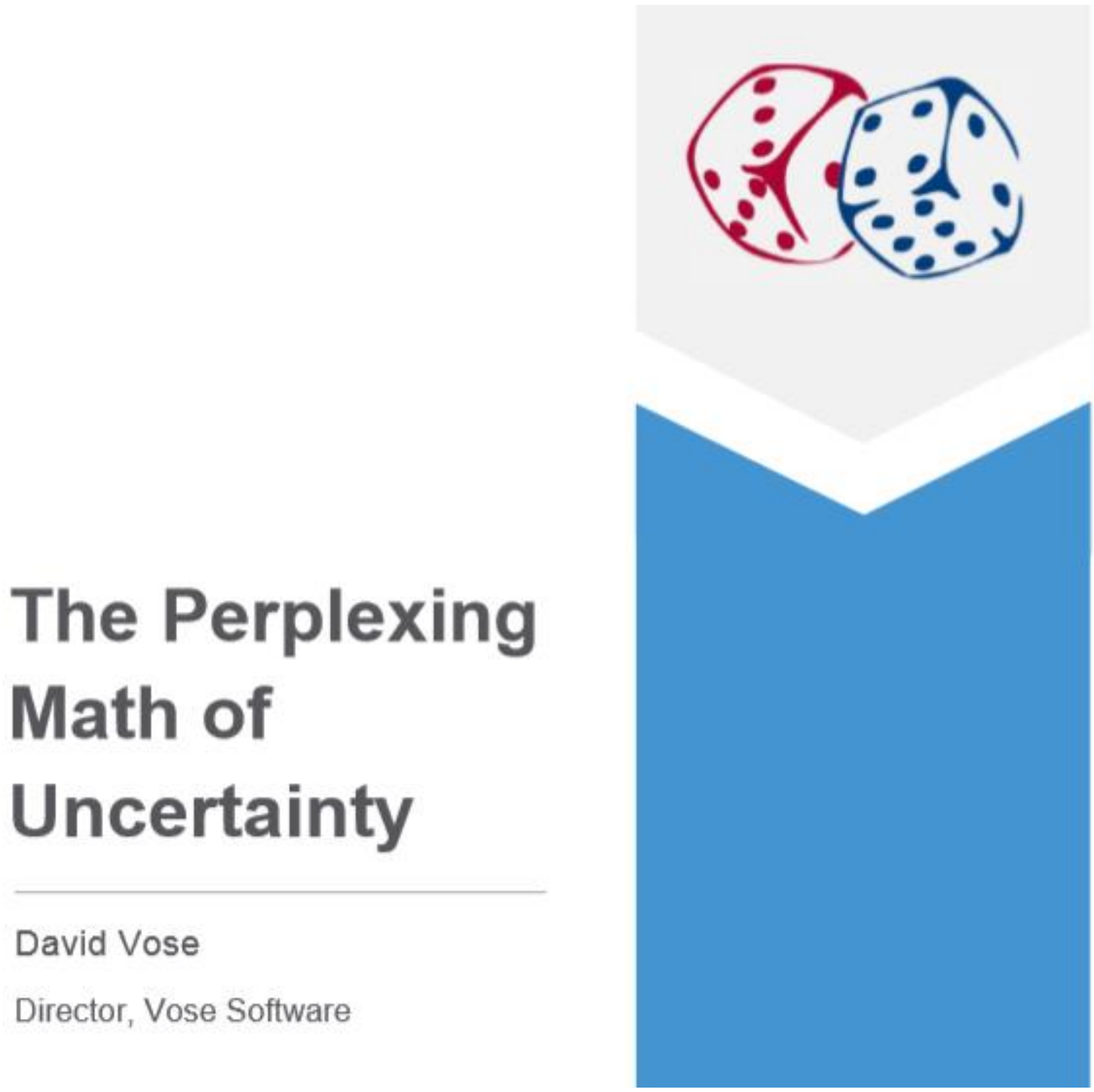


Озадачивающая математика неопределенности

Дэвид Восе

Перевод Алексея Белкова

ВВЕДЕНИЕ

Количественный анализ рисков требует от риск менеджера широкого спектра навыков. Одним из наиболее недооцененных навыков является умение работать с неопределенными переменным в имитационных моделях Монте-Карло, что является основой стандартной техники анализа рисков.

