridding with Kriging, grid spacing 3 by 2; compare to the previous two methods and see the improvements.</i>		78
<i>Figure 2-17 nugget and sill</i>		79
<i>Figure 2-18 Variogram</i>		82
<i>Figure 2-19 Variogram</i>		82
<i>Figure 2-20 An example of minimum curvature gridding, grid spacing 3 by 2 compare to the previous methods above.</i>		86
<i>Figure 2-21 Linear regression</i>	<i>Figure 2-22 Parabolic regression</i>	87
<i>Figure 2-23 Cubic regression</i>	<i>Figure 2-24 Quartic regression</i>	87
<i>Figure 2-25 Linear regression</i>	<i>Figure 2-26 Quartic regression</i>	88
<i>Figure 2-27 Quadratic regression</i>	<i>Figure 2-28 Cubic regression</i>	88
<i>Figure 2-29 grid spacing 3 by 2.</i>	<i>Figure 2-30 grid spacing 3 by 2.</i>	89
<i>Figure 2-31 grid spacing 3 by 2.</i>	<i>Figure 2-32 grid spacing 3 by 2.</i>	89
<i>Figure 2-33 grid spacing 3 by 2.</i>		89
<i>Figure 2-34 Radial basis functions</i>		90
<i>Figure 2-35 Shepard's method again compare with previous methods above</i>		91
<i>Figure 2-36 Delaunay Criterion a) violates the criterion and b) honors the criterion</i>		91
<i>Figure 2-37 Voronoi Triangulation</i>	<i>Figure 2-38 Triangulation contours with the same data set as before</i>	92
<i>Figure 2-39 Effect of tilting on structure</i>		93
<i>Figure 2-40 Three point solution (again the same data set was used)</i>		94

Figure 2-41 Manual trend surface drawing	95	
Figure 2-42 Computer mapped structure	95	
Figure 2-43 Machine Trend surface with structure	95	
Figure 2-44 Structure minus trend residual with structure overlain	96	
Figure 2-45 Second derivative of structure with structure overlain	96	
Figure 2-46 An example of polynomial linear regression gridding	Figure 2-47 Linear trend surface	103
Figure 2-48 First order residuals	Figure 2-49 Quadratic trend surface	103
Figure 2-50 Second order residuals	Figure 2-51 Cubic trend surface	104
Figure 2-52 3rd order residuals		104
figure 3-1 time series plot		107
figure 3-2 Life expectancy against birthrates for all countries in the world		108
figure 3-3 Variations of GE stock from the start of the year 2000 up to now		108
figure 3-4 Variations of Dow Jones from the start up to now; note the market crashes		109
figure 3-5 scatter plot of some sequence data, vertical axis are the observations horizontal axis is the time interval		110
figure 3-6 Line plot of some time series data, vertical axis are the observations horizontal axis is the time interval		111
figure 3-7 Line plot of data series with a time trend; vertical axis is the observations, horizontal axis is the time intervals		111
figure 3-8 from top left clockwise: Saw tooth function, Rectangular function, Heaviside function and Square wave function.		115
figure 3-9 A linear system		116
Figure 3-10 Top - Extrema: Minima and Maxima; Bottom - Discontinuity		117
Figure 3-11 Dispersion in a prism, or splitting of white light into its frequency components		119
Figure 3-12 Time function (upper left) transformed into its Amplitude (lower left) and Phase (lower right) Spectra - upper right is the combined display of amplitude and phase spectrum		119
Figure 3-13 A discrete Amplitude Spectrum and the various components		120
Figure 3-14 Harmonic decomposition and the principle of superposition		121
Figure 3-15 Saw tooth function created from the harmonics in fig 3-10		121
Figure 3-16: Complex harmonic motion (a) and simple harmonic motion (b)		122
Figure 3-17 Filter or Analog signal - top with its amplitude spectrum- middle and phase spectrum - bottom		123
Figure 3-18 convolution by using the cross multiply add method		125
Figure 3-19 Bit and byte multiples and their values		126
Figure 3-20 Time function and its Autocorrelation function		127
Figure 3-21 Time function and its Auto convolution function		127
Figure 3-22 data smoothing		128
Figure 3-23 data periodicity and trending		129
Figure 3-24 linear interpolation of data points, locations 3 and 4 have been filled in.		129
Figure 3-25 Moving averaging of data points		130
Figure 3-26 Data points and linear and quadratic curve fitting		131
Figure 3-27 Data points and least squares, spline and Hermite polynomial fitting		131
Figure 3-28 Stock trading data points and various curve fitting and smoothing applications		132
Figure 3-29 Fourier spectra for a spike		132
Figure 3-30 Aliasing and Nyquist frequencies		133
Figure 3-31 Spatial Aliasing due to incorrect sample intervals		133
Figure 3-32 Spectra for spike at zero time		134
Figure 3-33 Spectra for a spike at time other than zero		134
Figure 3-34 A boxcar creates a spectrum shaped like a sync function		134
Figure 3-35 A narrower boxcar or spike creates a sharper sync function as a spectrum		135
Figure 3-36 A broader spike or boxcar creates a wider sync function as a spectrum		135
Figure 3-37 Spike with 1D and 2D Fourier spectrum		136
Figure 3-38 Comparison of sine wave interference and plane wave interference		137
Figure 3-39 Superposition of wave vectors at time zero above and interference for five random sinusoidal waveforms below.		137
Figure 3-40 combined effect of two probability density functions; also called a bivariate distribution.		138

