Operational Risk Analytics

A FineIT publication to measure operational risks under Basel framework

FineIT PVT Limited

Table of Contents

Operational Risk

I. Introduction

Operational risk was defined and included in Basel framework in 2004 by BCBS. The first Regulatory Directive of BCBS which was issued in 1988 and is commonly known as Basel I addressed the issue of capital charge calculations on the basis of credit risk only, ignoring both market and operational risks of financial institutions. In 1993, BCBS issued its second Regulatory Directive as an amendment to Basel I which added market risk component to credit risk but still ignored operational risk component. Finally, the third BCBS Regulatory Directive of 2004 which is commonly known as Basel II recognized operational risk and included operational risk component in its capital charge calculations.

II. Definition of Operational Risk

Basel Committee on Banking Supervision (BCBS) defines operational risk in Basel II framework as "The risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events." The concepts within this definition are further interpreted as follows:

Process: Losses due to a deficiency in an existing procedure or the absence of a procedure. Losses in this category can result from human error or failure to follow an existing procedure. Process-related losses are unintentional.

People: Losses associated with intentional violation of internal policies by current or past employees. In some specific cases, the risk extends to people who are being considered for employment.

Systems: Losses that are caused by breakdowns in existing systems or technology. Losses in this category are unintentional. If intentional technology-related losses occur, they should be placed in either the people or external category.

External Event: Losses occurring as a result of natural or man-made forces or the direct result of a third party's action.

Source: Harmantzis (2004)

The term "operational risk" has undergone a certain evolution over last 10 years and its contents differ according to different interpretations and uses. Basel's definition is also far from perfect and it does not include all operational risks, which daily threaten financial institutions. It is estimated that the definition in Basel framework reduces operational risk to about half of the actual size.

For example, this definition excludes set of strategic and reputational risks, despite the fact that these risks meet the characteristics of operational risks. On the other hand it however includes legal risks which are defined as, but are not limited to, exposure to fines, penalties, or punitive damages which may result from supervisory actions or private settlements.

The reason for this imperfection is probably that it is difficult to identify all operational risk factors, and thus the extent of the risks and potential impacts. For example existing or potential client's opinion of a particular bank is featured by various influences and circumstances that cannot be always recognized. It is then impossible to quantify reputational risk because a possible loss of business with existing customers under the influence of reputation risk or lost profits from potential clients who have opted for another company after losing the bank's reputation would have to be established. Despite these obvious imperfections, Basel II definition of operational risk has now been accepted and adopted as a general standard by the financial industry worldwide, however certain variations do exist.

III. Categories of Operational Risks

Basel II has divided operational risk events into 7 distinct types commonly known as Level 1 risks. These are further divided into 20 additional sub categories which are called Level 2 risks. Following table shows the list of Level 1 and Level 2 risks.

OPERATIONAL RISK CATEGORIES	
I	**Internal Fraud**
1	Unauthorized Activity
2	Theft & Fraud
II	**External Fraud**
3	Theft & Fraud
4	Systems Security
III	**Employment Practices & Workplace Safety**
5	Employee Relations
6	Sale Environment
7	Diversity & Discrimination
IV	**Clients, Products & Business Practices**
8	Suitability, Disclosure & Fiduciary
9	Improper Business Practices
10	Product Flaws
11	Selection, Sponsorship & Exposure
12	Advisory Activities
V	**Damage to Physical Assets**
13	Disasters & Other Events
VI	**Business Disruption & System Failures**
14	Systems
VII	**Execution, Delivery & Process Management**
15	Transaction, Execution & Maintenance
16	Monitoring & Reporting
17	Customer Intake & Documentation
18	Customer Account Management
19	Trade Counterparties
20	Vendors & Suppliers

Source: BCBS: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version. June 2004. (P. 144)

A brief description of Level 1 risk events is given as follows:

1. Internal fraud: These are losses due to acts of a type intended to defraud, misappropriate property or circumvent regulations, the laws or company policy, excluding diversity/discrimination events, which involves at least an internal party.

2. External fraud: These are losses due to acts of a type intended to defraud, misappropriate property or circumvent the laws, by a third party.

3. Employment practices and workplace safety: These are losses arising from acts inconsistent with employment, health or safety laws or agreements, from payment of personal injury claims, or from diversity / discrimination events.

4. Clients, products, and business practices: These are losses arising from an unintentional or negligent failure to meet a professional obligation to specific clients including fiduciary and suitability requirements, or from the nature or design of a product.

5. Damage to physical assets: these are losses arising from loss or damage to physical assets from natural disaster or other events.

6. Business disruption and system failures: These are losses arising from disruption of business or system failures.

7. Execution, delivery, and process management: These are losses from failed transaction processing or process management, in relations to trade counterparties and vendors.

IV. Calculations of Capital Charge for Operational Risk

Under Basel framework, banks are required to adopt a methodology to assess the operational risk capital charge that would serve as a shield against potential future losses given a one-year horizon. Basel framework has suggested three quantitative methods for determining the capital requirement for operational risk ranging from very simple to very advanced models. These methods include:

- Basic Indicator Approach – BIA
- Standardized Approach – SA
- Advanced Measurement Approaches - AMA

Financial institutions have flexibility to choose the most suitable method in terms of their activities, capabilities and risk profiles as long as they meet some conditions. BCBS document *Sound Practices for the Management and Supervision of Operational Risk*, published in February 2003 provides detailed qualifying criteria that participating financial institutions should adhere to when adopting a particular approach.

Basel framework requires financial institutions to select the simplest approach to start with and gradually step-up with an objective to reach advanced approaches in medium to long term.

Financial institutions are also allowed to use a combination of approaches for different business lines which is known as Partial Use. Once an approach is chosen, a bank will not be permitted to revert to a less sophisticated approach.

More sophisticated approaches should in theory permit greater benefits, both in terms of the capital charge as well as in terms of more effective and efficient business practices. Transition from simple to advanced approaches technically requires availability of credible historic data as well as modeling & analytical expertise within financial institution. In certain cases, financial institution requires permission from supervisor before adopting a particular advanced method.

The Basel framework provides for supervisors to review the capital requirement produced by the operational risk approach used by a financial institution (whether Basic Indicator Approach, Standardized Approach or AMA) for general credibility, especially in relation to a firm's peers. In the event that credibility is lacking, appropriate supervisory action under Pillar 2 of Basel framework will be considered.

V. The Basic Indicator Approach - BIA

Capital charge under Basic Indicator Approach (BIA) is calculated as % of previous three years average positive annual gross income. Gross income under BIA has a specific definition which differs from standard accounting definitions. BIA is regarded as the simplest method and there is no criterion or condition for a bank to use it. Due to its easy construction and minimum data requirements the BIA method is considered most suitable for smaller banks with simple risk management systems.