Такие простейшие действия как сложение, вычитание, умножение и деление мы научились выполнять с простыми числами еще в начальной школе, они для нас само-собой разумеющиеся и не вызывают вопросов. Трудно представить какую-либо финансовую модель, где хотя бы одно из этих действий не присутствовало. Тем не менее все вышеперечисленные действия не работают с распределениями так, как они работают с числами. Тревожно, что большинство риск менеджеров не знает об этих закономерностях и/или неправильно выполняют манипуляции с неопределенными переменными в ходе моделирования в финансовых моделях, возможно, потому, что подходят к этому вопросу с точки зрения привычных вычислений с простыми числами.

Большинство финансовых моделей строится на базе Excel с использованием надстроек, позволяющих применять метод Монте-Карло, точно также устроен и Vose ModelRisk.

Бытует мнение, что можно взять исходную финансовую модель (например, модель Cash Flow с расчетом EBITDA или NPV) и просто заменить любое неопределенное значение, выраженное детерминистически (одним числом), на функцию, генерящуюся случайным образом из заданного распределения, чтобы отразить неопределенность вышеупомянутого значения. Именно здесь и возникает ошибочное мнение, что на этом работа по заданию неопределенности завершается и всю остальную логику модели можно оставить без изменений.

А это на самом деле имеет значение и оказывает определенный эффект. Некорректные (*частично правильно выполненные – прим. Белкова А.*) действия с неопределенными переменными в процессе построения модели приводят к получению «околоправильных усредненных результатов», которые в последующем используются для «тестирования реальности». Проблема в том, что разброс значений вокруг данного «околоправильного усредненного результата» является неверным. В результате лицам, принимающим решение, предоставляется крайне неточная оценка неопределенности, влияния риска на различные варианты принятия решений. Некоторые могут понять, что результаты моделирования нереалистичны и отклонить их, другие – нет и будут принимать ошибочные решения, основываясь на некорректных данных.

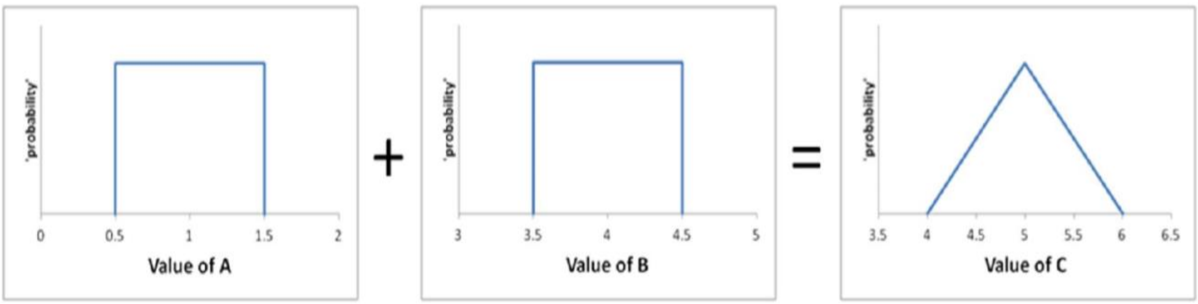
Дэвид Восе призывает вас внимательно изучить представленные ниже примеры и поделиться данной информацией со всеми коллегами, кто занимается построением рисковых моделей или планирует данную работу.

СЛОЖЕНИЕ

Из четырех рассматриваемых нами операций сложение представляется наименее проблематичным. **Единственным нюансом, который возникает при сложении является то, что две суммируемые переменные имеют определенную логическую связь между собой (то есть, существует определенная явная или не явная причинно-следственная связь, которая может приводить к корреляции значений переменных).** Иными словами, если мы хотим суммировать переменные, на которые воздействуют один или несколько внешних факторов, то простое сложение, скорее всего, нам не подойдет.

