ICEBERG RISK

An Adventure in Portfolio Theory

Kent Osband
kosband@nyc.rr.com

January 2002

PRE-PUBLICATION REVIEW MANUSCRIPT EXCLUSIVE TO READERS OF WILMOTT.COM

TO BE PUBLISHED BY TEXERE IN AUGUST 2002

TABLE OF CONTENTS

PREFACE: A DIALOGUE BETWEEN YOU AND ME

You: Hmmm. Iceberg Risk: An Adventure in Portfolio Theory. Strange title. What's this book about?

Me: Portfolio analysis. The theory of risk/reward tradeoffs in bundles of financial assets.

What's that have to do with icebergs?

I use icebergs as a metaphor for big, unusual, semi-obscured risks that standard portfolio theory can't deal with.

Can't deal with, or chooses not to?

Can't. The core assumption of normality rules them out by modeling overall risk as a bell curve.

I thought a big diversified portfolio is bound to be nearly normal. Isn't that what the Central Limit Theorem is about?

The Central Limit Theorem applies only when the underlying risks are independent.

Hold on a second. Normality can handle correlated risks.

Yes, provided the risks can be decomposed into lots of small independent building blocks. But often that's not possible.

For example?

For example, you own a hundred different Nasdaq stocks and Nasdaq crashes. More generally, whenever some pervasive common factor affects a huge chunk of your portfolio and that factor's not even approximately normal.

Oh, I knew that. I thought you were going to tell me something profound.

The profound part is that standard theory misses it.

You keep saying that but I don't believe it. I just came across a whole rack of books on "Value at Risk" methodology. Doesn't that tackle icebergs?

Yes, but in an ad hoc way that doesn't mesh with standard theory.

So what? Black cat, white cat, I don't care as long as it kills mice.

Bad analogy. Good portfolio managers don't exterminate all risks. They balance them against rewards. Typical value at risk methodology either doesn't measure rewards or implies some irrational investment decisions.

How do you define "irrational"?

By a willingness to make trades that are bound to lose you money.

Name me an investor who's never been irrational.

I can't. But that's no excuse for advising it. Theory's supposed to help, not make things worse.

OK, so I take it your book tries to integrate iceberg risk within standard theory.

The first half of the book does. Only it fails.

What do you mean?

I told you before. It can't be done. Standard finance theory assumes normality. Iceberg risks aren't normal.

If it can't be done, why did you bother trying?

Because I want to clear readers' minds of the illusion it can be done.

And that takes half a book?

"I wonder whether half a book's enough. Normality's an awfully entrenched prejudice in theory, even for people who reject it in practice."

Why do you think that is?

Because it's so simple. Normality allows you to calculate every possible risk/reward tradeoff just from data on means, variances, and correlations between assets.

I wouldn't call that simple. Won't 100 assets have around 5,000 distinct correlations?

Yes, but without normality even trillions of parameters might not suffice to model the tail risks of different portfolios. Because you still might not be able to answer questions like: If the first 50 assets go down the drain, what are the odds that the other assets go down the drain too?

What if I gather additional information about the tail risk on each asset?

That's less useful than it sounds. Fat-tailed assets can make for very smooth portfolios. Conversely, nearly normal assets can generate huge tail risks.

What matters then?

I told you. Common factors that make most of your assets go down together.

And how do you propose to model those?

The only way that's tractable. Conditional normality.

You just told me that normality doesn't work.

That was unconditional normality. Conditional normality is different.

How is it different?

Think of unconditional as overall averages. Think of conditional as averages under particular market regimes, like "bull market" or "bear market".

This is confusing. Do bell curves describe risks or don't they?

Conditional risks, yes. But the overall risk is an overlay or weighted average of different bell curves. That needn't resemble a single bell curve any more than a mountain range need resemble a single mountain.

With enough bell curves I imagine you can imitate any clump of risks.

You can imitate uniform risks too. But you would need a lot of different regimes. In practice it's more useful to segregate out a few regimes at a time.

I don't get it.

You will if you read the first half of the book.

Pardon my asking, but how much math does it involve?

About as much or as little as you want.

How is that possible?

I've divided every chapter into two sections. The first section explains the intuition and presents a couple of illustrative charts. The second section contains the supporting math, carved up into bite-sized nuggets.

So if I'm really smart I should focus on the math.

No, if you're really smart you should focus on the intuition. Once you understand that everything else is easy. Math's just a language for expressing things more clearly. I translate slowly in case you're interested but rusty. If you know the math already or couldn't care less, skip it.

Thanks for the confidence booster. I'll try to keep that in mind.

But I don't want to boost your confidence. I want to deflate it. Imagine discovering that the risks you've been hired to analyze and help manager are most likely bigger than you thought but harder to pin down. That the models you've been using don't work but you can't come up with a viable alternative.

Do you think that's really the norm in financial risk management?

No. I think the norm is trying to cover those truths up. So that at least you can manage the appearance of risk.

You sound cynical.

Just experienced. Besides, it is important to manage the appearance of risk. At least if you want to keep your job.

Then what's the point of deflating my confidence?

So that the second half of the book can rebuild it on sounder foundations.

How?

By modifying standard theory enough to incorporate iceberg risk.

Are the modifications simple?

As simple as they can be.

Simple enough for ordinary people to use them?

They'd probably need help with the math.

I thought you said the math wasn't essential.

It's not essential to understanding what's wrong. It is essential to getting things right.

And if I can't understand the math?

Then delegate the details to someone who does, and just stay focused on the intuition.

Great, I'll do that. By the way, can you summarize your core modifications in a few words?

Sure. Partition functions.

Never heard of them.

I'm not surprised. Partition functions come from thermodynamics. And what they bring to portfolio theory isn't obvious.

So what do they bring to portfolio theory?

The ability to incorporate regime-switching and options.

How?

I'm not telling yet.

What if I'm too busy to wait?

Then why are you browsing thru finance books?

Because I'm looking for answers.

Any answers? Or the right answers?

Don't get cheeky with me. The right answers, of course.

To what questions?

Practical questions. How to mitigate portfolio tail risks by shifting weights across assets. How to weigh risks against rewards.

Given what information?

The usual stuff. Information on the risks and rewards of the various assets and the overall correlations between them.

But it can't be done. That's what I've been trying to tell you.

Nonsense. People answer those questions all the time with no more information than I have.

I didn't say they couldn't answer those questions. They just can't answer them right.

Approximate answers are OK too.

The answers won't even be approximately right. Not without some hidden assumptions.

Surely you're exaggerating.

No, I'm not. For example, suppose you have three fair, uncorrelated coins. What's the probability they all turn all up tails?

One-eighth of course. One-half times one-half times one-half.

Nope. Chapter 1 gives some counterexamples.

I don't see how.

Good. That means the book can teach you something.

Even if the counterexamples are valid, why shouldn't I dismiss them as minor exceptions?

Because subsequent chapters in Part I show they're not.

But Part II shows how to deal with them.

Given the minimal extra information needed, yes.

You mean information about the conditional bell curves.

Yes, but that's still not quite enough information. We also need information on risk aversion. How willingly you would risk, say, a quarter of your wealth in return for a colossal gain.

And how do you tie those two together?

I told you: thru partition functions.

Which I told you mean nothing to me.

Alright. They're weighted sums of exponentials, where each regime contributes one exponential and the weights depend on the probabilities of the various regimes.

That's better. What determines the exponents?

The risk-adjusted expected returns for every regime times a factor related to the risk aversion.

Why are you adjusting twice for risk?

One factor adjusts for risk within regimes. The other adjusts for risk across regimes.

How do you adjust for risk within regimes?

Subtract a multiple of the conditional variance from the conditional mean.

How big a multiple?

That depends on the risk aversion too. Look, I said I'd give you a hint, not tour every nook and cranny.

Sorry. But if you don't mind my asking, why don't you just average out the various risk-adjusted returns?

That's basically what the method does. Only the weights don't just depend on the odds of the various regimes. They also depend on how bad the regime is. The worse the risk-adjusted return, the more weight the regime gets.

What's the point of that?

It makes you more cautious.