1. Numeric & Descriptive Formulae

Descriptive and numeric formulae for capital charge calculations under Basic Indicator Approach are as follow:

Descriptive Formula: Capital Charge = 3 Years Average Gross Income X Alpha Multiplier

Where,
Alpha Multiplier = Applicable % which is decided by the regulator. Currently Alpha is 15%.
Gross Income = Sum of net interest income and net non-interest income for previous 3 years.

Numeric Formula: $$K_{\mathrm{BIA}} = \frac{\sum_{i=1}^{n}\left(GI_i * \alpha\right)}{n}$$

Where,
K = Total Capital charge under Basic Indicator Approach
Σ = Summation sign
GI = Gross income, where positive over previous 3 years
α = Alpha (15% under current Basel guidelines)
n = Number of years for which gross income was positive (3 years under current Basel guidelines)

2. Analytical Points

- In the above formulae, the word gross income is actually misleading since it is calculated partly as net income or operational income in reality. However, Basel framework likes to call it gross income for some unknown reason.

- Several researchers and analysts have tried to ascertain the basis for taking 15% as Alpha Multiplier but its theoretical or practical rationale has not been established so far. Neither did BCBS give any rationale for this percentage in its directives.

- If the annual gross income is negative or zero for any year, figures for that year are excluded from both the numerator and denominator when calculating the average gross income.

- If a bank does not have the required historic information because it has been operational for less than three years, then the bank is allowed to use the gross income values assumed in its projected business plan.

- Basel framework is specific about what constitutes gross income with following general formula: Gross Income = Net Interest Income + Net Non-Interest Income.

- The incomes in formula are gross of any provisions including unpaid interest and gross of all other operating expenses including fees paid to outsourced service providers.

- Gross Income in formula does not include profit & losses from the sale of securities in the banking book or any extraordinary income including income derived from insurance.

VI. Standardized Approach - SA

Standardized Approach is very similar to Basic Indicator Approach, but in this case instead of taking total gross income of financial institution and multiplying it with a universal 15% Alpha, separate gross incomes are calculated for all business lines of financial institution and multiplied by specific assigned percentages which are called Beta Factors. Beta Factor for each business line is separate as per risk associated with that particular business operation.

The annual capital charge under this approach is the sum of the products of the relevant business line gross incomes and the beta factor, with the beta factor being a proxy for the assumed industry-wide relationship between the operational risk loss experience for a given business line and the aggregate level of gross income for that business line.

In order to qualify for the Standardized Approach, financial institutions need to comply with a set of minimum entry standards. Any financial institution that wishes to adopt Standardized Approach has to satisfy regulator that institution's senior management has active involvement in overseeing operational risk management framework; it has conceptually sound operational risk management policies, processes and systems which are implemented with integrity; and it has sufficient technical resources to use Standardized Approach across its business lines.

1. Basel II Business Lines & Beta Factors

Basel framework breaks out financial institution's operations into eight standard business lines. Each of these eight business lines has a different beta factor to calculate capital. The eight business lines and their respective beta factors are shown in the following table.

BUSINESS LINES

No	Business Lines	Beta
1	Corporate Finance	18%
2	Trading & Sales	18%
3	Payments & Settlements	18%
4	Commercial Banking	15%
5	Agency Services	15%
6	Retail Banking	12%
7	Asset Management	12%
8	Retail Brokerage	12%

Source: BCBS: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version. June 2004. (P. 147)

2. Numeric & Descriptive Formulae

Descriptive and numeric formulae for capital charge calculations under Standardized Approach are as follow:

Descriptive Formula: Capital Charge = 3 Years Simple Average of (8 Business Lines Gross Incomes X Applicable Beta Factors)

Where,
Beta Factors = Applicable % for each of the eight business lines as shown in above chart

Numeric Formula: $$K_{\text{TSA}} = \frac{\sum_{j}^{3} \max\left[\left(\sum_{i}^{8}\left(GI_i \ * \beta_i\right)\right), \ 0\right]}{3}$$

Where,
K = Total capital charge under the Standardized Approach
GI = Annual gross income in a given year for each of the eight business lines
β = Applicable % for each of the eight business lines as shown in above chart
3 = Number of years (3 years under current Basel guidelines)

3. Analytical Points

- In the Standardized Approach gross income is measured for each of eight business lines separately and not for the whole institution as in Basic Indicator Approach. For example, in corporate finance, the indicator is the gross income generated in the corporate finance business line.

- The total capital charge is calculated as the three-year average of the simple summation of the annual capital charges across each of the business lines in each year.

- For retail and commercial banking there is also an Alternative Standardized Approach (ASA) available, introduced to eliminate double counting of risks. Banks, at the national supervisor's discretion, may be permitted to substitute an alternative measure in the case of retail and commercial banking. In this case, the volume of outstanding loans will be multiplied by the beta factor and the result multiplied by 3.5%.

  Descriptive and numeric formulae for Alternative Standardized Approach are as follows:

  Descriptive Formula: Capital Charge = Outstanding Loans X Applicable Beta % X 3.5%

  Where,
  Outstanding Loans = 3 years average retail loans and advances gross of provisions

  Numeric Formula: KRB = βRB x m x LARB

  Where,
  KRB = Capital charge for the retail banking business line
  βRB = Beta for the retail banking business line
  LARB = Total outstanding retail loans and advances (non-risk weighted and gross of provisions), averaged over the past three years
  m = 0.035 or 3.5%

- In any given year, negative capital charges resulting from negative gross income in any business line may offset positive capital charges in other business lines without limit.

- However, where the aggregate capital charge across all business lines within a given year is negative, then the input to the numerator for that year will be zero.

- If it is assumed that gross incomes from all 8 business lines are equal for any financial institution, then the capital charge calculated using Standardized Approach will actually be equal to capital charge calculated using basic Indicator Approach as average beta factor is still 15%.

VII. Advanced Measurement Approaches - AMA

Advanced Measurement Approaches are fundamentally different from Basic Indicator Approach and Standardized Approach. In case of the BIA and the SA, all the parameters are determined by a regulator when the capital requirement for operational risk is calculated. In case of AMA methods, the financial institution's calculations and its real history of losses are taken into account.

Under the AMA the financial institutions are allowed to develop their own empirical models to quantify required capital charge for operational risk. Financial institutions wishing to use this approach need to meet certain conditions and require approval from their local regulators. Regulators give approval for the usage of AMA methodologies on the basis of financial institution's internal technical capabilities, soundness of risk management systems and strength of risk management framework. Once a financial institution has been approved to adopt AMA by a regulator, it cannot revert to a simpler approach without regulatory approval. Such approvals are given only in case of extra ordinary circumstances.

1. Data Requirements for AMA Models

Measurement of operational risk to determine the capital charge comes with a great challenge of collecting loss data. An operational risk is more difficult to measure than market or credit risk, due to the non-availability of objective data, presence of redundant data and the lack of knowledge of what to measure.