Помимо прочего, выполнение простого сложения двух неопределенных переменных достаточно ярко показывает, насколько не интуитивными являются результаты моделирования Монте-Карло и, следовательно, почему так опасно полагаться на свою интуицию при оценке правильности рассматриваемых результатов моделирования. Допустим, у нас есть две стоимости – A и B и мы хотим вычислить общую сумму – C. Если A = $1, а B = $4, то результат C будет равен $5.

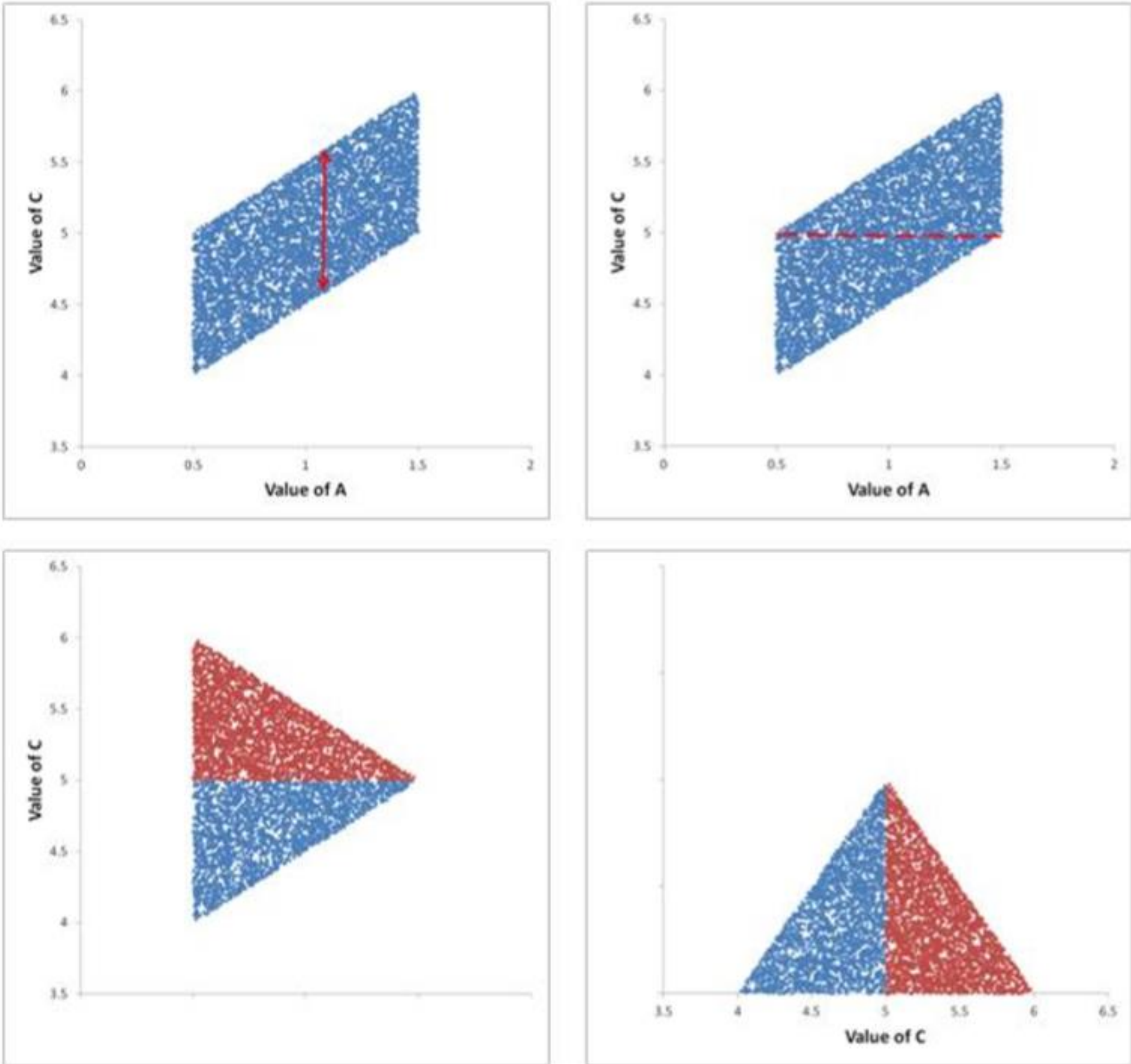
Теперь представим, что стоимости A и B являются неопределенностями. Значения стоимости A находятся где-то в диапазоне от $0.50 до $1.50, значения стоимости B – от $3.50 до $4.50. Это означает, что результат – значение C также является неопределенностью и находится в диапазоне от $4 до $6. На самом деле (технически, если значения A и B независимы), C скорее всего примет значение равное $5, что можно наглядно показать следующим образом:



При моделировании это будет выглядеть так:

Uniform (0.5,1.5) + Uniform (3.5,4.5) = Triangle (4,5,6)

Пользователи могут быть удивлены, что распределение неопределенности C принимает форму – Triangle вместо исходных форм сумм Uniform. И если неподготовленного человека спросить, как графически будет выглядеть сумма двух неопределенных переменных с распределением Uniform, скорее всего, ответ будет – Uniform. Но это не так. Приведенные ниже графики призваны объяснить факт получения распределения Triangle:



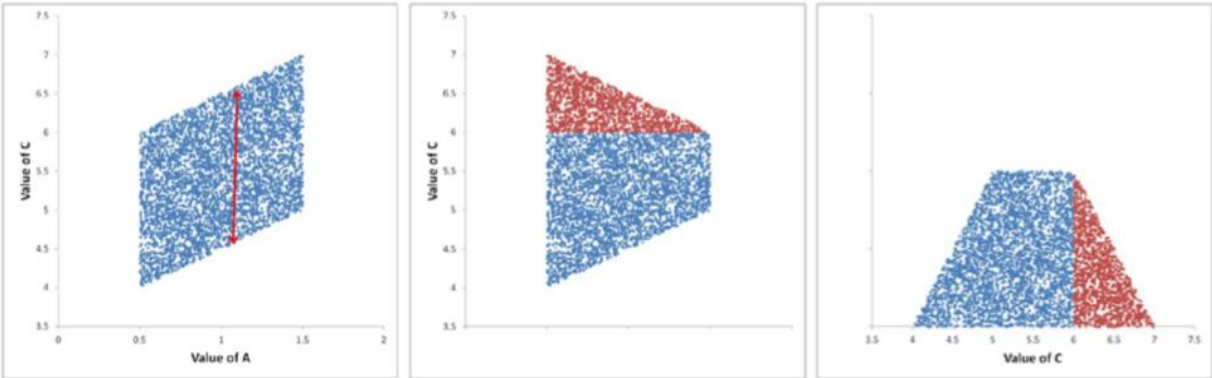
I. График смоделированных Монте-Карло значений А по отношению к итоговым значениям С. Например, в нижнем левом углу графика отображена выборка, где значения А и В стремятся к их минимуму (0,5 и 3,5 соответственно, что дает суммарное значение С равное 4). В верхнем правом углу отображаются значения выборки, где А и В стремятся к их максимуму (1,5 и 4,5 соответственно, что дает суммарное значение С равное 6). Диапазон значения В показан красными стрелками.

II. Точки статистически равномерно распределены по форме ромба, так как значения А и В распределяются равномерно. Представьте себе, что мы разделили смоделированные данные на две части в соответствии с показанной горизонтальной линией…

III. Затем переверните положение верхней части данных. Вы можете видеть, что распределение значения данных на ось OY отображается в виде треугольника.

IV. Заменив расположение осей OX и OY местами, мы получим отображение треугольного распределения. Обратите внимание, что вертикальная ось теперь отображает «вероятность» (если быть совсем точными, то даже не «вероятность», а «плотность вероятности», т.к. переменная непрерывна), так как высота треугольника пропорциональна количеству выпадающих значений, отраженных на горизонтальной оси по итогам моделирования.

В рассмотренном примере распределения суммируются в треугольное, потому что диапазон распределений A и B одинаков и равен $1, а именно: для A - Uniform (0.5,1.5) и для B - Uniform (3.5,4.5). Если бы ширина диапазонов распределений была разной, то графическое выражение результирующей суммы приобрело бы вид трапеции, как показано на нижеследующем примере графиков, где A - Uniform (0.5,1.5) и B - Uniform (3.5,5.5):



Не интуитивный характер такого простого вычисления подчеркивает сложность проверки правильности результатов построенной рисковой модели.