Figure 3-41 Bivariate normal distribution - 3D display and contour ellipsoid display.	138
Figure 3-42 Discriminant function separating three sample classes at top and many at bottom	141
Figure 3-43 two sample classes separate on the left and combined on the right	141
Figure 3-44 two overlapping sample classes, a discriminant function line and the separated result	142
Figure 3-45 A bubble chart produced from scatter matrices	143
Figure 3-46 Bivariate data distribution	145
Figure 3-47 Bivariate zero mean data distribution	147
Figure 3-48 Bivariate zero mean data distribution scaled to Discriminant function.	148
Figure 3-49 Bivariate data distribution scaled to Discriminant function	148
Figure 3-50 example of two dendograms; one horizontal type and one vertical or tree type.	149
Figure 3-51 random data set left and its clustered data set right	149
Figure 3-52 Three clusters left and five clusters right	149
Figure 3-53 Manhattan Distance clusters left and Chebychev distance clusters right	150
Figure 3-54 Cosine Distance clusters left and Euclidean distance clusters right	150
figure 4-1a Processes or ventures A and B shown separately and together	152
Figure 4-1b Set A , Set B, Set A "AND" B, and the most common logical conditions, AND, OR, Exclusive OR, Implies, Negated AND, Negated OR	154
Figure 4-2 Decision Tree for drilling prospect.	155
Figure 4-3 Decision Tree for 3 experiments with 5 outcomes	158
Figure 4-4a Decision Trees: monetary returns	159
Figure 4-4b Decision Trees: monetary returns and estimated probabilities	160
Figure 4-5 Decision Trees: Optimal strategy computation for exploration example with payout	161
figure 4-6 Types of success and their relationship to probability (illustration only and not to scale)	163
figure 4-7 independent factors and the resulting combined probability	165
figure 4-8 Geologic risk factors	165
figure 4-9 Reserve estimates	166
figure 4-10 Triangular distributions: left = skewed to the left; middle = symmetrical; right = skewed to the right	167
figure 4-11 Triangular distributions showing an acres distribution skewed to log normal	167
figure 4-12 Normal Distribution, Triangular Distribution and Log Normal Distribution	167
figure 4-13 estimated absolute minimum and maximum	168
Figure 4-14 Triangular Log normal distribution	169
Figure 4-15 Gaussian distribution	170
Figure 4-16 A Gaussian distribution and its first three derivatives	170
Figure 4-17 A Log Normal distribution; left - normal plot and right - logarithmic plot	171
Figure 4-18 Beta distribution	171
Figure 4-19 Poisson distribution	172
figure 4-20 Binomial distribution	172
figure 4-21 A Gamma distribution	172
figure 4-22 Chi-Square distribution	173
figure 4-23 Weibull distribution	173
figure 4-24 Triangular distribution	173
figure 4-25 Exponential distribution	174
figure 4-26 Choosing P90, P50 and P10	177
figure 4-27 logarithmic P90, P50 and P10 display	178
figure 4-28 Prospect risk flow.	180
figure 4-29 Reserve Types	181
figure 4-30a Cartoon view of Basin with kitchen, plays and prospects	183
figure 4-30b Factors necessary for prospect existence	184
figure 4-31 Uncertainty variation for prospects having increasing risk showing ranges for their mean	189
figure 4-32 Some reservoir trapping styles and seals	192
figure 4-33 Play setting with several prospects	195
figure 4-34 Three point log normal deterministic input	195
figure 4-35 3-point log normal stochastic results in bbls and MMbbls	196
figure 4-36 Monte Carlo simulation results - 10,000 trials 8 reservoir factors vs. 3 reservoir factors	196
figure 4-37 Prospect three point log normal computation	197

