

Selecting the Right Distribution in @RISK

October 2009

By Michael Rees
mrees@palisade.com

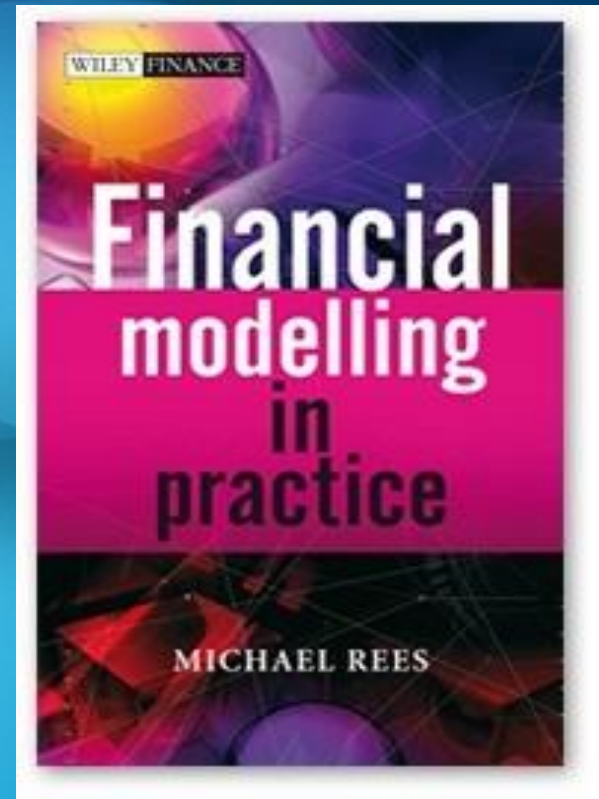
Michael Rees

Professional Experience

- » 20 years
 - Strategy consultant
 - Equity analyst
 - Independent consultant
- » Palisade Director of Training and Consulting

Qualifications

- » B.A. in Mathematics and D.Phil. in Mathematical Modeling from Oxford
- » MBA from INSEAD
- » Certificate in Quantitative Finance
- » *Financial Modelling in Practice* (John Wiley & Sons, October) [see Wiley.com for free .pdf of Ch 1; without the models]



Topics

- » A framework for selecting distributions
- » Distributions and properties that are new in v5.0/5.5
- » Examples

Is there a “right” distribution?

- » Theoretically correct? E.g. Based on some implicit model of the input's behavior
- » Conforms with industry standards?
- » Data driven? E.g. Data fitting or re-sampling
- Pragmatic? E.g. Intuitive to communicate and implement
- Mixtures of the above

Industry standards: Examples

Finance and Insurance

- » Normal distribution for change in price from one period to the next (for short periods; lognormal for longer periods)
- » Poisson distribution for number of events (crashes, earthquakes...)
- » Insurance: Pareto (or Lognormal) distribution for severity/magnitude of events

Oil and Gas

- » Lognormal distribution: Volumetric uncertainty on reserves, permeability
- » Normal distribution: Core porosity (%), % of abundant minerals, chemical elements or oxides in rocks

Data-driven approaches

Fitting

- » Data points are from the same random process (e.g. changes to, not the level of, house prices)
- » Distribution is one of those available in the fitting software
- » Fitting may not be appropriate where the data arises from compound processes, such as those involving several possible event risks

Resampling

- » (v4.5-5.0) RiskDUniform distribution (sampling with replacement) with historic or assumed data (sometimes: Discrete if data has already been grouped into frequency classes)
- » (v5.5) RiskResample to allow also for sampling without replacement, as well as re-sampling in order

Pragmatic Approaches

Continuous

- » Uniform
- » Triangular
- » PERT
- » Distribution Artist
- » Alternate parameter form
 - Incl. NormalAlt, LogNormAlt
 - i.e. “mixed approaches”

Discrete

- » Binomial
- » Discrete

Distributions: Summary

Uses	Conceptual	Pragmatic	Data driven
General	Binomial, Poisson (RiskCompound) Normal, Lognormal Ext Value, Hypergeometric	(Uniform) Triangular PERT Binomial Discrete Alternate Parameter RiskGeneral, RiskCumul (Distribution Artist)	Any fitted distribution Resampling
Waiting Time	Geometric, Exponential, Weibull Negative Binomial Gamma, Erlang Any: via intensity function for probs		Any fitted distribution Resampling
Parameter Estimation & Statistical	Beta, BetaGeneral, Gamma Chi-squared, Student, Pearson (F dist.)	PERT	Any fitted distribution Resampling
Other	Logistic, LogLogistic, InvGauss Pareto, Compound and MakeInput		Any fitted distribution Resampling