I thought penalizing for variance was supposed to make you more cautious.

That too. Remember, Risk aversion penalizes both for risk within a given regime and for risk across regimes.

OK, that makes sense. Any other twists I should know about?

Plenty. The most important is that that you can incorporate options and other nonlinear assets without resorting to Monte Carlo regimes.

How do you do that?

By adjusting the risk-adjusted returns in each regime for delta and gamma contributions from options.

That's neat how you've managed to boil all this down to a simple formula.

Like I told you, it's as simple as it can be. But it isn't simple. And unlike standard mean-variance theory there's generally no closed-form solution. Still, computers can calculate numeric answers very quickly.

If your theory's right, it could revolutionize mainstream portfolio analysis.

My theory is right. But it might not change the mainstream at all.

Why not?

I told you. The mainstream seems less interested in managing risk than the appearances of risk. And my theory, by acknowledging possible changes in regime, makes risks more visible.

Surely if someone came along and carefully explained the problems with standard theory and how to fix them they would be welcomed with open arms.

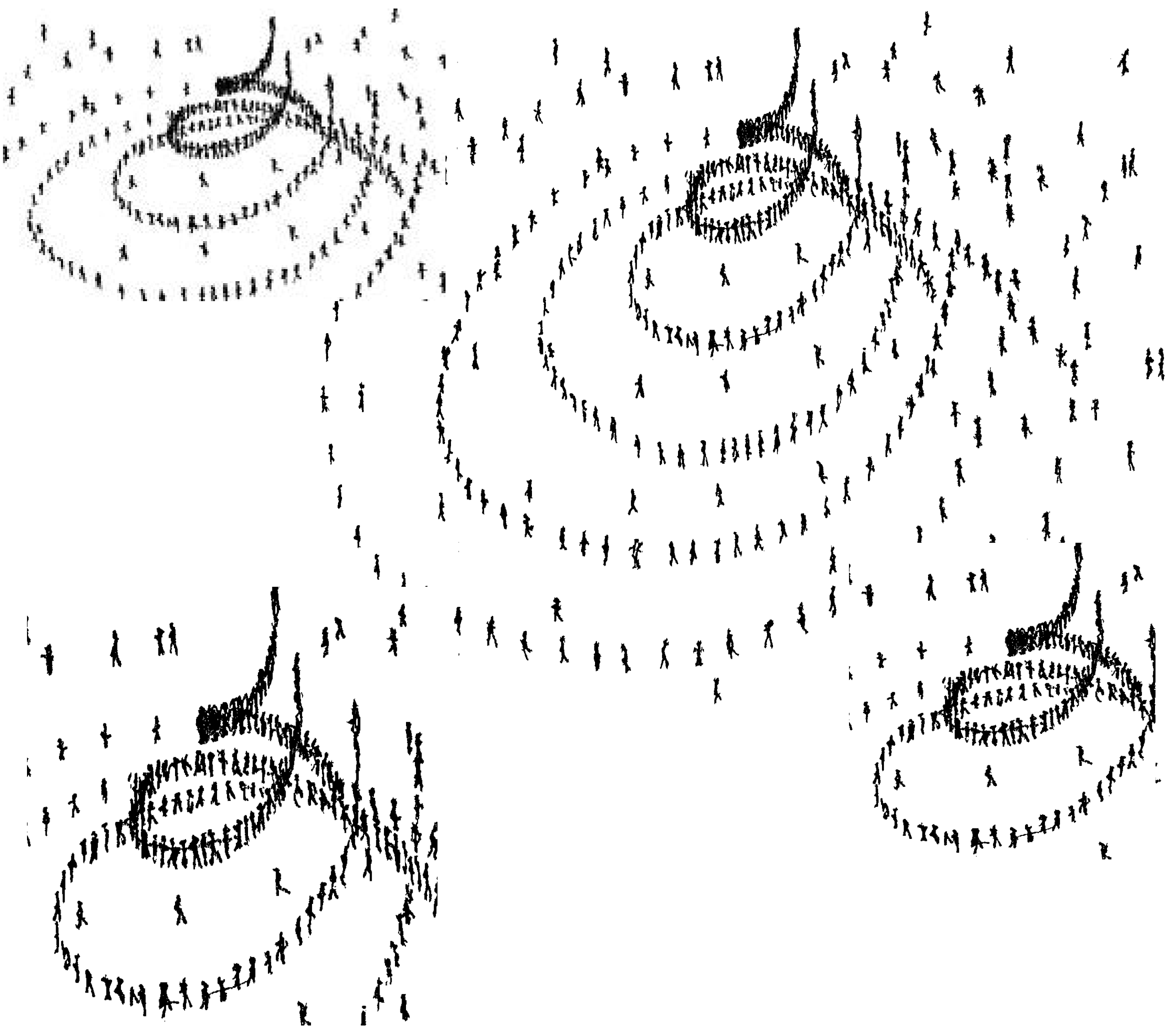
I take it you haven't met Conway or Devlin.

Who?

Never mind. I wouldn't want to spoil the adventure. Welcome to Iceberg Risk.

PART ONE

Standards Of Deviation



“MISSING THE ICEBERG FOR THE ICE CUBES”

Tuesday April 9th 1912

Captain Smith stepped out from his cramped office in the White Star Line Building on Southampton Dock. He was fuming. He had a huge amount of paperwork to complete before tomorrow, and was in no mood for interruptions.

“This had better be important.”

His clerk, who had moments ago reluctantly knocked on the Captain’s door, gestured nervously toward a dubious young man sitting in the reception, smoking a foul smelling cigarette. “I’m terribly sorry, Sir. I’ve been trying all morning to explain to this – gentleman – that you can’t be interrupted. But he insists on seeing you and won’t leave until he has. He seems to think that your ship is at risk.”

“My ship at risk.” Captain Smith slowly repeated this last phrase. “The Titanic. The biggest, fastest, safest ship ever built is at risk. I have no time for such foolishness.”

The clerk smiled smugly and turned to the visitor. “You see. I told you that the Captain would not be interested. Now if you would be so kind as to leave, we can get back to our work in peace.”

The visitor rose to his feet. “Captain, please give me just a few moments to explain. I am not here to criticize your wonderful ship. Our agents’ reports concur that The Titanic may indeed be the finest ship every built. To lose such a vessel would be a tragedy. I am here to prevent precisely that.”

“Your agents?” thundered the captain. “Prevent a tragedy? Are you trying to threaten me young man?”

“No, Captain, nothing of the sort.” The young man thrust out his hand. “Please allow me to introduce myself. My name is Jacques Bachelier and I represent a consortium of shipping insurers. We are willing to offer extremely favorable rates to The Titanic, provided you instigate the required scientific precautions to avoid icebergs.”

“Icebergs! Come now. I have been sailing the North Atlantic for twenty years. In all that time I cannot recall one single incident of a ship hitting an iceberg. Granted, on rare occasions icebergs do stray into the shipping lanes. We understand this smallest of risks, and take the necessary steps to guard against it. The Titanic has six lookouts amongst its crew. If anyone sees an iceberg we will simply sail around it. Satisfied?”

“An excellent system, Captain. Provided you spot the iceberg in time. The Titanic by your own admission is very fast, and its great size will prevent it from making any quick maneuvers.”

“The Titanic is also very tall Monsieur Bachelier. Our lookouts can spot potential problems from a great distance. We shall have plenty of time to take the required action.”

“In the midst of fog? On a moonless night?"

“Do you take me for a simpleton? Under such conditions I can do what sailors have done for centuries: slow down. Now, if you will excuse me Monsieur I …”.

"Certainly you can slow down Captain. But will you? Do you really think the risk of icebergs is such to justify slowing down on every occasion of imperfect visibility? Slow down and the crossing takes longer. Slow down too often and you make less trips per year than White Star Line had anticipated. Think of the cost to your employers, Captain. You cannot eliminate every risk. We understand that perfectly. We simply want you to manage the risks scientifically."

A quick glance at his pocket watch told the Captain he was already running late. But he was intrigued. "Manage the risk scientifically? What precisely do you have in mind?"