figure 4-38 plot of prospect three point log normal reserve size distribution	198
figure 4-39 log-plot of reserves by using the Rose method	199
figure 4-40 log plot of some Alberta Oil fields in millions of cubic meters (1 cubic meter = 8.65 Barrels)	199
figure 4-41 Geologic chance factors used in play risk assessment	200
figure 4-42 Use of geologic chance factors in play risk assessment	201
figure 4-43 Log normal plot with truncated reserve distribution	202
figure 4-44 Expected Value example	203
figure 4-45 Expected Value example	205
figure 4-46 3D Risk Matrix	206
figure 4-47 A real analysis with current gas prices for a one section prospect size produces a loss after taxes and royalties (or wellhead taxes), even though the risked expected value was quite favorable	208
figure 4-48 Cartoon of Hydrocarbon generation parameters	210
figure 4-49 Western Canadian Basin field size statistics computed from Gov data	211
figure 4-50 Western Canadian Basin field size Lognormal distribution in millions of cubic meters computed from Gov data	211
figure 4-51 Lognormal gas distribution for prospect	214
Figure 4-52 Lognormal field size distributions	214
figure 4-53 Lognormal distribution ratios for Swanson Mean	216
figure 4-54 Lognormal distribution Swanson Mean and Mean comparison	217
figure 4-55 Two zone prospect computation for independent parameters only	218
figure 4-56 Two zone prospect computation with dependent and independent parameters	219
figure 4-57 Two zone prospect computation with dependent, independent parameters and common factors	219
figure 5-1 different growth rates, blue -diminishing or logarithmic, red - linear, green - exponential	223
figure 5-2 different utility function shapes- Risk Lover - Risk Averse - Risk Neutral for the example above	224
figure 5-3 Value function in prospect theory after Kahneman and Tversky	228
figure 5-4 Decision weight versus objective probability according to prospect theory after Kahneman and Tversky	229
figure 5-5 Risk averse utility function and its derivatives	232
figure 5-6 Utility functions (Blue = $C + e^{2w}$, Red = $C + \log(w)$, Green = $Cw + cw^2$, Black = w) and their associated Arrow-Pratt Risk Aversions	233
figure 5-7 Certainty Equivalent for Risk Aversion	234
figure 5-8 Risk lover utility function and its derivatives	234
figure 5-9 Certainty Equivalent for Risk Lover	235
figure 5-10 Utility curves showing absolute and relative risk aversion	236
figure 5-11 Distributions showing different skewness and kurtosis	239
figure 5-12 Two Asset Portfolio Optimization for $R_a=5, 10, 15$ and 50 resp.	243
figure 5-13 Markowitz efficient frontier for two assets	247
figure 5-14 Iso-Variance Curves	247
figure 5-15 Eight Prospect Portfolio computation	249
figure 5-16 Covariance Matrix of the eight prospect portfolio	249
figure 5-17 Weighted Covariance Matrix of the eight prospect portfolio	250
figure 5-18 Exponential Utility function	251
figure 5-19 Exponential Utility function, first derivative (green), second derivative (blue) and Arrow-Pratt risk aversion (red)	252
figure 5-20 Risk Adjusted Value or Certainty Equivalent versus WI for various NPV with a Risk Tolerance of \$50MM.	252
figure 5-21 Risk Adjusted Value or Certainty Equivalent versus WI for various probabilities of success.	253
figure 5-22 Risk Adjusted Value or Certainty Equivalent versus WI for various Risk Tolerance values.	254
figure 5-23 Deterministic and Stochastic efficient frontier solutions for the eight prospect portfolio	255
figure 5-24 Deterministic efficient frontier solutions for the eight-prospect portfolio	256
figure 5-25a Deterministic and Stochastic efficient frontier solution for the eight-prospect portfolio	257
figure 5-25b Deterministic and Stochastic efficient frontier solution for the eight prospect portfolio	258
figure 5-26 Gambler's Ruin- Working Interest	260
figure 5-27 Gambler's Ruin - Number of Wildcats	260