The data requirements for measuring market risk are very straightforward, such as prices, volatility and other external data. These are packaged with significant history in large databases which are easily accessible and measurable. Similarly, credit risk relies on the assessment and analysis of historic and factual data, which is again easily available in banking systems.

However, operational risk events are in contrarily hard to detect. Operational risk, in theory is an ill-defined inside measurement, related to the measures of internal performance, such as internal audit ratings, volume, turnover, error rates, income volatility, interaction of people etc. On top of that uncertainty about which factors are important arises from the absence of a direct relationship between the risk factors usually identified and the size and frequency of losses.

In order to develop an AMA model, financial institutions needs a 3 years historic database of Internal Loss data and External Loss Data as a minimum requirement. Financial institutions collect this historic operational loss data and register it in a database which is called Loss Database.

The fundamental premise of collecting operational loss data is that each firm's operational loss is a reflection of its expected operational risk exposure. However, the challenge of measuring operational loss is the calculation of the unexpected loss. Unexpected operational losses are losses that have occurred, but they are not registered in the database. In fact, they are for the most part unknown. Given below is a brief description of loss database and how it is collected and cleaned.

2. What is Loss Data?

Operational loss databases are a collection of number of occurrence of operational risk events called "Frequency" and financial impact of these risks called "Severity". The Frequency is divided into 3 categories of High frequency, Medium Frequency and Low Frequency. Similarly Severity is also divided into 3 categories of High Severity, Medium severity and Low Severity. This can be represented in a 9 cell matrix showing 9 combinations of frequency and severity on high, medium and low scale.

For example, event with high-frequency/low-severity can be processing errors in a high-volume business. These types of operational losses are expected, as they are easier to be detected. Calculating capital for these expected events will be as easy as examining the historical data. Financial institutions can even budget them as an expected cost of doing business. It is only the larger than expected losses that create downside volatility in a financial institution's earnings and they are the high-severity incidents of major frauds and errors.

Even when the high-severity events are detected, it is hard to tell the size of the financial damage that has been caused. Most of the time, the actual losses will be known weeks or months after the first moment of observation. These rare events that threaten the solvency of financial institution will be referred as unexpected losses.

The most important part of a loss database is the integrity of data. Whether a financial institution builds its own database or buys a commercial database, it is important that the processes and systems must assure good quality. The point to remember is that working with bad data can produce false result, which is far more dangerous than having no data at all.

3. What is Internal Loss Data?

Internal loss data are the collection of loss data which are located in financial institution's internal sources. Proper record of internal loss data is a prerequisite for the development of AMA measurement models. Internal loss data is also important for tying a bank's risk estimates to its actual loss experience. To qualify for regulatory capital purposes, a financial institution's internal loss data collection processes must meet the following standards:

- **Mapping of historical internal loss data into relevant categories and business lines**
  Internal loss data must be mapped into the relevant business line and event type and objective criteria for allocating losses into specified business lines and event types must be properly documented.

- **Capturing all activities & exposures from all sub-systems and geographic location**
  Internal loss data must be comprehensive enough to capture all material activities and exposures form all appropriate sub-systems and geographic location. Financial institutions must decide a threshold for internal data collection which represents a minimum amount of severity and all risk events where severity is greater than the assigned threshold must be recorded.

  The appropriate threshold can vary between financial institutions and even between business lines and event types within a financial institution. Also, when any activities and exposures are excluded individually and in combination, financial institution must justify that this would not have any material impact on the overall risk estimates in order to keep the integrity of analytics.

- **Dating of event and recoveries of gross loss amounts**
  In addition to gross loss amounts relating to severity of risk events, banks must also collect and record information about the data of events and recoveries of gross loss amounts together with some descriptive information about the drivers and causes of the loss event.

- **Treatment of credit risks events**
  Operational risk losses that are related to credit risk and have historically been included in the credit risk database of financial institutions are treated as non-operational losses. For example collateral management failures are treated as credit risk for the purpose of calculating minimum regulatory capital. Therefore, such losses are not subject to the operational risk charge.

- **Treatment of market risk events**
  Market risk related operational risk would be subjected to operational risk. Operational risk losses that are related to market risk are treated as operational risk for the purpose of calculating minimum regulatory capital, and therefore will be subjected to the operational risk capital charge.

4. What is External Loss Data?

It seems to be generally accepted in the finance industry that internal loss data alone is not sufficient for obtaining a comprehensive understanding of the risk profile of a financial institution. This is the reason why additional data sources have to be used to capture, in particular external losses.

Source: Basel Committee on Banking Supervision, 2006.

External loss data is basically collection of internal loss data of other financial institutions within the local industry. External loss data therefore should have same characteristics as of internal loss data described above. External loss data is used to capture relevant operational loss events within industry and assess infrequent and potentially severe losses.

External data should include data on the actual loss amount, information on the scale of business operations where the event occurred, information on the causes and circumstances of the loss events.

There are many ways to incorporate external data into the calculation of operational risk capital. External data can be used to supplement an internal loss data set, to modify parameters derived from the internal loss data, and to improve the quality and credibility of scenarios. External data can also be used to validate the results obtained from internal data or for benchmarking.

In LDA models, external data is used as additional data source for modeling tails of severity distributions. The obvious reason is that extreme loss events at each financial institution are so rare that no reliable tail distribution can be constructed from internal data only.

5. Loss Data Collection Procedure

A clear data collection procedure is a key to robust collection of loss data. Data collection procedure must clearly define what data is to be collected together with standards, roles and responsibilities for its collection.

Considering the basic definition that an operational risk loss is the amount charged to the profit & loss statement net of recoveries, the following table shows the essential elements of operational loss database.

REQUIREMENTS FOR RECORDING LOSS DATA

No	Data Inputs
1	Date of Event Occurrence
2	Date of Event Discovery
3	Date of Event Write-off
4	Location of Event Occurrence
5	Name of Financial Institution
6	Level 1 Type of Event Category
7	Level 2 Type of Event Category
8	Amount of Loss
9	Severity of Loss
10	Loss Recovery Amount
11	Loss Recovery Source
12	Cause of Event

6. Loss Data Cleaning Procedure

Loss data collected from internal as well as external resources is generally dirty data which needs to be cleaned before its use in analytics. Internal data needs to be audited, classified, scaled, weighted and truncated and external data needs to be cleaned from scale bias, truncation bias and data capture bias.

- Data auditing is the process of checking accuracy of data points and incorporating missing values.

- Data classification refers to checking the distribution of loss in respective categories of business lines and event types. This is especially relevant in case of Split Losses where one loss amount is distributed between two different business lines on the bases of weights.

- Data scaling refers to converting historic nominal loss amounts into real inflation adjusted amounts today. A 3 years earlier loss of $100 will be recorded as $100 plus compounded effect of 3 years inflation.