"I propose to apply a theory developed by my brother Louis more than ten years ago. It concerns the statistics of interactions between millions of floating particles. Although individual interactions are nearly impossible to predict, most deviations wash out in the aggregate. The residue is random motion that is normally distributed, with a mean drift and variance that rise linearly with time."

Now the Captain had him. He smiled inwardly. A career spanning many years of long sea voyages brings with it one particular luxury. Time to read. And with gambling popular amongst seamen, the Captain had undertaken some study of chance and probability. "Surely you refer to the theory of Brownian motion. Discovered, I believe, by Albert Einstein in 1905. I do not recall your brother's name in that connection."

Jacques Bachelier was at once astonished and indignant. "Let me assure you, Captain, that my brother developed this theory five years before Mr. Einstein, in a doctoral thesis during his time at the Sorbonne. His 'Theorie de Speculation" offers magnificent insight into the random movement of securities prices. But my brother's advisors were blind. They couldn't see the value in his work and it has languished in obscurity. I mean to change that." His voice softened. "But I am most impressed with your knowledge, Captain. Impressed and delighted. You, of all people, must surely appreciate the merits of my new system for forecasting encounters with icebergs. I call it IceMetrics."

The Captain's fascination with mathematics had now firmly got the better of him. "Please step into my office and explain in more detail if you will."

Warming to his subject, Bachelier followed the Captain into the office, continuing to speak as he went. "Basically, IceMetrics estimates the mean frequencies of encounters with large and small lumps of ice in the ocean, the variances in their frequencies, and the pairwise correlations between large and small lumps. Assuming that ice lumps follow multivariate normal distributions, we can calculate the probability of encounters with icebergs."

"Would you not expect these probabilities to vary with location and season?"

"Of course. I propose to sample ice concentrations every hour in the ocean surrounding the ship, and update the mean frequency and correlation estimates using the new information. Granted, encounters with large icebergs are fairly rare. But bear in mind that an iceberg can be viewed as a highly correlated collection of much smaller lumps. So by using information on smaller lumps of ice and their correlation, we can estimate the frequency of very large lumps."

"Very clever, in theory. How accurate have your forecasts been?"

Bachelier reddened a bit and paused to light another cigarette. "In most cases very respectable. Extreme events do tend to occur significantly more frequently than theory predicts. Fortunately, colleagues at the Bureau of International Ship Insurance, or BISI, have devised a practical remedy."

The Captain raised a questioning eyebrow. “Which is?”

“They multiply the IceMetric estimates by three or a bit more.”

“Why three?”

At this, Bachelier had to smile. “Really, I don’t know. I suppose the BISI favored a simple system, in order not to exaggerate its own prowess, so chose to multiply by an integer. At the same time, it did not want to be accused of cutting corners on safety, so it skipped two to arrive at three.”

“Not four?”

“As I explained, we insurers do not wish to avoid all risk, just to manage it. The BISI multiplies stated risks by four only for very poorly run ships. In any case, the patch seems effective. All our insurers using IceMetrics feel quite comfortable underwriting iceberg risk.”

The captain was puzzled. “You call this science, Monsieur Bachelier?”

“Captain, IceMetrics lays a scientific foundation for monitoring iceberg risk. But it does not purport to give exact assessments. Rather, it is a practical tool. On the one hand, it conveys to insurers the order of magnitude of the risks they insure. On the other hand, it helps deflect ex-post criticism from regulators and shareholders should disaster strike. We call the latter the ‘CYD effect’, for “cover your derrière”, and believe me it is very effective. In this new century, no responsible manager should lack a good scientifically-grounded excuse for his mistakes.”

“Somehow,” mused the captain, “I doubt that posterity would forgive me should I let The Titanic hit an iceberg, even if my excuse was firmly grounded in science.”

“Captain, I can understand your skepticism. But allow me the chance to convince you. Let me and my research assistant, Fleur, join you on the voyage tomorrow. We will forecast ice concentrations hourly using the IceMetrics methodology and you can compare our predictions against the ice you actually encounter. If you are satisfied, we would ask you to recommend that White Star Line adopt IceMetrics as their risk management technique, and permit us to advertise that to other prospective customers. If you are not satisfied, then you have no further obligations. In either case, to show our appreciation, our consortium will insure the maiden voyage at our lowest rates.”

“Your offer is very tempting I must admit. But I must also consider my passengers. There are a number of highly influential people on board this maiden voyage, including the Astors no less. I cannot have you taking ice samplings on deck and raising alarm.”

“Captain, I fully endorse your sentiments. We have no desire to draw attention to ourselves. I assure you that my assistant and I will be discreet.”

On board the ship, Jacques and Fleur set about their task with great enthusiasm. Often they perched on the very brow of the great ship to better haul in their ice collection nets. Such apparent derring-do caused quite a stir amongst the other passengers, but was soon explained away as the antics of youthful lovers and they were left in peace.

The hourly sampling proved uneventful. The only tricky part was the updating of forecasts. Originally, IceMetrics forecasts had been based on a simple rolling average of the preceding 20 samples. However, Jacques quickly noticed that ice tended to come in clusters, so that more recent observations deserved extra weight. So he introduced the so-called

Exponentially Weighted Moving Average or EWMA technique, which allowed past observations to fade gradually in influence. On further study, he noticed additional wave-like tendencies in the variance. He incorporated this through a technique he called IGARCH, for Ice-Guarding Auto-Regressive Conditional Heteroskedasticity.

The captain cringed at the terms but was impressed with the results. The hourly forecasts were far more reliable than could be ascribed to blink luck. Granted, they were only four days into the voyage and The Titanic had not yet encountered any patches with high concentrations of ice, but Jacques assured the Captain that the results under average conditions could be extrapolated to the extremes, and even proved this mathematically using multivariate Normal distributions. In none of the samples did the ratio of the actual frequencies of extreme events to the predicted values even approach the BISI-recommended cushion of three.

That evening, Quartermaster Hitchin reported to the Captain from the bridge that a heavy fog had set in. A rather old-fashioned chap, he advised the Captain to slow down for safety. But just as the Captain was about to concur, Jacques appeared with the latest IceMetrics forecast. Ice concentrations had been negligible for the past 24 hours, so both the estimated means and standard deviations were very low.

“Monsieur Bachelier,” said the Captain, “it is time for a practical test. What are the current chances of striking an iceberg big enough to sink The Titanic?”

Jacques was thrilled. “According to my calculations, Captain, a Titanic-sinking iceberg currently represents a ten standard deviation event. For all practical purposes, the probability is zero.”

“Then let us maintain our speed,” said the captain, brushing aside Hitchin’s objections. “From here on, the Titanic will use IceMetrics to gauge its iceberg risks. I believe this decision will be remembered as a historic moment both for us and for the science of risk.”

Within a few hours, the captain’s prophecy would be proved catastrophically right…

Let us not fret too much over poor Jacques Bachelier. His brother’s scientific contributions were indeed rediscovered, and the family name won honor. Indeed, a thrilling though distorted account of Jacques’ relationship with Fleur on that fateful voyage became one of the world’s most popular films. Even Jacque’s failure proved fruitful, for it motivated other innovations in ocean travel that greatly reduced the risk of big things that go bump in the night.

But I worry that the world has not yet grasped the broader lessons of IceMetrics. No financial radar yet penetrates the fog that shrouds the risk of big financial crashes, and distressingly few lifeboats are at hand. Moreover, while managers and regulators of big financial institutions continually monitor their vulnerability to large losses, the methodologies they rely on are distressingly similar to IceMetrics.

Consider, for example, the conventional use of means, standard deviations, and pairwise correlations between assets to predict the risks of high losses on broad asset portfolios. Technically, this is justified only when assets are multivariate normal; that is, when every possible portfolio is normally distributed. As is increasingly recognized, this condition rarely holds. Much effort is spent estimating alternative distributions that can capture the fat tails of individual assets. But hardly any effort is spent working through the implications of these distributions for overall portfolio risks.