$figure\ 5-28\ Exponential\ Utility\ 1 - p * e^{-\left(\frac{NPV}{RT}\right)}$	262
$figure\ 5-29\ Certainty\ Equivalent\ Value\ CE\ or\ RAV$	263
$figure\ 5-30\ Optimal\ Working\ Interest\ and\ Certainty\ Equivalent\ Value\ CE\ or\ RAV$	264
$figure\ 5-31\ Certainty\ Equivalent$	264
$figure\ 5-32\ Certainty\ Equivalent\ for\ different\ NPV;\ curves\ $25\ MM\ and\ up\ have\ positive\ CE\ ranges$	265
$figure\ 5-33\ Certainty\ Equivalent\ for\ different\ RT\ curves$	265
$figure\ 5-34\ Optimal\ Working\ Interest\ for\ different\ RT\ curves$	266
$figure\ 5-35\ Apparent\ Risk\ Tolerance\ for\ different\ Working\ Interest\ rates$	266
$figure\ 5-36\ Certainty\ Equivalent\ for\ prospect\ 5$	267
$figure\ 5-37\ Apparent\ Risk\ Tolerance\ and\ WI\ as\ a\ function\ of\ P(success)$	268
$figure\ 5-38\ Expected\ Hyperbolic\ Utility\ versus\ NPV\ as\ a\ function\ of\ P(success)$	269
$figure\ 5-39\ Certainty\ Equivalent\ versus\ NPV\ as\ a\ function\ of\ P(success)$	270
$figure\ 5-40\ Decision\ Tree\ and\ various\ options$	274
$figure\ 5-41\ Decision\ Tree\ for\ cost\ matrix$	276
$figure\ 5-42\ Decision\ Tree\ for\ new\ cost\ matrix$	276
$figure\ 5-43\ Decision\ Tree\ for\ binomial\ probabilities\ for\ a\ sample\ of\ 2$	279
$figure\ 5-44\ Decision\ Tree\ for\ prior\ probabilities\ for\ a\ sample\ of\ 2$	280
$figure\ 5-45\ Decision\ Tree\ for\ terminal\ action\ costs$	281
$figure\ 5-46\ Reduced\ Decision\ Tree\ for\ terminal\ action\ costs$	282
$figure\ 5-47\ Linear\ Programming\ and\ Geometric\ truth\ set$	284
$figure\ 5-48\ Linear\ Programming\ Extreme\ Points\ of\ a\ Polyhedral\ convex\ set$	284
$figure\ 5-49\ Linear\ Programming\ Extreme\ Points\ for\ a\ practical\ maximization\ problem$	286
$figure\ 5-50\ Certainty\ trends\ for\ the\ projected\ NPV\ of\ the\ oil\ field\ example\ with\ hyperbolic\ decline$	287
$figure\ 5-51\ 10\ year\ cash\ flow\ analysis\ for\ a\ producing\ oil\ well$	289
$figure\ 5-52a\ Net\ Present\ value\ for\ the\ oil\ field\ example$	291
$figure\ 5-52b\ Net\ Present\ value\ for\ the\ oil\ field\ example$	292
$figure\ 5-53a\ Monte\ Carlo\ oil\ price\ prediction$	293
$figure\ 5-53b\ Monte\ Carlo\ oil\ price\ prediction$	294
$figure\ 5-54a\ Volatility\ of\ logarithmic\ cash\ flow\ approach$	295
$figure\ 5-54b\ Volatility\ of\ logarithmic\ cash\ flow\ approach$	296
$figure\ 5-55a\ Volatility\ of\ logarithmic\ present\ value\ approach$	297
$figure\ 5-55b\ Volatility\ of\ logarithmic\ present\ value\ approach$	298
$figure\ 5-56a\ PV\ for\ time\ period\ 0\ of\ logarithmic\ present\ value\ approach$	299
$figure\ 5-56b\ PV\ for\ time\ period\ 0\ of\ logarithmic\ present\ value\ approach$	300
$figure\ 5A-1\ Compounding\ periods\ versus\ Future\ Value\ factors$	305
$figure\ 6-1\ Simple\ binary\ options\ choose\ between\ a)\ -$150\ and\ +$200\ and\ b)\ -$200\ and\ +$150$	320
$figure\ 6-2\ two\ level\ binary\ option$	321
$figure\ 6-3\ two\ level\ real\ option$	322
$figure\ 6-4\ trend\ with\ uncertainty$	323
$figure\ 6-5\ Geometric\ Brownian\ Motion\ with\ drift\ used\ to\ model\ project\ return\ for\ one\ year$	324
$figure\ 6-6\ A\ simple\ Binomial\ lattice\ for\ a\ $100\ stock\ without\ growth\ and\ uncertainty$	324
$figure\ 6-7\ A\ simple\ Binomial\ lattice\ for\ a\ Call\ option\ of\ a\ $100\ stock\ with\ a\ 5%\ risk\ free\ rate\ and\ a\ 25%\ volatility$	325
$figure\ 6-8\ The\ uncertainty\ cone\ for\ the\ up\ and\ down\ factors$	326
$figure\ 6-9\ Computational\ flow\ for\ the\ up\ and\ down\ factors$	327
$figure\ 6-10\ Lattice\ computation\ for\ the\ actual\ option\ value$	328
$figure\ 6-11\ Uncertainty\ cone\ behavior\ as\ the\ number\ of\ time\ steps\ change\ (counting\ from\ center\ 100,\ 50,\ 25)$	329
$figure\ 6-12\ Uncertainty\ cone\ behavior\ as\ the\ volatility\ changes\ (counting\ from\ center\ 12.5%,\ 25%,\ 50%)$	330
$figure\ 6-13\ Seismic\ data\ acquisition\ option$	331
$figure\ 6-14\ present\ value\ cost\ of\ perfect\ seismic\ information$	332
$figure\ 6-15\ Seismic\ data\ acquisition\ option\ part\ 1$	334
$figure\ 6-16\ Seismic\ data\ acquisition\ option\ part\ 2$	334
$figure\ 6-17\ Seismic\ data\ acquisition\ option\ part\ 3$	335
$figure\ 6-18\ Alternate\ seismic\ data\ acquisition\ option\ flow\ part\ 1$	335