- Data weighing gives weights to historic data on a time scale basis. Last year data is considered more relevant and has more weight as compared to 10 years old data.

- Truncation is the process of establishing a minimum threshold of loss amount and ignoring all values that fall below established threshold.

- Scale bias refers to the fact that operational risk is dependent on the scale of operations of a financial institution. A bigger institution is exposed to greater operational failures and therefore to a higher level of operational risk. The actual relationship between the size of the institution and the frequency and severity of losses is dependent on the measure of size and may be stronger or weaker depending on the particular operational risk category.

- Truncation bias refers to fact that financial institutions collect data above certain thresholds which may be different from each other. It is generally not possible to guarantee that these thresholds are uniform

- Data capture bias refers to a systematic bias that a positive relationship exists between the loss amount and the probability that the loss is reported. If this relationship does exist, then the data is not a random sample from the population of all operational losses, but instead is a biased sample containing a disproportionate number of very large losses and statistical inferences based on such samples can yield biased parameter estimates.

7. Loss Data Mapping Procedure

Once internal and external loss data is collected and cleaned, these databases need to be mapped. This process is done into following 2 steps:

Mapping of loss data to loss events: This step involves distribution of collected loss data into 7 categories of Level 1 risk events. Level 1 risk events, mentioned earlier, include internal frauds; external frauds; employment practices and workplace safety; clients, products, and business practices; damage to physical assets; business disruption and system failures; and execution, delivery, and process management.

Mapping of loss data to business line: This step involves distribution of collected loss data into 8 categories of business lines. Business lines, mentioned earlier, include corporate finance; trading & sales; payments & settlements; commercial banking, agency services; retail banking; asset management; and retail brokerage.

8. Components of AMA Models

AMA itself does not specify the use of any particular modeling technique to calculate capital charge for operational risk. Financial institutions are at liberty to develop their own empirical models as long as these adopted models meet certain conditions. However, in the development of these systems, financial institutions must have and maintain rigorous procedures for operational risk model development and independent model validation.

All AMA models must meet a certain minimum criterion in order to be qualified as proper models. This criterion includes the following:

- AMA model should ideally be based on 4 data sets which are called elements of AMA model. These data sets include internal data, external data, scenario analysis and business environment & internal control factors. Any AMA model must at least use internal data and scenario analysis as a minimum requirement.

- Internal data refers to financial institution's historical data of operational loss events. The data should have 2 components i.e. frequency and severity. Frequency represents the number of times a particular risk event occurred and Severity represents the financial impact of the risk occurrence. The internal data needs to be ideally recorded across 3 timelines i.e. date of occurrence, date of discovery and date of accounting record. A minimum of 3 years of historic internal data is required for an AMA model.

- External data refers to either public data and/or pooled industry data. These external data should include data on actual loss amounts, information on the scale of business operations where the event occurred, information on the causes and circumstances of the loss events.

  A financial institution must have a systematic process for determining the situations for which external data is used and the methodologies used to incorporate the data e.g. scaling, qualitative adjustments etc. The conditions and practices for external data use must be regularly reviewed, documented, and subject to periodic independent review.

- Scenario analysis refers to assessment of plausible severe losses under assumed statistical loss distribution. A financial institution must use scenario analysis of in conjunction with external data to evaluate its exposure to high-severity events. Scenario analysis should also be used to evaluate potential losses arising from multiple simultaneous operational risk loss events. Such assessments need to be regularly validated through comparison to actual loss experience to assess their reasonability.

- Business environment and internal control factors refer to those elements that are key drivers of risks. Any improvement in the control of these drivers will result in decrease of risk probability and any deterioration in the control will cause an increase of risk probability. Numerical impact of the management of these factors needs to be incorporated in the empirical estimates of risk in model.

- AMA model must be able to calculate capital charge as the sum of expected loss (EL) and unexpected loss (UL). In case financial institution can demonstrate that it is adequately capturing expected loss (EL) in its internal business practices, capital charge will be equal to unexpected charge (UL) only.

- AMA model must demonstrate that its operational risk measure meets a soundness standard comparable to that of the internal ratings-based approach for credit risk. This means model must be able to calculate capital charge for one year holding period with a 99.9th percentile confidence interval.

- AMA model must be sufficiently 'granular' to capture the major drivers of operational risk affecting the shape of the tail of the loss estimates.

- Capital charge calculated by AMA models should not be less than 75% of capital charge calculated under Standardized Approach. This floor needs to be maintained unless approved and allowed by the regulator.

9. Categories of AMA Models
The models developed under AMA approaches fall into following three categories depending upon underlying methodology:

- Internal Measure Approach – IMA
- Loss Distribution Approach – LDA
- Score Card Approach – SCA

a. Internal Measurement Approach - IMA
IMA models are basically modified versions of Standardized Approach. Standardized Approach calculates capital charge by multiplying gross income of 8 business lines with pre-decided Beta Factors. IMA models are developed along the same lines.

In the IMA Models, financial institution decides their own indicator of exposure i.e. gross income, number of transactions, trading volume etc. and determines individual capital charge for all 56 combinations of 8 business lines and 7 risk events. Total capital charge for operational risk is calculated as simple sum of 56 individual capital charges.

The Internal Measurement Approach provides discretion to individual banks in the use of internal loss data. In this approach banks estimate the operational risk capital based on the measurement of the total expected losses. The IMA approach assumes a fixed, direct relationship between expected loss which is the mean of the loss distribution and the unexpected loss which is the tail of the distribution. The relationship can either be linear or non-linear. A linear relationship implies that the capital charge is a simple multiplication of the expected loss with a fixed number. While a non-linear relationship implies that total capital charge will be a more complex function of expected losses.

The IMA approach calculates the capital charge based on a framework that divides a bank's operational risk exposure into 56 combinations of business lines and event. In such a framework separate expected losses are calculated for each business line and event type combination. Such an approach, calculates the expected losses generally by estimating the loss frequency and the size of the amount for various business line and event combination by using internal loss data and, where appropriate, relevant external loss data, along with a measure of the scale of business activities for the particular business line in question.

1. Calculation of Capital Charge in IMA Models

The capital charge is determined in IMA models as the product of three parameters:

- The Exposure Indicator – EI
- Probability of Event – PE
- Loss Given the Event – LGE

The product EI × PE × LGE is used to calculate the expected loss (EL) for each business line/loss type combination. The EL is then rescaled to account for the unexpected losses (UL) using a parameter γ (gamma). Gamma is different for each business line/loss type combination and its values are predetermined by the supervisor.

2. Descriptive and Numeric Formulae

The total one-year regulatory capital charge is calculated as per following descriptive and numeric formulae:

Descriptive Formula – Calculation of Expected Loss: Expected Loss = Exposure Indicator X Probability of Event X Loss Given the Event

Where,
Exposure indicator = Values of gross income or number of transactions or trading volume etc.
Probability of event = Statistical probability for risk event occurrence.
Loss given event = Financial impact of risk event.