ВЫЧИТАНИЕ

Допустим, у нас есть две стоимости – A и B, которые суммируются в итоговое значение C, но нам известны только значения A и C. Если A = $1, а C = $5, мы можем вычислить значение B: = $5 - $1 = $4. Но, когда значения A и B являются неопределенностями, подобный расчет не допустим. Из предыдущего примера мы имеем:

A – Uniform (0.5,1.5)

B – Uniform (3.5,4.5)

C – Triangle (4,5,6)

И мы увидели, что A + B = C, т.е.:

Uniform (0.5,1.5) + Uniform (3.5,4.5) = Triangle (4,5,6)

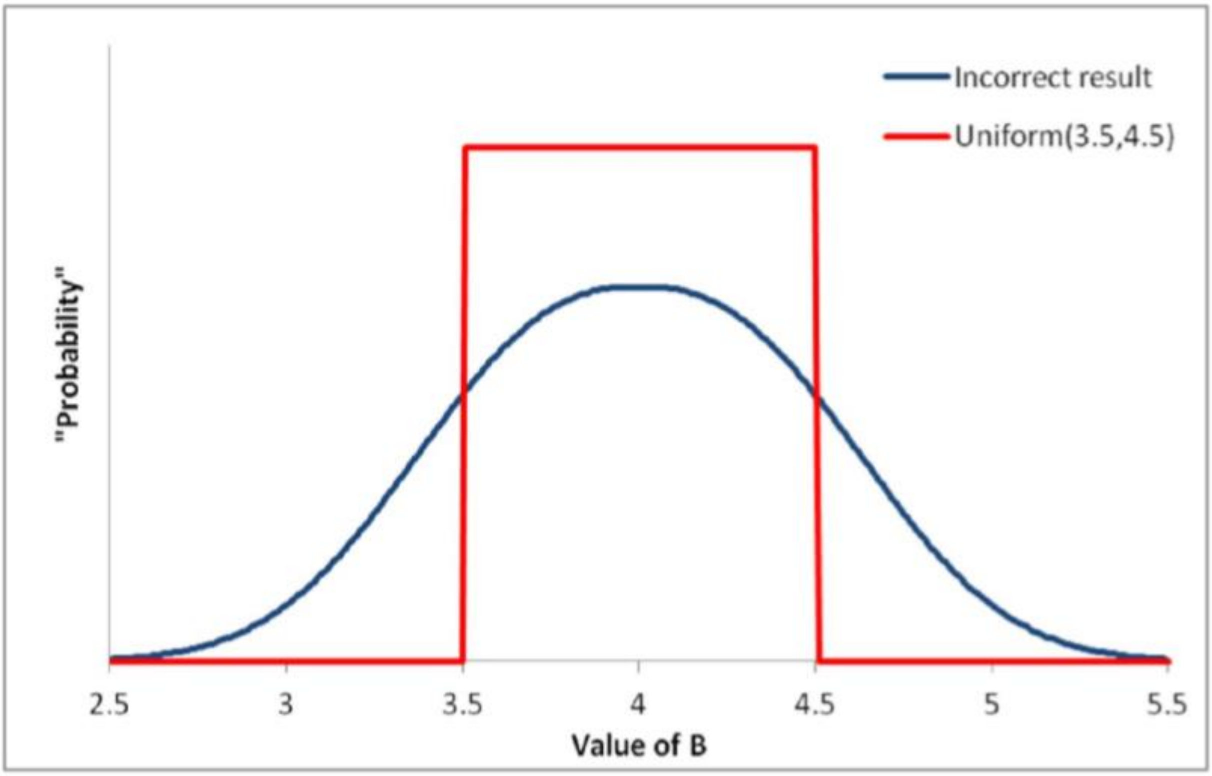
Законы простой алгебры заставили бы нас поверить, что C – A = B, т.е.:

Triangle (4,5,6) - Uniform (0.5,1.5) = Uniform (3.5,4.5)

На самом деле левая и права части этого уравнения очень разные. Если мы рассчитываем в модели:

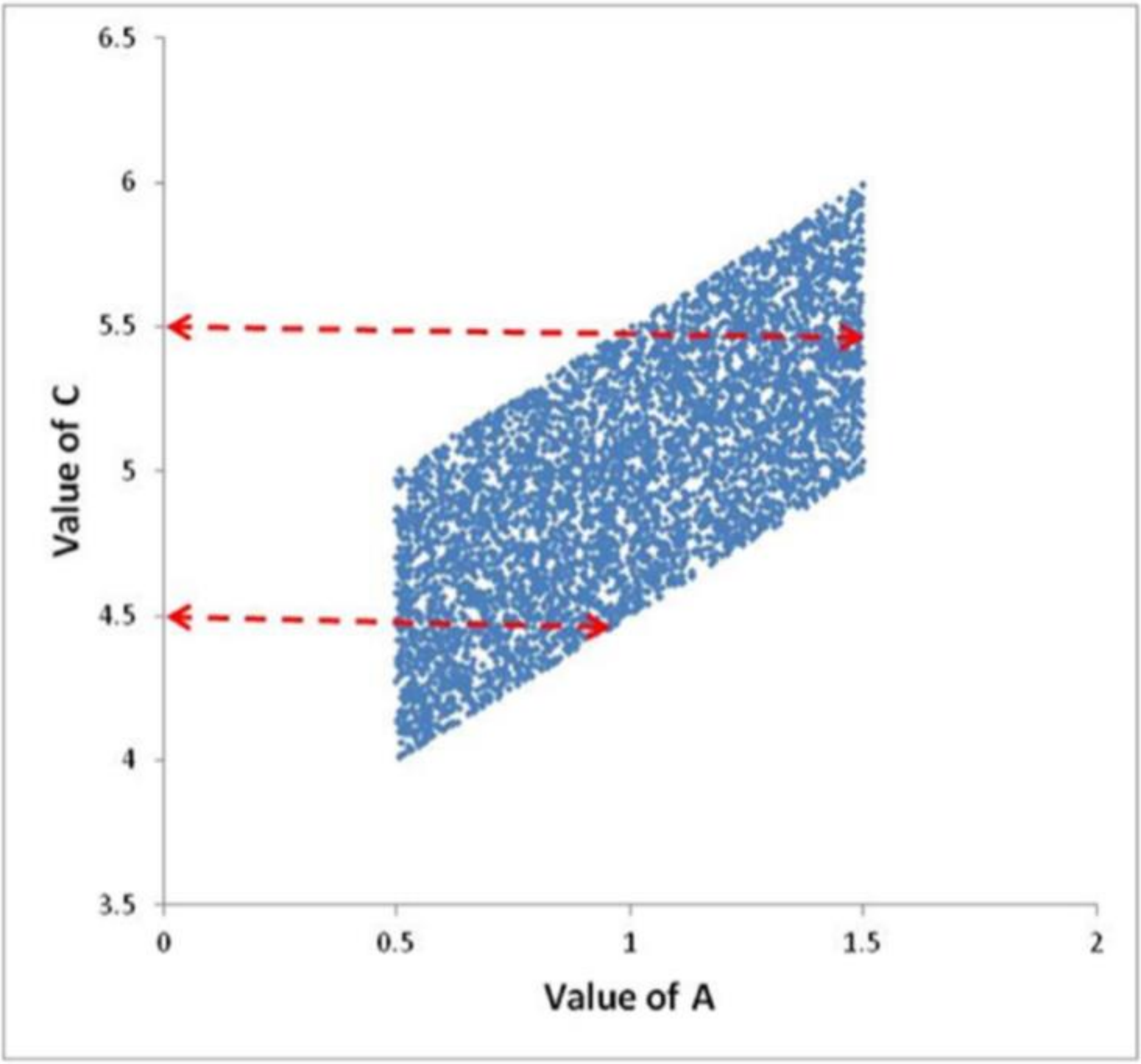
Triangle (4,5,6) - Uniform (0.5,1.5)

В надежде получить правильное распределение для B, мы фактически переоценили бы неопределенность значений B, как показано на рисунке ниже:



Что пошло не так?