Lognormal distribution

- » Multiplication of random processes creates a positive asymmetry (skew)
 - Small multiplied by small is small
 - Large multiplied by large is very large
- » Conceptual models: Lognormal arises from the multiplication of many random factors
- » Examples of applications
 - Future value of an asset with random independent periodic growth rates
 - Uncertainty of oil reserves
- » $\text{Log}(A.B.C. \dots) = \text{Log}A + \text{Log}B + \text{Log}C + \dots = \text{Normal distribution}$
 - Product of many random variables is Lognormally distributed i.e. the logarithm of the product is normally distributed

Normal distribution

- » Addition of random processes creates symmetry e.g. Binomial process with $n=100$ and $p=0.1$; average no. of heads/tails = 10/90. Around this, there are:
 - 90 ways for a tail to become a head, each with probability 0.1
 - 10 ways for a head to become a tail, each with probability 0.9
 - So around the average, the distribution is approximately symmetric
- » Examples of applications
 - NPV of cash flows over several years
 - Goals in a soccer season
 - Earthquakes in the world during next 30 years
 - RTAs in the world per day/in a city per year
 - People arriving in a queue per day (or per minute over many queues)
 - Electricity used by all houses in a city
 - Travel time for a long journey
 - Total company expenditure, built from departmental expenditure
 - Total sales, sum of individual products
 - Total amount of oil in the world, based on amount in each field

Poisson distribution

- » Answers “How many?”, rather than “yes/no?”
- » Goals in football/soccer
 - Number of goals in a **short period of a game?**: RiskBinomial(1,p)
 - Number of goals during **the entire game?** RiskBinomial(n,p) where n is large and p small =>RiskPoisson(np)
 - Number of goals **over the whole season?** RiskBinomial(nm,p) where m is large, and np constant; RiskPoisson(mnp)=>RiskNormal(mnp, \sqrt{mnp})
- » Other examples
 - Breakdown/maintenance modelling
 - Queuing/service modelling (e.g. arrival rate)
 - Earthquakes in a country per year
 - Road traffic accidents in a city per day

Time to occurrence (constant probability or intensity)

Discrete Time

- » Questions
 - Which period will first goal occur in soccer (or: periods between goals)?
 - How many coin tosses of coin before getting a head?
 - How often will a gambler think she is “ahead of the game” (before losing!)
- » Simulation model (e.g. using MATCH function) to verify Geometric distribution

Continuous Time

- » Questions
 - Time until first goal (or between goals)?
 - Etc.
- » =>Exponential distribution

Time to occurrence – other examples

Multiple occurrences with constant intensity

- » Negative Binomial: No. of failures before s successes = $\text{RiskNegBin}(s,p)+s$ can be derived from mathematics (strictly: this is the *definition* of the negative binomial distribution)
 - $\text{RiskNegBin}(1,p) = \text{RiskGeomet}(p)$
- » Gamma: inter-arrival times for multiple events from a Poisson process
 - $\text{Gamma}(1, \beta) = \text{ExponDist}(\beta)$

Variable intensity processes

- » Typical examples assumed:
 - Weibull distribution
 - $\text{Weibull}(1, \beta) = \text{Exponential}(\beta)$ has constant probability
 - $\text{Weibull}(3.5, \beta)$ is approximately Normal
 - Gamma distribution
- » Often easy mathematics based on “intensity functions” for any assumed distribution

Parameter Estimation

- » The Beta distribution represents the uncertainty distribution for probability of a binomial process given observed data
 - You toss a coin 10 times and get 4 heads.
 - True probability of a head is uncertain, and given by $\text{Beta}(4+1, 10-4+1)$
 - If you score 40 heads with 100 coin tosses, $\text{Beta}(40+1, 100-40+1)$ has much narrower range
- » Similarly, the Gamma distribution represents the uncertainty for the intensity of a Poisson process