figure 6-19 Alternate seismic data acquisition option flow part 2	336
figure 6-20 Option payoff chart representing the option's NPV	337
figure 6-21 Option payoff chart for a \$100 stock with a strike price of \$100, a risk free rate of 3% and a volatility of 30%	338
figure 6-22 Computation chart for a \$100 asset with a project cost of \$100, a risk free rate of 3% and a volatility of 30%	338
figure 6-23 Computation chart for a project delay decision	340
figure 6-24 plot of option value sensitivity against exercise time figure 6-25 plot of option value sensitivity against underlying asset value	340
figure 6-26 Delay option binomial lattice computation	341
figure 6-27 Expansion option computation	343
figure 6-28a Abandonment option computation	344
figure 6-28b Abandonment option computation	345
figure 6-29 Chooser option computation - Expand, Contract, Abandon	347
figure 6-30 Option sensitivity measures Delta, Gamma, Rho, Vega and Theta	349
figure 6-31 Initial distribution for Monte Carlo Simulation	350
figure 6-32 Monte Carlo Simulation example	351
figure 6-33 Monte Carlo Simulation example	351
figure 6-34 Monte Carlo Simulation portfolio maximizing example results	352
figure 6-35 Monte Carlo Simulation portfolio maximizing example results	353
figure 6-36 Monte Carlo Simulation portfolio maximizing example results	353
figure 6-37 Monte Carlo Simulation showing certainty of not losing any money	354
figure 6-38 Monte Carlo Simulation portfolio showing certainty of making a return of at least \$5,000	354
figure 7-1 Fractals	357
figure 7-2 Fractal Brownian Motions with different scaling parameters showing persistency and anti-persistency	357
figure 7-3 Fractal dimension calculation	358
figure 7-4 Mandelbrot-Weierstrass function calculated with different Fractal dimensions	358
figure 7-5 Hurst Rescaled Range	360
figure 7-6 Local Hurst Exponent distribution, box sizes are 512, 1024, 4096 and 8192 points (adapted from Bartolozzi,2010)	362
figure 7-7 Pareto distribution	362
figure 7-8 Density function for a stable distribution	363
figure 7-9 Stable random variables used to simulate stock prices	363
figure 7-10 Logistics map	366
figure 7-11 Two discrete time series	367
figure 7-12 Hénon attractor map	368
figure 7-13 Hénon Lag	368
figure 7-14 log-log plot of the estimated slope of the correlation dimension D	369
figure 7-15 Fractal Brownian (solid blue line) and Geometric Brownian simulations of a Black-Scholes real option	370
figure 7-16 Corrected Black-Scholes(solid blue line) and Black-Scholes(dashed red line) real option	371
figure 7-17 Corrected Black-Scholes (solid blue line) and Black-Scholes (dashed red line) real option with a kurtosis of 7	372
figure 7-18 Different kurtosis and skewness of a PDF	372
Figure 7-19 Spot prices for Cushing, WTI , Top: unaltered, Middle: Rescaled, Bottom: Frequency plot	374
Figure 7-20 Normally Distribution and Fractional Distribution for the timing option.	375
Figure 7-21 Option value curves versus time, Deterministic NPV curve and Flex value curves (difference between option value and NPV)	376
figure 7-22 Diagram of a typical myelinated vertebrate motoneuron (from Wikipedia and has been released into the public domain by its author)	378
figure 7-23 Diagram of a neuron (from Wikipedia and has been released into the public domain by its author)	378
figure 7-24 Diagram of various activation functions	380
figure 7-25 Diagram of a three layer network	380
figure 7-26 Diagram of a three layer network; layer 1 is input, layer 2 is the hidden layer and layer 3 is the output layer	381