Descriptive Formula – Calculation of Capital Charge: Capital Charge = Sum of (Expected Loss x Gamma) for 56 business line & risk events combination.

Where,
Expected loss = As calculated above.
Gamma = Applicable % for each business line & risk type combination as decided by supervisor.

Numeric Formula – Calculation of Expected Loss: EL = EI X PE X LGE

Where,
EL = Expected loss
EI = Event indicator
PE = Probability of event
LGE = Loss given event

Numeric Formula – Calculation of Capital Charge: $\sum (\gamma \times EL)$

Where,
$\sum$ = Summation mark
Y = Gamma value
EL = Expected loss

The main drawbacks of this approach are the assumptions that there is a perfect correlation between the business line/loss type combinations and there is a linear relationship between the expected and unexpected losses.

b. Score Card Approach - SCA
In the Score Card Approach, financial institutions first determine operational risk capital charges for each business line and then modified the amounts of these capital charges according to an operational risk scorecard.

Scorecard approach differs from IMA and LDA approaches in a way that it relies less exclusively on historical loss data in determining capital amounts. Instead of this, after the size of the regulatory capital is determined, its overall size and its allocation across business lines are modified on a qualitative basis. However, historical operational risk loss data must be used to validate the results of scorecards.

The calculation of operational risk capital charges for business units on theoretical basis follows five basic principles:

- **Principle-1:** If the world gets riskier, the business units need more capital;
- **Principle-2:** If a business unit's size increases, so does its capital;
- **Principle-3:** If the business of a business unit is more complex, it needs more capital;
- **Principle-4:** If the level of control of a business unit is higher, it needs less capital;
- **Principle-5:** If the business units' losses from internal incidents exceed the level of expected loss accounted for in the first four framework principles, it needs more capital.

While IMA and LDA models are based on first 3 principles, SC models are developed and calculate capital charge on the basis of fourth principle and it combines quantitative calculations with qualitative judgment.

1. Calculation of Capital Charge in SC Models
Operational risk capital charge in Score Cards models is calculated in following 3 steps:

- Calculation of initial capital charge
- Development of score card & risk scoring
- Adjustment of initially calculated capital charge on the basis of score card ratings

Calculation of initial capital charge: Under SC approach, initial capital charge can be calculated by using a variety of methods that include:

- Standardized Approach
- Loss Distribution Approach
- Benchmarking proportions of total capital e.g. 20%
- Benchmarking vs. other peer institutions
- Benchmarking vs. capital for other internal risk types
- Explicit consideration of low frequency, high severity loss scenarios

The choice of an appropriate method for the calculation of initial capital charge depends upon the basic risk profile of a financial institution. An essential prerequisite for such capital level to be "right" for a particular financial institution is that it must be accepted and used by the Executive Management of that financial institution. This is an appropriate executive accountability.

Development of score card and risk scoring: Development of score card is the most critical and time consuming issue in SC approach. Scorecards aim to measure the quality of key operational risk management processes within a financial institution. The scorecard procedure is based on questionnaires that require quantitative data, qualitative judgments or simple yes/no questions e.g. indicating compliance with certain group policies. These questionnaires are developed by experts with 2 key objectives which are assessment of firm's exposure to specified risk drivers and quality of firm's internal control system and processes to control these risk drivers

Separate questionnaires are developed for each of 8 business lines incorporating business line specific operational risk questions with each question having different weight. These scorecards questionnaires are completed by all business units using self-assessment and reviewed by an expert panel who determines the final score for each business unit.

In order to understand score calculation of a score card questionnaire, let us assume that a specific business line questionnaire contains 5 questions with separate weights and these questionnaires are filled by 5 relevant personnel of business unit.

Assumed scoring bands and questions weights are shown in following tables:

SCORING BANDS

Band	Score
Very High	10
High	7
Moderate	5
Low	3
Very Low	1

QUESTIONS WEIGHTS

Questions	Weights
Question 1	10%
Question 2	20%
Question 3	40%
Question 4	20%
Question 5	10%

The scoring band refers to the interpretation of score given by personnel to each question with 1 being most optimal score point and 10 being most suboptimal score point. A score point of 1 would represent very low risk level while a score point of 10 will represent a very high level of risk.

The weight of each question refers to magnitude of risk contribution of that particular question in overall risk. In the above chart, question 3 has the highest contributing factor in assessment of risk and questions 1 & 5 have lowest contribution factors in the assessment of overall risk within business unit. Responses of 5 personnel within business unit are shown in following table:

RESPONSES

Questions	Response 1	Response 2	Response 3	Response 4	Response 5	Average
Question 1	7	4	3	6	8	5.6
Question 2	8	5	8	8	7	7.2
Question 3	6	8	7	6	8	7.0
Question 4	8	6	5	7	9	7.0
Question 5	7	8	6	5	10	7.2

Simple multiplication of average scores of each question with its respective weight gives the Weighted Scores. The sum of these weighted scores gives the total score for the business unit which is called Residual Risk Score (RRS). The Residual Risk Score for the business unit is 6.9 as shown in following table:

WEIGHTED SCORE

Questions	Average Score	Weights	Weighted Score
Question 1	5.6	10%	0.6
Question 2	7.2	20%	1.4
Question 3	7.0	40%	2.8
Question 4	7.0	20%	1.4
Question 5	7.2	10%	0.7
TOTAL	**N/A**	**100%**	**6.9**

Adjustment of initially calculated capital charge on the basis of score card ratings: This is a relatively simple and straight forward step after the calculation of initial capital in step one and estimation of residual risk score in step two. Let us assume that the initial capital for our assumed business unit was estimated at $10,000,000 using Standardized Approach. As the Residual Risk Score of business unit is 6.9, therefore Capital Charge per RRS point can be calculated by dividing $10,000,000 by 6.9 which comes to $1,449,275. This is shown in the following table.

INITIAL CAPITAL CHARGE

Parameters	Amount
Initial Capital Charge for Business Unit	**10,000,000**
Initial Residual Risk Score of Business Unit	6.9
Capital Charge Per RRS Point	1,449,275

Let us further assume that Residual Risk Score of business unit changes to 6.2 in the scorecard exercise of next quarter. As capital charge per point was $1,449,275 which was established in initial exercise, therefore new capital charge can be calculated by multiplying capital charge of per point of $1,449,275 with new residual risk score of 6.2 which will generate new capital charge of $8,985,507. This is shown in the following table.