This curious omission has several causes:

- On a practical level, many value-at-risk calculations only get used the way a drunkard uses a lamppost: more for support than illumination. When justifying a risk decision that's already essentially been made, it is not worth fitting one's best analytic bulb: normal approximations will do. The Bureau of International Settlements inadvertently encourages this in banking by requiring that the calculated values at risk be multiplied by a factor of three for safety: This means that a truly diligent risk manager will look for legitimate ways to report only a third of the true value at risk.

- Dropping assumptions of multivariate normality opens up a can of worms. Most alternative assumptions are either theoretically unappealing or computationally intractable. So even where analysts concede the unrealism of multivariate normality, they may retain it for practical work. This is especially true in finance, which puts a premium on decisiveness.

- Most people blithely assume that portfolios are bound to suffer less than individual assets from high standard-deviation outliers. That's not true. Granted, according to the Central Limit Theorem, portfolios of many independent, identically distributed assets have approximately normal distributions even when the distributions of individual assets are very fat-tailed. But even a small degree of dependence can render the Central Limit Theorem inapplicable.

Let me be more specific. The single most widely used measure of financial risk is the volatility, or annualized standard deviation. It's an attractive measure because the volatility of a portfolio depends only on the volatilities of the individual assets, their weights, and their correlations. Moreover, we can always approximate the portfolio's risks by a normal distribution having the same mean and the same volatility, just as we can approximate the individual asset's risks by a normal distribution. The relevant question is: how good an approximation is this likely to be for the portfolio, compared with the approximation for the individual assets?

While this question is rarely posed explicitly, it frequently comes up implicitly. Every time the normal distribution is used for portfolio analysis even when the constituent assets are clearly non-normal, the implicit answer is "yes, the approximation is relatively good". Indeed, the implicit answer is "yes" virtually any time an analyst gauges portfolio risks simply by the mean and volatility, for with most other distributions the mean and volatility would not suffice to characterize risks.

In reality, the normal approximation tends to be reasonable for the central region of portfolio risks, where it improves with the number of assets. In the tails, the approximation is much worse. Even in large portfolios, high standard-deviation events tend to occur far more often than a normal distribution suggests. This seems especially true on the downside, which investors fear most. However, tail risks vary enormously from one scenario to another.

I will call this variable excess risk "iceberg risk", because it is mostly hidden from view but threatens major damage. It might also be called "Noah risk", after the proverbial flood that drowned the world. However, "Noah risk" conventionally refers to any extreme risk, whereas I want to focus on the tail risk in portfolios. Portfolios can and generally do have large iceberg risks even when the constituent assets have tail risks that are nearly normal or thinner than normal. Also, in the Biblical story, the world was amply warned about the flood but refused to listen. In contrast, icebergs reflect a type of risk that people look for but may not see.

Iceberg risk has helped drive a huge wedge between portfolio theory and portfolio practice. For example, theory frequently recommends massively long positions in a few assets offset by a few nearly equally massive short positions. Such portfolios brook such obviously massive iceberg risks that no practitioners will consider them, apart from a few heading for the title of “ex-“. Modelers who want to keep their jobs learn to restrain this by imposing fairly tight floors and ceilings on position sizes. Things look much better that way, but privately most modelers will admit to dismay. Ad hoc constraints often end up driving the core result, reducing the theory to a window dressing

In most fields of modern science, such a huge disjuncture between theory and practice would prompt a serious rethinking of theory. Quantum theory, for example, arose from the classical physics’ false prediction that a black box should burn your eyes out with X-rays. Unfortunately, finance theorists have a crutch to fall back on that most other scientists lack. That crutch comes in various fancy carrying cases but basically it amounts to saying that people are stupid. Why don’t markets obey theory? People are stupid. Why don’t portfolio managers accept theory’s advice? People are stupid.

“People are stupid” arguments are hard to refute because, well, people are stupid. There, I’ve fallen into it myself. What I mean is, rather, that the charge of stupidity can itself be rather mind-numbing. Think about it. Suppose a child darts across the street to fetch his ball and gets run over. There are several good candidates for stupidity here: the child, adults who should have been monitoring, and the driver. However, it might be fairer to just call it an accident. The charge of stupidity presumes more information was readily at hand than participants may have realized.

Hence, at risk of appearing foolish, I would rather talk about “ignorance” and “learning” than stupidity. Unlike conventional finance theory, I will not assume that market risks and rewards are clear traffic lights to investors. Rather, I will allow the risks and rewards to occasionally change without giving clear warning of what they’ve changed to or of when and how they may change again. It follows that investors will rarely know exactly what the current risks and rewards are. Instead, investors have to make guesses based on past data and their current theories about the world.

Most readers, I suspect, will find the preceding comments totally plausible. So plausible, perhaps, that some may dismiss them as banal. Or regard them as implicit in conventional theory. Or doubt their potential for application. This book proves otherwise, at least to those willing to read it, and maybe even to those willing to skim it. But if you’re a busy and impatient person like me, perhaps you’re not even willing to invest that much time absent more concrete evidence of why this approach matters.

So let me give two examples. The first concerns the abundant evidence that stock markets “overreact” to news about earnings. More precisely, they overreact if you assume that investors by and large understand the risk parameters of the markets they’re investing in. Some people, most notably Robert Schiller at Yale University, consider this proof that investors are economically irrational, aka stupid. However, if you allow for the possibility that the risk parameters evolve (note that I’m not even requiring the risk parameters to change, but simply that investors think they might change), you can easily account for excess volatility.

The second example is personal. My flirtation with what I’ll call the “theory of ignorance” began back in economics graduate school in the early 1980s, where I wrote my dissertation on the problems of motivating and evaluating economic forecasters. However, I ended up marrying the mistress I knew best–Soviet economic reform–and watched her blossom into post-Soviet reform. I was in love until some stuffed-shirt advisors stole her heart. Having seen no good deed go unpunished, I defected to Wall Street in early 1994 and immediately stumbled into my old flame, looking more radiant than ever.

You see, my new business card gave me recognized expertise in calling post-Soviet debt markets as magically as the Wizard of Oz's diploma had given Scarecrow a brain. Too bad it didn't give me Scarecrow's peace of mind. For example, within a few weeks of joining, I was asked to estimate future default risks for Bulgaria's forthcoming Brady bonds. Despite having worked a lot on Bulgarian economic reform at the IMF and World Bank and knowing some key players, I didn't have much of a clue. I didn't know where to find clues. I didn't even know enough to know that on Wall Street, no analyst should ever say he doesn't know.

It was a rocky start. How I envied my colleagues specializing in Latin America. Compared with the post-Soviet bloc, a country like Mexico had much better data, a much more stable and experienced political regime, and a much longer relevant debt-servicing record. Based on extensive consultations with the IMF, with major investors, and with Mexican government officials, my guru colleagues confidently predicted the Mexican peso would hold to its announced exchange rate band against the US dollar. They were right too. They were right down to the date that they were proved dead wrong.

The Mexican iceberg was a revelation for me. Wow, I thought, some of these gurus are just as ignorant as me. More productively, I realized that the markets I specialized in were bound to move largely on perceptions and changes in perceptions, with only occasional wake-up calls from reality. Ever since, I have focused on better understanding the role of ignorance in financial markets, both from a theoretical and a practical perspective. Call it an odyssey in ignorance.

This book, if you let it, will launch you on that odyssey too. Part One explores the basic mathematics of iceberg risk. Its main intent is destruction: destruction of the presumption that iceberg risk is typically small. It begins with some simple examples, so simple that readers wedded to the conventional view will be tempted to dismiss them as exceptional. I urge sceptics to maintain an open mind. The examples will progressively be elaborated until the supposedly exceptional case becomes the norm and the supposedly normal case the exception.

Fortunately, in the midst of destruction a more robust modelling approach starts to emerge. It is founded on the division of a complex dependence into common and independent parts. The resulting probability distributions are known as mixed (conditional) multivariate normal. Allowing even a limited degree of mixing can make financial risk models far more plausible.

Part Two of the book applies the new approach to portfolio analysis. After reviewing the strengths and weaknesses of various alternatives, I opt for a hybrid system that grafts extra "regime-switching" branches onto the rootstock of the conventional theory. While the extra switches add some math complications, they make the theory much easier to apply. The results are more satisfying as well, since they tend to avoid implausibly extreme recommendations.