Summary: Key “first line” Distributions

Continuous

- » **Symmetric uncertainty**
 - **Triangular:** Pragmatic/use for (approx.) symmetric ranges/double bounded
 - **Normal:** Conceptually the sum of many random processes/Symmetric/double unbounded
- » **Asymmetric uncertainty**
 - **PERT:** Pragmatic/use for non-symmetric ranges/double bounded
 - **LogNormal:** Conceptually the product of many random processes/Non-symmetric/unbounded on one side
 - **Distribution Artist**

Discrete

- » **Event risks**
 - **Binomial** for occurrence or not (and as approximation to low intensity Poisson)
 - **Discrete** for several outcomes
 - **Poisson** for multiple occurrences (e.g. use within RiskCompound)
- » **Data**
 - **Distribution fitting**
 - **Resampling from data** (RiskResample or RiskDuniform)
- » **Industry standards**
- » **Alternate parameter formulation can allow “convergence” of methods** e.g. LognormAlt

Using RiskTheo functions

- » General exploration of properties of distributions
 - Properties of alternate parameter distributions e.g.
 - Implied volatility of a process that is calibrated with percentiles
 - Effect of alternate parameters on implied non-symmetry of distributions
 - Standard deviation of inputs e.g. when looking at tornado graphs
 - Implied biases of base cases in the context of non-symmetric uncertainty e.g.
 - Proportion of distribution above/below the modal values
 - Assessing the effect of sample sizes on model calibration

- » Approximation of one distribution by another
 - Normal by PERTAlt
 - Continuous by discrete

V5.0 and 5.5: Some new distributions and features

- » RiskTheo statistics functions
 - Cross-calibration when using alternate parameters
 - Distribution matching (e.g. Swanson's rule)
- » RiskResample
 - Allows for sampling with/without replacement, or stepping through historic data
- » RiskCompound (RiskMakeInput)
 - Frequency-severity modelling (creation of single sources of risk from multiple sources, e.g. for tornado charts)
- » RiskMakeInput
 - Event-Impact model
 - Aggregation of small risks into categories (e.g. annual operating expenses, where opex is made up of several line items)
- » RiskSplice
 - Creates single continuous distribution formed by joining two distributions together at the splice point

- » Johnson family

RiskCompound

- » Frequency-Severity Modelling e.g.
 - Damages from accidents, fires (including forest fires)
 - Operational losses e.g. resulting from accounting problems, fraud, rogue trading etc.
 - Medical injury modelling
 - General insurance and reinsurance pricing: Accidents, fires, property, catastrophes, earthquakes, weather damages, directors' liability insurance etc.
 - Capital requirements (e.g. using value at risk methods to determine e.g. the P99.5 point of a loss distribution)
 - Credit loss modelling
 - Other finance e.g. Modelling extreme exchange rate movements
 - Other: customer service costs
- » RiskCompound has optional arguments to layer the severity distribution (e.g. for excess of loss treaty)

RiskMakeInput

- » Treats argument to the function as a distribution for the purposes of sensitivity analysis, and excludes the distributions to this from all sensitivity analysis (even for that of another output)
- » Examples of uses
 - Event-Impact model with binomial distribution with $n=1$ for each event (could also use RiskCompound)
 - Aggregation of small risks into categories (e.g. annual operating expenses, where opex is made up of several line items)

Selection of Distributions: Recap

- » Conceptual?
- » Industry standards? e.g.
 - Oil and Gas: Reserves estimation using Lognormal distribution
 - Insurance: Number of events using Poisson distribution etc
- » Availability of data and its use?
 - Fitting a distribution to data (Fitting icon)
 - Re-sampling from the data (RiskResample or RiskDuniform distribution)
 - Use data to calibrate a pre-selected distribution
- » Pragmatic
 - Discrete versus continuous?
 - Symmetric or skewed?
 - Accuracy/defensibility versus communication objectives?
 - Requirement to capture un-boundedness?

- » Parameterisation using percentiles?

Last words

- » Don't panic!!! There is generally no absolutely correct distribution: Correct distributions exist in only idealized or conceptual situations
- » Be optimistic!!! As long as you use some common sense, your @RISK model will be a large improvement over your static model!!
- » Use a combination of techniques:
 - Think about the conceptual side
 - Think about the pragmatic side
 - Check whether historic data is valid for fitting etc