Давайте снова обратимся к диаграмме распределения из предыдущего примера.



Если значение C ожидается 5,5 как показано верхней стрелкой, то значение A может находиться в диапазоне от 1 до 1,5. Подобным же образом, если C ожидается на уровне 4,5, A может находиться в диапазоне от 0,5 до 1. Иными словами возможное распределение A зависит от значения C, которое не учитывается в простой формуле:

C – A = Triangle (4,5,6) - Uniform (0.5,1.5)

Общее правило в данном случае такое, что необходимо избегать выполнения таких вычитаний как C – A при использовании моделирования Монте-Карло, в случае если суммирующее значение C включает в себя значение A.

Например:

- C = общая стоимость эксплуатации завода, A = затраты на персонал, C – A не будет объективно отражать затраты, не связанные с персоналом. **Вместо этого вы должны построить свою модель наоборот – рассчитать расходы на персонал и не связанные с персоналом, а затем сложить их вместе, чтобы получить общую стоимость.**
- C = выручка, A = стоимость, тогда C – A будет рассчитывать прибыль, если в модели учтены любые взаимосвязи между C и A (например, отношение к объему проданных товаров).

УМНОЖЕНИЕ

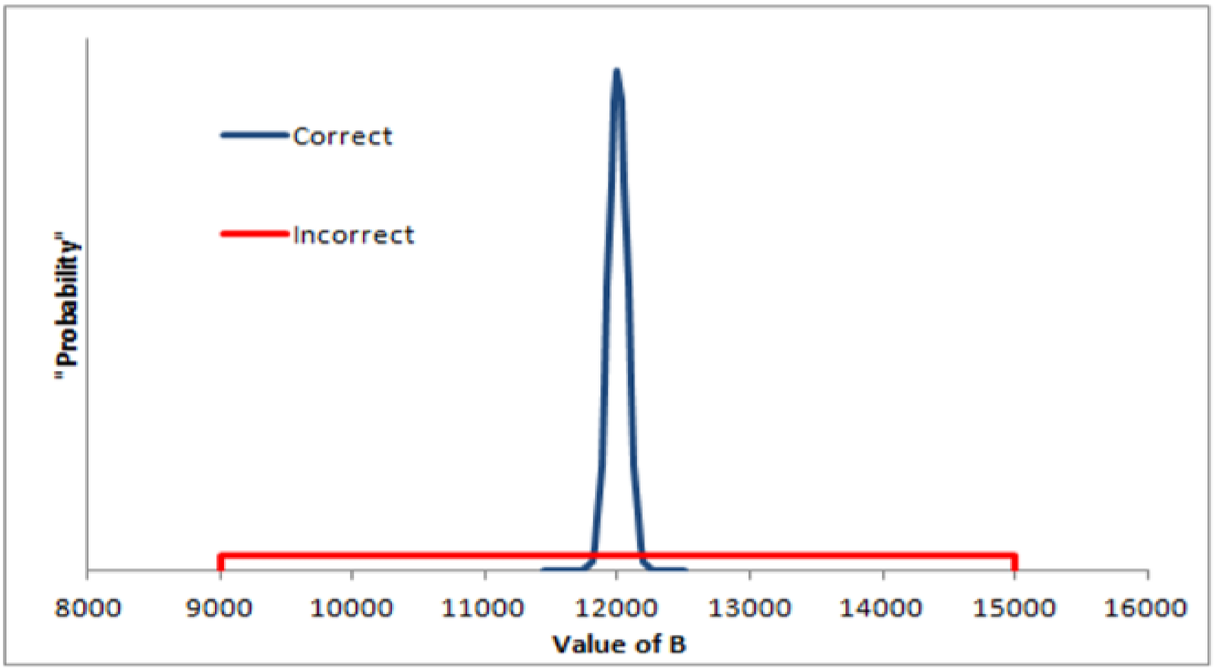
Допустим наш магазин посетило 600 покупателей и каждый из них потратил $20. Общая выручка магазина составит 600 * $20 = $12 000. Большинство бизнес моделей будут иметь множество подобных типов вычислений – умножений количества единиц на стоимость или доход за единицу.