However, if you think that the new approach justifies all the conventional rules of thumb, guess again. Convention has turned risk management into a lumbering elephant. The new approach says the elephant needs to dance. To put it more prosaically, funds need dynamic "risk overlays", the risk counterpart to the "currency overlays" increasingly in vogue. These overlays will vary not only with managers' reading of an ever-changing environment, but also with their attitudes to risk.

But, before setting sail, let me warn you about the ride ahead. I don't present either a polished academic treatise or a comprehensive "how-to" guide. Instead I introduce a new way of thinking about portfolio risk. Actually, I should say “unfamiliar” rather than “new”, because none of the elements is new, just the way they are put together, and even the latter may be disputed.* Still, this is a daunting task, because one of the hallmarks of conventional thinking is a dismissal of criticisms as either uninteresting or impractical.

Thus, if I plunge into a formal mathematical treatment of iceberg risk, few potential readers will follow me in. If I downplay the math, the approach will look hokey. If I harp on criticisms of the status quo, I will be begged for constructive suggestions. If I focus on remedies, I will be challenged to prove there is a serious disease.

This recalls an old joke: How many preachers does it take to change a light bulb? The answer is: Just one, but the light bulb has to really want to change. Unfortunately, most financial risk analysts would bristle at the suggestion their light bulbs are burning dim. Either they have grown accustomed to working in the gloom or reject other light bulbs as offering too little improvement for the price.

I was in despair about how to proceed until I chanced into two old friends who worked in risk management for a major investment bank. As they told me about the practical problems they were encountering and the remedies they were developing, I saw that their intellectual evolution paralleled my own. Yet their adventures and lively debates were far more entertaining than my dry theory. One day it dawned on me that readers would likely feel the same. My friends kindly gave me permission to recount their experiences and to use them as a foil for theoretical reflection—provided, of course, that I camouflage their names and the business-sensitive details. So, without further ado, let me introduce you to my friends and the chance incident that first shook their faith in chance.

---

*Let me use this one and only footnote in the book to lament its most shameful feature, namely the dearth of footnotes citing others’ work. There are two main reasons for this. First, reflecting my own distance from academic finance, I am largely unfamiliar with who said what first when. Second, most of the readers I want to reach don’t care. Still, I regret not giving others toiling in the same vineyard more credit for the wine. To help make amends, the book closes with suggestions for further reading. Between those readings and the books and articles they cite you can plug into a rapidly expanding field at whatever mathematical level you prefer.

1. Odd Odds of Odds

Conway, head of risk management for Megabucks Investment Bank, shuffled again through the auditors' report. On the bright side, the auditors praised the new monitoring system he had introduced. For every asset in its portfolio, Megabucks now had up-to-date information not only on the mean and standard deviation but also on the skewness and kurtosis. Conway's team also meticulously tracked the correlations between assets. But the auditors went on to fault the team for sloppy implementation at a portfolio level. "In one particularly egregious example," wrote the auditors with thinly disguised contempt, "an analyst argued that three completely uncorrelated assets bore absolutely no risk of plunging together."

The report did not name the analyst and didn't need to. Only Devlin was reckless enough to argue openly with the auditors. Devlin was the quintessential Devil's Advocate. At first this had put Conway off, as it did most people. Over time he had come to appreciate Devlin's challenges. They kept him on his toes. And more often than not Devlin turned out to be right.

But this time Devlin has clearly gone too far. Why defend the indefensible? Conway shook his head and sighed. Two years previously, when the Asian crisis lured the two of them into risk management, Conway figured they would rise or fall together. Instead, only Conway had thrived. Devlin was too much of a perfectionist in a business that craved quick results. Indeed, Devlin only kept his job thanks to Conway's intervention. *Now they'll be gunning for him again. And maybe for anyone who harbors him. I better look into this.*

Conway found Devlin hunched over his desk, looking even more frazzled than usual. Peering over his shoulder, Conway watched him pull coins out of a bowl on his desk, rearrange them into groups of three, scribble furiously in his notebook, and then toss the coins back in the bowl to start all over again. After a few minutes of this spectacle, Conway spoke up. "I certainly admire your conversion to the more practical aspects of risk management, but I hardly think those coins represent much of a threat to the profit margins of Megabucks."

A startled Devlin spun around. "Oh, it's you, Conway. Ha-ha. Very funny. But I am being practical. I was just running some simulations on the options book."

"With three coins?"

"Granted, it's a bit crude. But I like to start with the basics. It sharpens my intuition."

"Imagine that. And here was I idly worrying that my old sparring partner had blown a fuse. You need to lighten up, my friend. You're beginning to remind me of the young psychiatrist in the joke."

"What joke?" asked Devlin. He smiled in spite of himself. Conway was a charmer.

"An old and a young psychiatrist worked in the same building. Every day after work the old psychiatrist would come by, bursting with energy, and propose some interesting activity: tennis, jogging, attending a concert, etc. But the young psychiatrist was exhausted and kept turning the old man down. After a few weeks, the young psychiatrist couldn't stand it any longer. 'I just don't understand how you can do it. All day long people tell me their problems. Unbelievable problems. It just wears me out. You're thirty years older and must hear the same kind of problems that I hear. And yet you're full of energy. What's your secret?' he asked. 'I don't listen,' said the old psychiatrist.

Devlin laughed, but at the same time he felt an all too familiar anxiety welling up within him. Conway cares more about having a neat system for labelling risk than about the accuracy of the labels, he thought. If it's talked and written about often enough, then Conway thinks it must be right. Mr. Conventional Wisdom. Faced with a problem, Conway's first inclination—and usually his second and third—was to follow the accepted authorities.

Nevertheless, Devlin had come to respect Conway for his pragmatism. Conway did not idolize conventional wisdom; he simply considered it far more likely than not to be approximately correct. If Devlin or someone else could demonstrate otherwise, Conway would change his mind. Those qualities were the secret of Conway's success in risk management.

Devlin, on the other hand, was born to question and doubt. The more he studied stochastic financial variables, the more questions and doubts he had. And the more Devlin doubted, the more depressed and frazzled he became. At times he wished he had entered a field with more scientific certainty, say, the study of the origins of the universe.

"Eh, Devlin. Hello."

"Sorry. I was daydreaming. Come on, Conway, tell me what you really want to talk with me about." Devlin smelled trouble. Devlin had a good nose for trouble once he was in it. Too bad he rarely smelled it sooner.

"The auditors wrote you up for unreasonable valuations."

"Those bozos!" exclaimed Devlin. "I tried to set them straight but they wouldn't listen."

If that's the way you talked to them then no wonder. "Why don't you tell me what happened?"

"Do you remember the problems I was having last month in valuing the aggregate risk exposure on three different binary options? You told me to use some common sense and for once I followed your advice."

"Wonders will never cease. Now if I recall correctly, you valued the payoffs, leaving aside fixed fees, at plus or minus \$10 million each, with 50-50 odds of success, and zero correlations.

"You have a good memory."

"How could I forget? The options desk was screaming to be allowed to keep them, despite a strict injunction not to risk aggregate losses of more than \$20 million. Frankly, I couldn't understand the fuss. The matter seemed cut and dry."

"That's how I felt too, so I let the desk keep its positions." "You what?" cried Conway. This was far worse than he feared. Devlin had gone beyond spouting gibberish. He had acted on it. "I'm sorry, Devlin, you need help, but I've got a business to run and now some damage to limit. Go home while I sort things out and try to figure out what to do with you."

"Conway, hold your horses. Just hear me out. I'm not crazy. Not yet anyway"

"You could have fooled me. Now listen, how much simpler can a problem be? You're the one flipping coins, so you tell me: If heads and tails are equally likely, what are the odds that three tails will appear in a row?"

"That depends. If the coins are independent, then $\frac{1}{8}$. That's what most people assume."

"Of course. What else is there? If the probability of two tails is ¼ and a third uncorrelated coin has a probability ½ of tails, the probability of three tails has to be 1/8."

"That's not the only possibility. For example, suppose the coins come in two equally prevalent types, with one weighted to always come up heads and other weighted to always come up tails. If you choose one coin randomly but then toss it three times, the odds of three tails will be ½."