NEW CAPITAL CHARGE

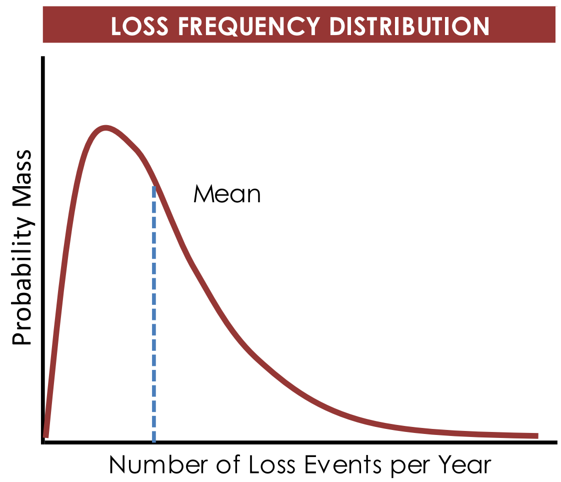
Parameters	Amount
New Residual Risk Score of Business Unit	6.2
Capital Charge Per RRS Point	1,449,275
New Capital Charge for Business Unit	**8,985,507**

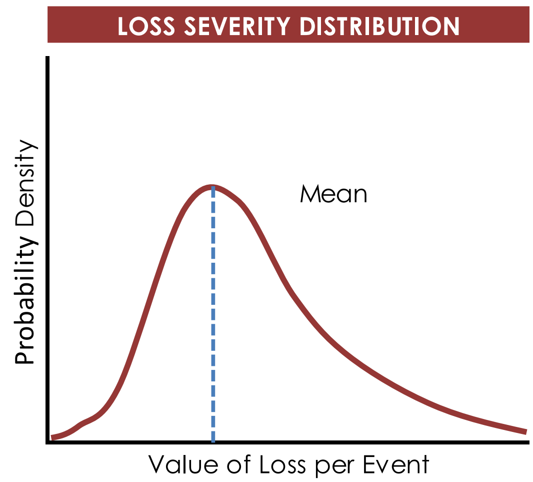
As Score Card Approach combines quantitative as well as qualitative methods to calculate capital charge, the scorecard adjustment reflects the level of quality of control in a specific financial institution. The set of scorecards then leads to an increase or decrease of the capital of the specific institution on the basis of changes in residual risk scores. Financial institutions aim to improve the risk control environment that will reduce both the frequency and severity of future operational risk losses. By identifying a number of risk indicators for particular risk types within business lines, institution can capture the underlying risk profile of various business lines. These risk indicators represents indirectly the altitude of the operational risk.

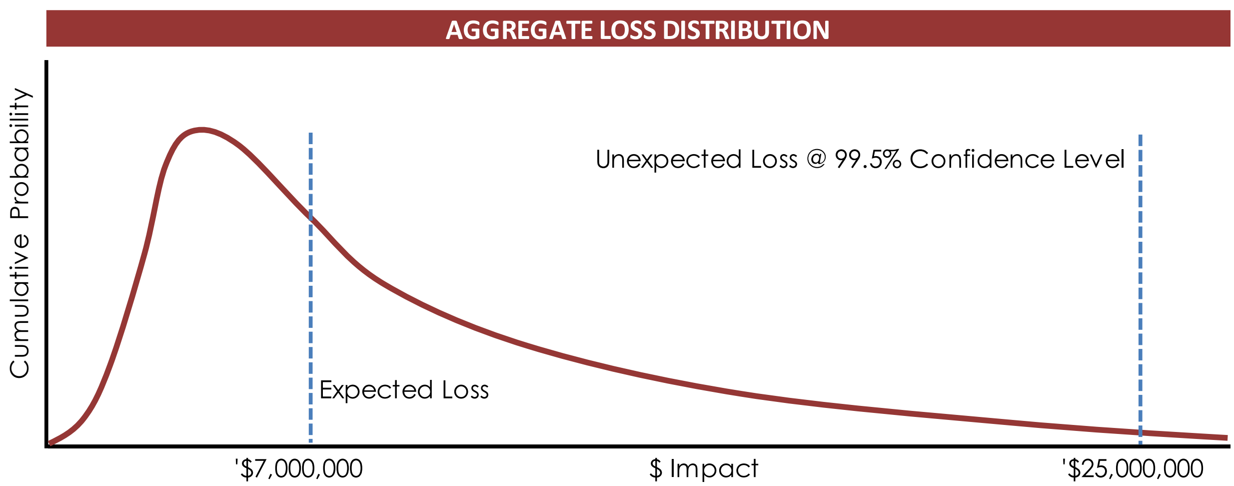
c. Loss Distribution Approach - LDA

Loss Distribution Approach is a statistical approach which is very popular in actuarial sciences for computing aggregate loss distributions. This is also the most complicated approach among AMA models and it requires decent amount of quantitative and statistical skills.

The LDA approach involves modeling of Loss Frequency Distribution and the Loss Severity Distribution separately and then combining these distributions via Monte Carlo simulations or other statistical techniques to form an Aggregate Loss Distribution for each loss type and business line combination, for a given time horizon. Capital charge is then estimated by calculating the expected and unexpected losses from Aggregate Loss Distribution.







The 5 sequential steps involved in the capital charge estimation from Loss Distribution Approach are as follows:

- Modeling of Loss Frequency Distribution
- Modeling of Loss Severity Distribution
- Modeling of Aggregate Loss Distribution
- Calculation of Expected and Unexpected Losses
- Calculation of Capital Charge

QUICK REVIEW OF STATISTICAL CONCEPTS - I

What is Random Variable? A random variable does not mean that the values can be anything (a random number). Random variables have a well-defined set of outcomes and well defined probabilities for the occurrence of each outcome. The random refers to the fact that the outcomes happen by chance and it is not known which outcome will occur next.

What are Discrete & Continuous Variables? A random variable is a discrete random variable if it can assume values, which are finite and countable For example, tossing of a die is a random experiment and its outcomes 1, 2, 3, 4, 5 and 6 are discrete random variable. Values of discrete random variable are whole numbers and do not contain any fractional numbers. If a variable can take on any value between two specified values, it is called a continuous variable. For example, value between 1 and 2. Values of continuous random variable contain fractional numbers.

What is Probability? Probability refers to the calculation of the likelihood that a given event will occur. Probability is expressed as a number between 1 and 0. An event with a probability of 1 is considered a certainty and an event with a probability of 0 can is considered an impossibility.

What is Probability Distribution? A probability distribution is the theoretical counterpart to the frequency distribution. A frequency distribution simply shows how many times a certain event occurred. A probability distribution says how many times it should have occurred. Probability distribution is a table, graph or equation that shows all possible values of a random variable with their corresponding probabilities.

What is Discrete Probability Distribution? Discrete probability theory deals with events that occur in countable sample spaces. If a random variable is a discrete variable, its probability distribution is called a discrete probability distribution.

What is Continuous Probability Distribution? Continuous probability theory deals with events that occur in a continuous sample space. If a random variable is a continuous variable, its probability distribution is called a continuous probability distribution.