Но если объем / стоимость покупок варьируется случайным образом между всем количеством покупателей, простое умножение является полностью неправильным и сильно переоценивает неопределенность дохода. Для наглядности предположим, что сумма, которую покупатель тратит в магазине в равной степени может составлять от $15 до $25. Неправильный расчет будет выглядеть следующим образом:

Суммарный доход = 600 * Uniform (15,25)

Проблема с этой формулой заключается в том, что для всех 600 клиентов используется только одна случайная выборка из равномерного распределения: так, например, если один клиент тратит 16 долларов США, все они делают - согласно формуле. Используя эту формулу, общий доход будет соответствовать Uniform (9000, 15000) распределению.

Что мы действительно хотим сделать, так это сложить 600 отдельных, независимых потраченных сумм, каждая из которых взята из равномерного (15,25) распределения, чтобы смоделировать поступление денег. Если мы сделаем это, ответ будет почти точно Normal (12000, 71) распределение. На графике ниже сравниваются результаты, которые делают этот расчет правильными и неправильными способами:



Разница между правильным и неправильным результатами моделирования очень существенная и тем больше, чем больше количество переменных (в данном случае, клиентов), которые суммируются. Если неправильный вариант предоставляет слишком драматичные результаты, обычно риск менеджеры это замечают. В менее драматичных ситуациях ошибка останется незамеченной, и рецензенту / менеджеру / лицу, принимающему решение, будет представлен анализ денежных потоков, который иллюстрирует гораздо большую неопределенность (риск), чем есть на самом деле. Это заставляет задуматься о том, сколько прекрасных бизнес-возможностей было упущено из-за простой ошибки моделирования.

Обратите внимание, что, когда вы добавляете несколько независимых переменных (в данном случае Uniform), сумма стремится к нормальному (то есть в форме колокола) распределению.

ДЕЛЕНИЕ

Почти каждый, кто начинает заниматься построением финансовых моделей и моделированием рисков, допускает ошибки, когда они включают в свои модели деление. Обычно это используется при расчете КПЭ. Это очень запутанно и интуитивно не понятно с самого начала. **Рекомендуется избегать использование деления в моделях.**

Разберем данный кейс с использованием предыдущего примера про магазин. Как мы выяснили ранее, общий доход магазина соответствует Normal (12000, 71) распределению. Допустим у нас есть эти данные и мы хотим выяснить сколько при этом тратит каждый покупатель.

Мы могли бы выразить это так:

Normal (12000, 71) / 600

Это средняя сумма, которую тратит каждый человек в магазине и это фактическая сумма если все посетители тратили одинаково. Данный расчет не делает разницы между 2 вариантами интерпретации. Однако, если каждый клиент тратит сумму, которая отличается и не зависит от других клиентов, нет никакого способа пересчитать распределение отдельных расходов (которое было Uniform (15,25)). С помощью приведенной выше информации мы не можем знать, каково распределение суммы, потраченной отдельными клиентами, но оказывается, что мы можем указать среднее значение и стандартное отклонение, если все они независимо принимают решения о покупке.

Есть и еще один не самый простой способ выполнить интересующую нас оценку. Если вы изучали математику, то рано или поздно должны были наткнуться на логарифмы. Представьте, что у нас есть две неопределенные переменные, X и Y, и что Z = X / Y. Уравнение можно записать в логарифмах следующим образом:

Log Z = Log X – Log Y

Если X и Y являются случайными переменными, Log X и Log Y также являются случайными переменными, и мы уже выяснили выше, что не можем вычесть одну из другой, поэтому деление одной независимо моделируемой случайной переменной на другую также, весьма вероятно, будет ошибочным действием.

Есть много проявлений этой проблемы. Например, всякий раз, когда слово «средний» встречается в вашем анализе (средняя стоимость за единицу, среднее время обслуживания клиента, средняя цена за сырье), появляется неявный знаменатель. Распределение и форма распределения среднего значения очень сильно зависит от числа значений, по которым мы усредняем. Та же проблема возникает, когда у нас есть данные, которые мы хотим использовать отражающие средние значения, а не набор отдельных наблюдений. Правильный анализ и использование таких данных требует хорошего знания вероятностей и статистики.