"Come on Devlin, I don't have time for this. This isn't a schoolyard experiment." Conway was determined not to get sucked into another of the great Devlin Theoretical Debates. But then a thought sprung up. "In any case, that's irrelevant. In your last example, the three tosses would be perfectly correlated. The options we're dealing with were uncorrelated. So we're back to $\frac{1}{8}$."

Devlin allowed himself a brief smile. He knew he had Conway now. "Slow down, slow down. While independent assets are never correlated, uncorrelated assets can be dependent. Granted, with only two Bernoulli assets, one definition is equivalent to the other: namely, that the probability of two tails is the square of the probability of one tail. But that is a special case. Besides, correlation applies only to pairs of assets, not larger groups."

"What about the three options we discussed? Each was pairwise uncorrelated with every other. That makes three pairs of uncorrelated assets, versus only one pair in the two asset case. Doesn't that amount to independence?"

"No, it does not. Let me give you an example." Devlin fumbled excitedly on his litter-strewn desk for the pad he had had earlier. In the process he tipped an old half-eaten sandwich off the desk and into the waste bin below. *Now was that random chance or a calculated move?* "Suppose we start by tossing three independent, unbiased coins but add the following condition: Every time an odd number of tails appears, all three outcomes are disqualified and the coins are tossed again. In that case, the only feasible orderings are HHH, TTH, THT, and HTT, each of them equally likely. You can readily check that the odds of heads and tails are equal for every coin, and that, for any given pair of coins, the probability of two tails equals ¼."

"Why doesn't the 'no odd tails in triples' restriction affect the correlation?"

"Suppose the third coin is heads. Then the first two coins must either both turn up tails or both turn up heads. Conversely, if the third coin is tails, then one of the first two coins must be heads and the other tails. Hence, the pairwise correlation is either +1 or –1 conditional on the outcome of the third coin, but since the third coin is equally likely to be heads or tails, the overall or unconditional correlation is zero. Yes?" Devlin looked up expectantly.

Conway felt dizzy. He glanced at his watch and hesitated. He had a management meeting in a couple of minutes. With a sigh he slumped down at Devlin's desk and began to play with the coins. After a few minutes, he looked up. "Remarkable," he said. "I wouldn't have thought this could be. How many possible unbiased, uncorrelated solutions are there?"

"Too many to count. Note that, by symmetry, 'no odd heads in triples' must also be a solution. Now suppose that at every triple toss, one of the restrictions 'no odd heads' or 'no odd tails' is randomly imposed with probabilities l and $1-l$ respectively. No matter which restriction is imposed, we have an uncorrelated fair solution. Hence each randomisation must also represent an uncorrelated fair solution, with probability $\frac{l}{4}$ of three tails."

Conway took a deep breath. “OK, I will concede your theoretical point. But the options traders are playing with serious money, not petty coins. I still don’t see why, out of all the possible solutions, you assumed the least possible tail risk for their trades. It is quite implausible. Why not be conservative and assume the most tail risk, or choose the average which corresponds to independence?”

“Theory told me I needed more information. So I went to talk with the traders. It turns out that the options applied to three companies vying to establish a new industry standard for computer networking. Four choices for a new standard were considered equally likely: the three different companies’ in-house solutions, or continuation of the no-standard status quo. Each company was expected to thrive unless a rival’s solution was adopted. You can check that this corresponds exactly to the ‘no odd tails’ scenario we discussed.

Conway’s jaw dropped open and his eyes stared at the coins on Devlin’s desk…

While Conway regains his composure, let me clarify a few of the terms that he and Devlin used, for they are important building blocks for what follows. In the process I will also list a few useful formulas. This is fairly elementary stuff, so feel free to skim through it. The only material likely to be unfamiliar concerns higher-order cross-moments.

Univariate Random Variables

When we identify the return on an investment with a toss of a coin, we are abstracting from every feature other than risk. Mathematically, this is formalized in the notion of a stochastic or random variable X, which can take on various outcomes k from a set K. X is univariate when each outcome k is associated with a single real number x_k.

A scalar function h of a univariate random variable X is also a univariate random variable, taking the values $\{h(x_k)\}$ for the various k in K.

Bernoulli Examples

In our coin tossing examples, X corresponds to an individual coin and K to the outcomes heads and tails. We will generally assign the numbers 1 to heads and 0 to tails, although this can be altered without loss of generality.

Any random variable that takes outcomes of only 0 or 1 is known as Bernoulli, in honor of one of the fathers of probability theory. With Bernoulli variables other than coins, the outcome one is generally labeled “success” and the outcome zero “failure”.

Probability Measures

Every subset of K defines a possible event. An event is said to occur if and only if one of the outcomes in the associated subset occurs, so we can use the same label for both. The probability $Pr\{L\})$ of any event L denotes the limiting relative frequency or “likelihood” of L in unbiased samples, assuming that limit exists. However, this definition is not adequate, since bias is typically defined in terms of probabilities. Formally, probability is defined as a measure of the subset relative to the whole set, which measure in turn is associated with a limiting likelihood. By definition, all probabilities range between 0 and 1 with $Pr\{K\}=1$. Probabilities are also additive in the following sense. For any countable union of mutually exclusive events (ie, no outcome is associated with more than one of the events), its probability equals the sum of the constituent probabilities.

Probability Point Distributions

For a set K of discrete outcomes, the probability distribution is a function p associating each value x with some probability $p(x) \equiv Pr\{x\}$. From the definition of probability, each $p(x)$ must be nonnegative and the sum $\sum_{k \in K} p(x_k)$ over all k in K must equal 1. If $p(x_k)=1$ for some x_j, then the outcome is certain and the distribution is said to be degenerate. The next simplest probability distribution pertains to a Bernoulli distribution, which is completely characterized by $p(1)$, since the only other probability $p(0)$ must equal $1\text{-}p(1)$.

The probability $Pr\{X \leq x\}$ is called the cumulative probability at x and is generally denoted $F(x)$. Viewed as a function of X, F is non-negative, non-decreasing, and takes a maximum value of 1. Given such a function, the point probabilities $p(x)$ can be calculated as $p(x)=F(x)\text{-}max_{y<x}F(y)$ when K is discrete.

Support of a Distribution

A probability distribution may attach zero weight to some outcomes in K. The subset of K that carries positive weight is called the support of the distribution. A distribution with support on n distinct outcomes is called an n-point distribution.

Tail Risk

Tail risk refers to the probability of unusually low outcomes. It is defined in terms of a threshold T and the cumulative probability $F(T)$- that X lies in the tail. When interested in the probability $Pr\{X{}^3T\}$ of unusually high outcomes, I shall speak of the upper tail risk.

Multivariate Random Variables

The notions of probability can be easily extended to variables taking on multivariate outcomes. Each outcome is associated with a point k in multidimensional space and the probability measure is again taken over the set K of all outcomes. For discrete outcomes, the joint probability distribution is given by a function p associating each outcome $(x_{1k},x_{2k},\ldots,x_{nk})$ with a nonnegative number $p(x_{1k},x_{2k},\ldots,x_{nk})$, such that $\sum_{k \in K} p(x_{1k}, x_{2k}, \ldots, x_{nk}) = 1.$

Conditional Probabilities

Conditional probabilities measure the probability $Pr\{L_1|L_2\}$ of one event L_1 given that a second "conditioning" event L_2 has occurred. This is defined as the probability that both events occur divided by the probability that the conditioning event occurs:

$$\Pr(L_1 \mid L_2) \equiv \frac{\Pr(L_1 \cap L_2)}{\Pr(L_2)}.$$

In many applications, the probability of the second event is not readily at hand but must be calculated from data on the joint distribution. For example, let X and Y be two random variables characterized by a joint probability $p(x,y)$. Then the probability distribution of Y alone can be calculated as $p_Y(y) \equiv \sum_{k \in K} p(x_k, y)$. Hence

$$p(x \mid y) = \frac{p(x,y)}{p_Y(y)} = \frac{p(x,y)}{\sum_{k\in K} p(x_k, y)}.$$

Conversely, given the probability distribution of *Y* and the conditional probability distribution of *X* given *Y*, we can calculate the joint distribution as $p(x,y) = p(x \mid y)\, p_Y(y)$.