What is Cumulative Probability Distribution? A cumulative probability refers to the probability that the value of a random variable falls within a specified range.

What is Probability Function? A probability function is a function or equation which assigns probabilities to the values of a random variable. All the probabilities must be between 0 and 1 in a probability function and the sum of the all probabilities must be 1. If these two conditions aren't met, then the function isn't a probability function. There is no requirement that the values of the random variable only be between 0 and 1, only that the probabilities be between 0 and 1.

What is Probability Density Function? As a continuous random variable can take infinite number of values, its continuous probability distribution cannot be expressed in tabular form. Instead, an equation or formula is used to describe a continuous probability distribution. This equation that describes probability function of a continuous random variable is called Probability Density Function.

What is Probability Mass Function? As a discrete random variable can only take finite number of values, its discrete probability distribution can be expressed in tabular form as well as in equation form. The equation that describes probability function of a discrete variable is called Probability Mass Function. It is a function whose domain contains the set of discrete values that the random variable can assume, with the probabilities of the random variable assuming the values in the domain as its range.

1. Modeling of Loss Frequency Distribution

Frequency refers to the number of times an operational risk event has occurred during past. A minimum history of at least 3 years of frequency data is required for loss frequency modeling. A Frequency Distribution is a representation in a graphical form which displays number of times an event has occurred within a given interval over a time horizon. Loss Frequency Distribution is composed of Discrete Values which means its data will not contain any fractional numbers.

Loss Frequency Distribution is modeled in 2 stages. In first stage a graph is constructed by using internal historic data of risk event occurrence with x-axis showing the intervals of time horizon and y-axis showing the number of risk events during those intervals.

The steps in preparing frequency distributions manually are as follows:

- Take cleaned data of event frequency record. If data are nominal, list the classes into which a data point might fall. If data are interval, select an appropriate number of data classes.

- Calculate the absolute frequency of each class, i.e. the raw number of data points in each class. The sum of all absolute frequencies must be equal the sample size.

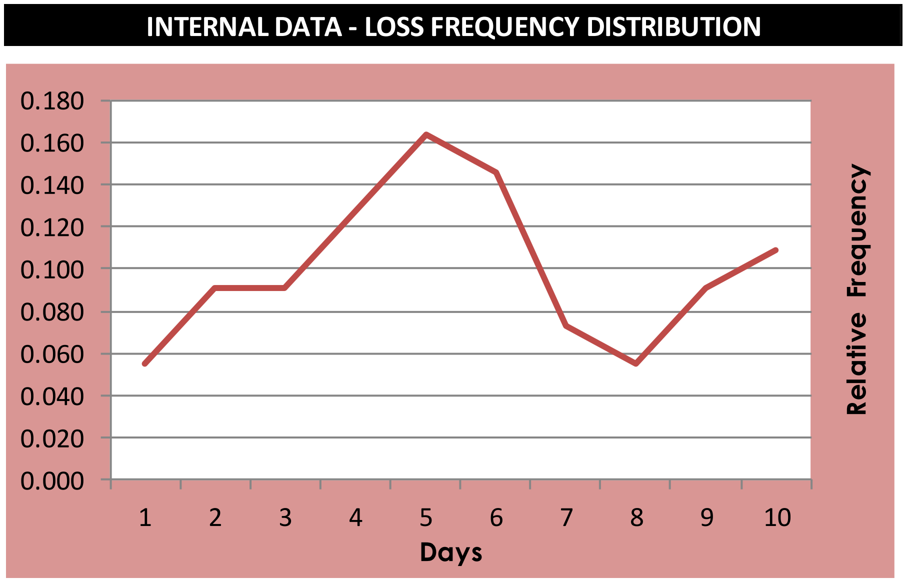
- Calculate the relative frequency by dividing the absolute frequency of each class by the sample size. This reveals the proportion or percent of data points in each class. The sum of all relative frequencies must be 1.

- Calculate the cumulative frequency for each class by adding the number or proportion/percentage of data points in that class to similar quantities for all preceding classes. When number of data points is accumulated, the last number should equal the sample size and when proportions/percentages are accumulated, the last number should be 1 or 100 percent. Given below are the tables for absolute, relative and cumulative frequency calculations together with frequency mean and its standard deviation together with line graph of relative frequency of loss events:

FREQUENCY CALCULATIONS

Days	Loss Events	Absolute	Relative	Cumulative
1	3	3	0.056	0.056
2	7	7	0.130	0.185
3	5	5	0.093	0.278
4	7	7	0.130	0.407
5	4	4	0.074	0.481
6	8	8	0.148	0.630
7	4	4	0.074	0.704
8	3	3	0.056	0.759
9	5	5	0.093	0.852
10	8	8	0.148	**1.000**
TOTAL	**54**	**54**	**1.000**	**N/A**

Mean	Standard Deviation
1.00	0.0370



Stage one concludes with the development of relative frequency graph and calculation of frequency's mean and standard deviation. In stage two, frequency data is remodeled on the basis of some comparable statistical distribution pattern. The reasons why it is done is because loss data is not available in sufficient quantities in any financial institution to permit a reasonable assessment of exposure; therefore it is necessary to put in more data points to supplement loss data, in particular for tail events.

These additional data points cannot be punched in randomly into existing data. They need to be generated on the basis of some formula or statistical function. There are a number of statistical functions or formulae that can generate data but the trick is to find a function or formula that uses some parameter of existing data as input and then generate numbers that have characteristics and pattern similar to existing data.

a. Characteristics of Loss Frequency Data

- Events are always in whole numbers. There is no such thing as 1.5 fraud incidents. It would either be 1 fraud or 2 frauds.
- Events occurrences are independent. One event does not increase or decrease the probability of another event.
- Events are countable. Total number of loss events can be a large number but they cannot be infinite.
- Mathematical Average of events occurrence is known. If events are countable as per above point, their mathematical average can also be counted.

The shape of frequency data graph will differ from institution to institution. Graph can be light tailed or heavy tailed, negatively skewed or positively skewed etc. therefore; statistical tests are conducted to determine which particular type of standard distribution function should be used to model data.

Graphical plots are also used to determine whether the data show light-tailed or heavy-tailed behavior, it also shows whether certain data portions can be modeled using the standard empirical distribution and what the possible thresholds for modeling might be, and whether one dataset or cell needs to be divided into and modeled across multiple segments. Popular statistical distributions to model loss frequency include Poisson Distribution & Binomial Distribution.

The seven types of operational risk can be categorized in terms of frequency which represents the number of loss events during a certain time period and severity which represents the impact of the event in terms of financial loss.