<i>figure 7-27 Diagram of a five layer Feed Forward Multilayer Perceptron Neural Network</i>	381
<i>figure 7-28 Diagram of a five-layer Feed Forward Multilayer Perceptron Neural Network, a five-layer Recurrent Neural Network and a five-layer Boltzmann Neural Network</i>	382
<i>figure 7-29 Diagram of a Radial Basis Neural Network</i>	383
<i>figure 7-30 Gaussian Bell in 1D and 2D</i>	383
<i>figure 7-31 Input function (left) and desired output function (right)</i>	385
<i>figure 7-32 Input function (sine) and desired output function (cosine)</i>	385
<i>figure 7-33 Error function (left) and the convergence of the weight vector (right)</i>	386
<i>figure 7-34 Error function with the convergence of the weight vector (left) 3-D display of the error function (right)</i>	386
<i>figure 7-35 A simple Hopfield network</i>	387
<i>figure 7-36 A simple pattern recognition using a Hopfield network</i>	387
<i>figure 7-37 A color pattern recognition using a Hopfield network and showing connected weights</i>	388
<i>figure 7-38 Simulated Annealing</i>	388
<i>figure 7-39 Solving the traveling salesman problem (TSP) using simulated annealing</i>	389
<i>figure 7-40 Genetic Algorithm results for the concession development option problem as was solved by Dias</i>	391
<i>figure 7-41 Genetic Algorithm flow chart as used by Dias to solve a concession development option problem</i>	391