Mixtures of Distributions

The preceding formula can also be interpreted as a way to concoct new probability distributions out of old ones. Namely, any weighted average $\sum_i l_i p_i(X)$ of a set of probability distributions $\{p_i(X)\}$ is also a probability distribution, provided the weights $\{l_i\}$ are non-negative and sum to 1. To verify this, just associate each subscript *i* with a value of *Y*, so that $l_i \equiv l(y)$ and $p_i(X) \equiv p(X|y)$.

A probability distribution formed in this way is called a mixture of distributions. Note that any expectation of a mixture equals the mixture of its expectations. Mixtures are also called convex combinations (the formal name for non-negative weighted averages summing to one) or randomizations. The probability distribution that determines the weights for the randomization is called the mixing distribution.

Independence

Two random variables are perfectly dependent if one can be expressed as the deterministic function of the other. At the other extreme, two random variables are said to be independent if and only if the outcome of one has no bearing on the outcome of the other. In other words, the conditional probability $Pr\{x|y\}$ must be a function of *x* alone. Recalling the definition of conditional probability, this means that *X* and *Y* are independent if and only if, for all outcomes *x* and *y*,

$$p(x,y) = p_X(x) \cdot p_Y(y)$$

Hence, if *X* and *Y* are independent, their joint probability is the product of a factor that depends only on *X* and a factor that depends only on *Y*. The converse is readily shown as well. In other words, independence is mathematically equivalent to multiplicative separability of the joint probability measures.

Expectations

The expectation *E* of a random variable X denotes the weighted average of the various values that *X* can take, with weights proportional to the probabilities. The formula for a discrete set of outcomes is:

$$E[X] = \sum_{k\in K} x_k\, p(x_k).$$

Expectations need not be finite. Suppose that $p(k) = \frac{1}{k} - \frac{1}{k+1} = \frac{1}{k(k+1)}$ for all positive integers *k*. This is a legitimate probability measure, as it is nonnegative and sums to one. However, the expectation is $\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$, which is unbounded. When the expectation is infinite, the average from any sample—also known as the "sample mean"—remains finite, but tends to grow without bound as the sample increases in size.

Expectations of Functions

Since a function *h* of a random variable *X* is also a random variable, we can calculate its expectation $E[h(X)]$ as $\sum_{k\in K} h(x_k)\,p(x_k)$. For example, the affine transformation $aX+b$, where *a* and *b* are constants, has the expectation $aE[X]+b$, which is the affine transformation of the expectation.

Likewise, we calculate a conditional expectation $E[X|Y]$ as $\sum_{j\in J} x_j\,p(x_j \mid Y)$. The expectation $E[X|Y]$ is in turn a function of *Y* so we can take its expectation with respect to the probability distribution of *Y.* It is readily demonstrated that $E_Y[E[X|Y]]=E[X]$. In other words, the expectation of the conditional expectation equals the unconditional expectation.

Expectations are also readily applied to mixtures of distributions. To calculate the probability of every outcome under a mixture of distributions, find the outcome's probability under each individual distribution and take the expectation over the mixture: $p_X(x)=E_Y[p(x|y)]$. Indeed, the expectation of any function over a mixture of distributions equals the mixture (more precisely, the expectation over the mixture) of the individual expectations.

Bias and Fairness

Devlin and Conway used the word "unbiased" to indicate a coin that is equally likely to come up heads or tails. By extension, they will use the word "unbiased" to describe any Bernoulli random variable with 50% probability of success.

Technically, this usage is sloppy. Bias is not a purely objective property. Rather, it refers to the difference between the supposed and actual values. If a Bernoulli variable is expected to never fail and never does so, there is nothing inherently biased about it.

In gambling and game theory, fairness is interpreted more properly as equivalence between expected returns and entrance fees. That is, after including entrance fees and conditioning on current information, a fair game's expectation should be zero. In theoretical finance, this concept of fairness is enshrined in the concept of a "martingale".

Unprepared to reject either conventional conversation or finance theory, I ask the reader to tolerate the dual usage. Fairness will mean a 50% probability of success in Bernoulli games and a zero expectation in other games.

First Moment

For *n* a positive integer and *X* a univariate random variable, the expectation $E[X^n]$ is known as the n^{th} moment of *X*. The first moment $E[X^1]$ is the mean or expected value $E[X]$ itself, and is often denoted by m. For a Bernoulli random variable, the mean m equals the probability of success. Indeed, every moment of a Bernoulli variable equals m.

The mean of a weighted sum of random variables equals the weighted sum of their means. That is, if $\{a_i\}$ are constants:

$$E\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i E\left[X_i\right] \quad .$$

Second Moment

The second moment $E[X^2]$ adds information about the dispersion of X around its mean, so it is often reported in the form of a so-called variance

$$Var[X] \equiv E\left[(X-m)^2\right] = E\left[X^2\right] - m^2 , \quad \text{where } m \equiv E[X] .$$

The variance is always positive except for a degenerate distribution, where it is zero. For Bernoulli variables, the variance equals $m(1-m)$.

The square root of the variance is known as the standard deviation $Std[X]$ and has the same units as X, which facilitates comparison. It is often denoted by s or s_X, which makes the variance s^2 or s_X^2 .

Adding a constant to a random variable does not affect its variance. Multiplying a random variable by a constant multiplies the variance by the constant squared. Hence, when a and b are constants, $Var[aX+b]=a^2Var[X]$ and $Std[aX+b]=a\times Std[X]$.

Higher-Order Moments

To highlight the marginal information value, higher-order moments of X are generally expressed in terms of the standardized transformation $\frac{X-E[X]}{Std[X]} \equiv \frac{X-m}{s}$, which has a zero mean and unit standard deviation.

The third standardized moment $E\left[\left(\frac{X-m}{s}\right)^3\right]$ is called skewness. Skewness, like all standardized odd-numbered higher moments, will be zero for distributions that are symmetric around their means. The skewness will tend to be negative if the distribution tilts toward the left of the mean, and positive if the distribution tilts toward the right of the mean.

The fourth standardized moment $E\left[\left(\frac{X-m}{s}\right)^4\right]$ less three is called kurtosis. Kurtosis measures the relative fatness of the outer values or tails. The minimum kurtosis is -2, which occurs when X has a two-point distribution with equal weights. Kurtosis is close to zero for the average of a large number of independent, identically distributed random variables. Distributions with distinctly positive kurtosis are called leptokurtotic or fat-tailed.

Multivariate Moments

Multivariate moments take the form $E\left[X_1^{k_1} X_2^{k_2} \cdots X_n^{k_n}\right]$ where the various exponents k_i are nonnegative integers and at least one is positive. The sum $\sum_{i=1}^{n} k_i$ of the exponents is called the order of the moment. For example, $E[X_1X_2X_3]$, $E[X_1^2X_2]$, and $E[X_3^3]$ are all third-order moments. If at least two exponents are positive, multivariate moments are called cross moments. The rest are essentially univariate and are called own moments.

Cross moments arise naturally in the calculation of moments of weighted sums of random variables. Each additive term of $E\left[\left(\sum_{i=1}^{n} a_i X_i\right)^J\right]$ consists of own moments or cross moments of order *J*.

Covariance

A second-order cross moment takes the form *E*[*XY*]. As with second-order own moments, measuring the variables as deviations from the mean helps to distil the marginal information. The resulting quantity is called the covariance *Cov*[*X*,*Y*] and can be expressed in various forms. Denoting the means of *X* and *Y* by m_X and m_Y respectively,

$$\begin{aligned} Cov[X,Y] &= E[(X - E[X])(Y - E[Y])] = E[X(Y - E[Y])] \\ &= E[(X - E[X])Y] = E[XY] - E[X]E[Y] \quad . \end{aligned}$$

Adding a constant to either of two random variables leaves the covariance unchanged, while multiplying either variable by a constant multiplies the covariance in the same proportion. Hence, for constants *a*, *b*, *c*, and *d*:

$$Cov[aX + b, cY + d] = ac\,Cov[X,Y] \quad .$$

The conventional shorthand for *Cov*[*X*,*Y*] is s_{XY}. Since *Cov*[*X*,*X*] equals *Var*[*X*], s_{XX} is also written s_X^2.