LIST OF TABLES

<i>Table 1 standard deviation</i>	30
<i>Table 2 covariance</i>	32
<i>Table 3 t distribution</i>	37
<i>Table 4 F distribution 5%</i>	40
<i>Table 5 F distribution 1%</i>	41
<i>Table 6 eigenvectors and Principal Component Analysis</i>	54
<i>Table 7 of Example computation for the nearest neighbor method</i>	72
<i>Table 8 of Example computation for the inverse distance method</i>	75
<i>Table 9 Semi Variance computation example</i>	80
<i>Table 10 Semi Variance computation examples for EW and NS data</i>	81
<i>Table 11 regional trend</i>	101
<i>Table 12 Original data and variance-covariance computations</i>	144
<i>Table 13 vector, variance, covariance and Discriminant function computations</i>	147
<i>Table 14 Reserve estimates; errors increase because of multiplicative conditions.</i>	191
<i>Table 15 Reserve estimates for the world's top Giant Oil Fields</i>	215
<i>Table 16 Function derivative signs and slope</i>	233
<i>Table 17 Types of Risk Aversion</i>	235
<i>Table 18 Moment computations</i>	241
<i>Table 19 Process cost matrix</i>	275
<i>Table 20 predicted NPV from 3-point lognormal input to DCF</i>	288
<i>Table 21 Compound interest PV and FV formulas</i>	305
<i>Table 22 A PV and FV Application</i>	306
<i>Table 23 A Discount and Opportunity cost example</i>	309
<i>Table 24 Monte Carlo sample computation</i>	350

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 **variave** 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

$$\text{follows: } \begin{cases}]a,b[= \{x \mid a < x < b\} \\ [a,b[= \{x \mid a \leq x < b\} \\]a,b] = \{x \mid a < x \leq b\} \\ [a,b] = \{x \mid a \leq x \leq b\} \end{cases}$$

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 $n! = n * (n-1) * \dots * 1$. 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 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:

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

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. $P(\text{success}) = P(\text{dry}) = 0.5$, would be computed as follows:

$$\begin{aligned}\text{Probability of success} &= P_r = C_r^n P^r (\text{success}) (1-P)^{(n-r)} \\ P(3 \text{ successes}) &= C_3^3 P^3 (\text{success}) P^{(3-3)} (\text{failure}) = \frac{3!}{3!(3-3)!} (0.5)^3 (1-0.5)^{(3-3)} = 1 * 0.125 * 1 = 0.125 \\ P(2 \text{ successes}) &= C_2^3 P^2 (\text{success}) P^{(3-2)} (\text{failure}) = \frac{3!}{2!(3-2)!} (0.5)^2 (1-0.5)^{(3-2)} = 3 * 0.25 * 0.5 = 0.375 \\ P(1 \text{ success}) &= C_1^3 P^1 (\text{success}) P^{(3-1)} (\text{failure}) = \frac{3!}{1!(3-1)!} (0.5)^1 (1-0.5)^{(3-1)} = 3 * 0.5 * 0.25 = 0.375 \\ P(3 \text{ Dry}) &= C_3^0 P^0 (\text{success}) P^{(3-0)} (\text{failure}) = \frac{3!}{3!(3-3)!} (0.5)^0 (1-0.5)^{(3-0)} = 1 * 1 * 0.125 = 0.125\end{aligned}$$

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

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

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.