Correlation

An even more distilled variant of a second-order cross moment is the correlation, sometimes known as the partial correlation or pairwise correlation. Correlation is the covariance between standardized random variables, and can be expressed as:

$$Cor[X,Y] = E\left[\left(\frac{X - m_X}{s_X}\right)\left(\frac{Y - m_Y}{s_Y}\right)\right] = \frac{Cov[X,Y]}{s_X s_Y} = \frac{E[XY] - m_X m_Y}{s_X s_Y}$$

Correlation is unitless. Adding a constant to one of the variables doesn't affect it. Multiplying one of the variables by a constant multiplies it only by the sign of the constant:

$$Cor[aX + b, cY + d] = \operatorname{sgn}(a) \cdot \operatorname{sgn}(c) \cdot Cor[X,Y] \quad .$$

The conventional shorthand for *Cor*[*X*,*Y*] is r_{XY}. Like covariance, correlation is symmetric in its arguments, so $r_{XY}=r_{YX}$. Note that the correlation r_{XX} of a variable with itself equals one.

Two variables with zero correlation are said to be uncorrelated. All independent variables are uncorrelated, but, as Devlin demonstrated, variables can be uncorrelated without being independent. To give another example, if *X* is distributed symmetrically around the origin and *Y* equals X^2, then despite their perfect dependence *X* and *Y* will be uncorrelated.

The Main Use of Covariance and Correlation

By far the most common use of covariance and correlation involves the calculation of the variance of a sum of random variables. The variance of the sum equals the sum of all the individual variances and the covariances between every pair of distinct random variables. The formula is readily extended to incorporate constant-weighted sums:

$$Var\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i^2 \boldsymbol{s}_i^{\,2} + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} a_i a_j \boldsymbol{s}_{ij}$$

where s_i^2 and s_{ij} denote *Var*[X_i] and *Cov*[X_i,X_j] respectively. Denoting *Cor*[X_i,X_j] by r_{ij} and taking account of the symmetry r_{ij}=r_{ji}, this can be rewritten as:

$$Var\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i \boldsymbol{s}_i^{\,2} + 2\sum_{i=1}^{n} \sum_{j=1}^{i-1} \boldsymbol{r}_{ij} a_i a_j \boldsymbol{s}_i \boldsymbol{s}_j \quad .$$

Correlation and the Variance of Averages

One particularly interesting weighted sum is the simple average of *n* variables having a common variance s^2 and a common correlation r. In that case the previous formula simplifies to:

$$Var\left[\sum_{i=1}^{n} \frac{X_i}{n}\right] = \left(\frac{1}{n} + r\,\frac{n-1}{n}\right)\boldsymbol{s}^2 = \left(\frac{1-r}{n} + r\right)\boldsymbol{s}^2 \quad .$$

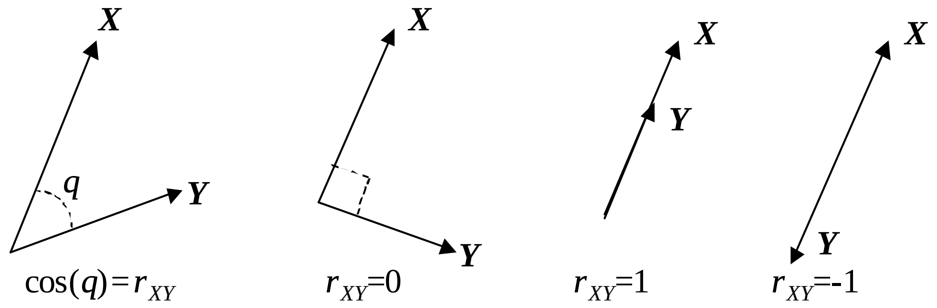
In other words, the variance of the average equals the common individual variance times $\frac{1-\boldsymbol{r}}{n} + \boldsymbol{r}$. The latter approaches r as *n* grows large, so diversifying across positively correlated variables can reduce risks only to a limited extent.

This same formula applies if the outcomes of the variables are identical with probability r and independent with probability 1-r. Thus, in at least some applications, correlation can be viewed as reflecting a mixture of independence and perfect dependence. However, correlation is a broader concept, and unlike a probability mixture it allows r to be negative.

Limits on Correlation

If each variable *X* and *Y* is treated as a vector, then r_{XY} indicates the cosine of the angle between the two vectors. Hence the correlation is always a number between –1 and 1. A value of +1 is called perfect positive correlation and implies an affine relationship *Y=aX+b* for some positive constant *a*. A value of -1 is called perfect negative correlation and implies an affine relationship with *a* negative.

THE GEOMETRY OF CORRELATION



The pairwise correlations of a set of variables are also restricted by the need for internal consistency between the angles of the corresponding vectors. For example, suppose *X* points exactly opposite to both *Y* and *Z*, so that it's perfectly negatively correlated with either. Then *Y* and *Z* must point in the same direction and be perfectly positively correlated with each other.

These restrictions on correlation have a straightforward statistical interpretation: namely that weighted sums of random variables can never have negative variance. For example, when *n* variables have a common correlation, that correlation can never be less than $-\dfrac{1}{n-1}$.

Conditional Covariances

The conditional covariance of *X* and *Y* conditional on a common event *Z* can be written as:

$$Cov[X,Y|Z] = E[XY|Z] - E[X|Z] \cdot E[Y|Z] \ .$$

By taking the expectations of each side with respect to *Z* and undertaking some additional manipulation, we can establish that:

$$Cov[X,Y] = Cov_Z\left[E[X\mid Z], E[Y\mid Z]\right] + E_Z\left[Cov[X,Y\mid Z]\right].$$

In other words, the unconditional covariance equals the covariance of the conditional means plus the expectation of the conditional covariances. In the special case *Y*=*X*, the unconditional variance equals the variance of the conditional mean plus the expectation of the conditional variance:

$$Var[X] = Var_Z\left[E[X\mid Z]\right] + E_Z\left[Var[X\mid Z]\right].$$

These formulas are readily applied to mixtures by letting Z denote the mixing distribution. For example, consider the "no odd tails" and "no odd heads" scenarios that Devlin discussed. Each scenario has mean 1, variance ¼ and covariance 0. The formulas above imply that any mixture matches these moments and therefore fills all the requirements of a solution.

Higher-Order Cross Moments

As a first approximation, the average correlation among a set of standardized random variables measures the average tendency to move in (linear) synchrony. However, this average may be formed in various ways. One set of random variables may alternate between independent and identical movements. Another set of random variables may react more uniformly. Higher-order cross moments help describe the fluctuations in synchrony.

Third-order standardized cross moments are known as co-skewnesses. They take two generic forms: $E\left[\left(\frac{X-m_X}{s_X}\right)\left(\frac{Y-m_Y}{s_Y}\right)\left(\frac{Z-m_Z}{s_Z}\right)\right]$ and $E\left[\left(\frac{X-m_X}{s_X}\right)^2\left(\frac{Y-m_Y}{s_Y}\right)\right]$.

Fourth-order standardized cross moments are known as co-kurtoses. They take four generic forms: $E\left[\left(\frac{X-m_X}{s_X}\right)\left(\frac{Y-m_Y}{s_Y}\right)\left(\frac{Z-m_Z}{s_Z}\right)\left(\frac{W-m_W}{s_W}\right)\right]$, $E\left[\left(\frac{X-m_X}{s_X}\right)^2\left(\frac{Y-m_Y}{s_Y}\right)\left(\frac{Z-m_Z}{s_Z}\right)\right]$, $E\left[\left(\frac{X-m_X}{s_X}\right)^2\left(\frac{Y-m_Y}{s_Y}\right)^2\right]$ and $E\left[\left(\frac{X-m_X}{s_X}\right)^3\left(\frac{Y-m_Y}{s_Y}\right)\right]$.