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

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: $P^B = (s, n, p)$

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

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

If we change the chance of success from 50% to only 25%, we then have:

$$P^B = \binom{3}{3} S^3 F^{(3-3)} = \frac{3!}{3!(3-3)!} * 0.25^3 * 0.75^{(3-3)} = 1 * 0.25^3 * 1 = 0.015625 \quad \text{Our chance}$$

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:

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

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

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

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

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?

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

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:

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

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:

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

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

$$\text{the area from } [z_1, z_2], \text{ where } z_1 = \frac{a - \frac{1}{2} - \mu}{\sigma} \text{ and } z_2 = \frac{b + \frac{1}{2} - \mu}{\sigma}.$$

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.

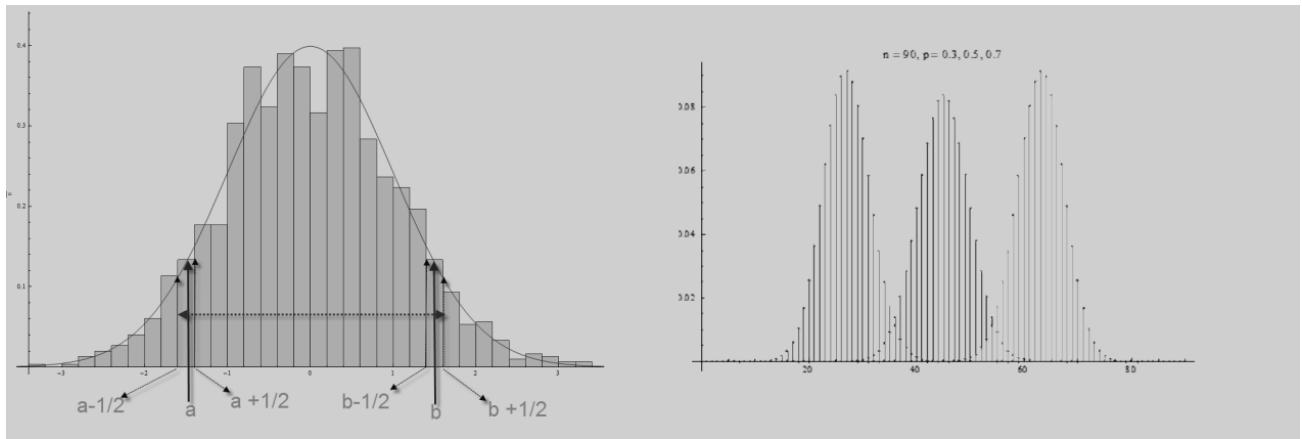


Figure 1-1 Left: interval $\left[a - \frac{1}{2}, b + \frac{1}{2}\right]$ Right: Binomial distributions where $n=90$ and $p=0.3, 0.5, 0.7$ respectively

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:

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

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

Standard Deviation

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

Thus:

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

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!

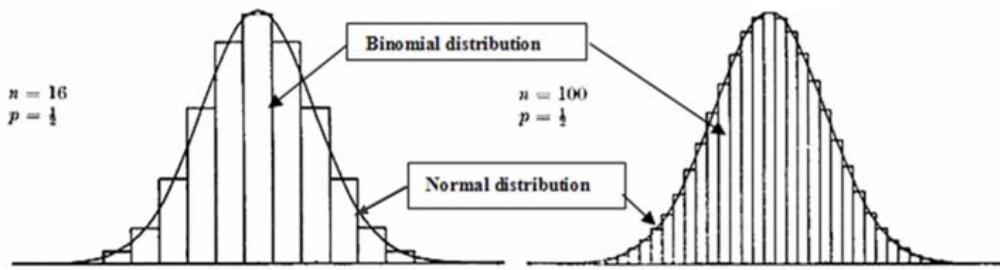


Figure 1-1A Normal and binomial distributions