The following table, which is based on the results from the Basel Committee (2002), indicates the typical frequency and severity of each risk type that may arise for the operations of a typical financial institution:

FREQUENCY & SEVERITY OF OPERATIONAL LOSSES			
No	**Risk Categories**	**Frequency**	**Severity**
1	Internal Fraud	Low	High
2	External Fraud	High/Medium	Low/Medium
3	Employment Practices & Workplace Safety	Low	Low
4	Clients, Products & Business Practices	Low/Medium	High/Medium
5	Damage to Physical Assets	Low	Low
6	Business Disruption & System Failures	Low	Low
7	Execution, Delivery & Process Management	High	Low

QUICK REVIEW OF STATISTICAL CONCEPTS - II

What is Mean? Mean is an indicator of Central Tendency of data. The arithmetic mean is equal to the sum of the values divided by the number of values. It is represented by the symbol Meu μ.

What is Median? Median is measure of central tendency of data. It is calculated by arranging data in ascending order and then selecting the one in the middle. If the total number of values in data sample is even, then the median is the mean of the two middle numbers.

What is Mode? Mode is the number that appears most often in a set of numbers.

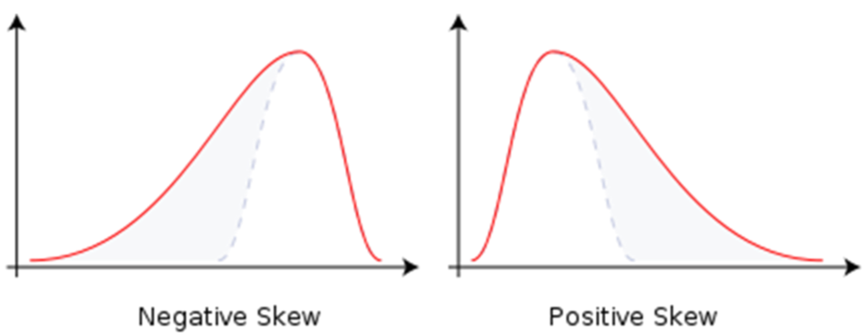
What is Dispersion? Dispersion, also called statistical variability or variation, is variability or spread in the values of a variable or a probability distribution.

What is Variance? Variance is a measure of the average distance between each of a set of data points and their mean value. Mathematically it is calculated as the sum of the squares of the deviation from the mean value.

What is Standard Deviation? Standard Deviation shows the dispersion of data points from data mean. A low standard deviation indicates that the data points are very close to the mean, whereas high standard deviation indicates that the data points are spread out over a large range of values. It is represented by the symbol Sigma σ.

What is Coefficient of Variation? Coefficient of Variation (CV) is a normalized measure of dispersion of a probability distribution. It is also called Unitized Risk or the Variation Coefficient. The coefficient of variation (CV) is calculated as the ratio of the standard deviation to the mean.

What is Skewness? Skewness is a measure of the degree of asymmetry of a distribution. The skewness can be positive or negative or undefined. A negative skew indicates that the tail on the left side of the distribution is longer than the right side and the bulk of the values lie to the right of the mean. A positive skew indicates that the tail on the right side is longer than the left side and the bulk of the values lie to the left of the mean. A zero value indicates that the values are evenly distributed on both sides of the mean.



What is Kurtosis? Kurtosis is the degree of peakedness of a distribution. Kurtosis can be high or low for a distribution. A high kurtosis distribution has a sharper peak and longer, fatter tails, while a low kurtosis distribution has a more rounded peak and shorter, thinner tails. There are 3 types of Kurtosis:

- **Mesokurtic:** When the peak of distribution is neither high nor low, as in Normal Distribution.
- **Leptokurtic:** When the peak of distribution is thin and tall.
- **Platykurtic:** When the peak of distribution has certain flatness and has slender tail.

b. Poisson Distribution

The Poisson distribution is probably the most popular distribution to model loss frequency data within financial industry. Poisson distribution gives the expected frequency profile for event that fulfills the following criteria:

- The event is something that can be counted in whole numbers.
- Occurrences are independent, so that one occurrence neither diminishes nor increases the chance of another.
- The average frequency of occurrence for the time period in question is known.
- It is possible to count how many events have occurred.

Loss frequency data fulfills all of the above conditions. The formula or the function of Poisson distribution is as follows:

Numeric Formula: P(k) = r^k / (k!)(e^r)

Where,
P = Probability
r = Mean or Average rate of occurrence (Also known as expected value or Lambda λ)
k = Expected number of events
k! = Factorial of k. (Product of all positive integers less than or equal to k)
e = Base of natural logarithm (Constant 2.71828)

It is clear from the above formula that the only variable required to construct a Poisson distribution is average rate of occurrence. As long as mean of loss frequency is known, Poisson distribution can be constructed showing the event occurrence probability of any number of events. Given below is a table that shows Poisson probabilities P(k) for number of events (k) where the mean of loss frequency data is 2.

POISSON PROBABILITY DISTRIBUTION

k	A. r^k	k!	e^r	B. ki (e^r)	P(k) = A/B
0	1	1	7.38905	7	0.13534
1	2	1	7.38905	7	0.27067
2	4	2	7.38905	15	0.27067
3	8	6	7.38905	44	0.18045
4	16	24	7.38905	177	0.09022
5	32	120	7.38905	887	0.03609
6	64	720	7.38905	5320	0.01203
7	128	5040	7.38905	37241	0.00344
8	256	40320	7.38905	297926	0.00086
9	512	362880	7.38905	2681337	0.00019
10	1024	3628800	7.38905	26813371	0.00004
TOTAL	**N/A**	**N/A**	**N/A**	**N/A**	**0.99999**

As per table, probability that no loss event happens is 0.13534, probability that 1 event happens is 0.27067 and so on.

The Poisson distribution has 2 distinct qualities i.e. it is not symmetrical and is skewed toward the infinity end, and the mean of any Poisson distribution is equal to its variance.

c. Binomial Distribution

The Binomial distribution gives the expected frequency for event that fulfills the following criteria:

- The experiment consists of n repeated trials.
- Each trial can result in just two possible outcomes. One of these outcomes can be called a success and the other, a failure.
- The probability of success, denoted by P, is the same on every trial.
- The trials are independent; that is, the outcome on one trial does not affect the outcome on other trials.

The formula or the function of Binomial distribution is as follows:

$$f(k; n, p) = \Pr(K = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

For k = 0,1,2,3,4,5……n, and

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Where,
Pr = Probability
n = Total number of events
k = Expected number of events
p = maximum likelihood (p = k/n)

Given below is a table that shows Binomial probabilities Pr(k) of 0 to 10 loss events (k) where total number of events (n) is 10.

BINOMIAL PROBABILITY DISTRIBUTION											
k	n	n-k	k!	n!	(n-k)!	p	1-p	p^k	A. 'n!/k!(n-k)!	B. p^k(1-p)^n-k	Pr = A x B
0	10	10	1	3628800	3628800	0.1	0.9	1.00000	1.0	0.34868	0.3486784401
1	10	9	1	3628800	362880	0.1	0.9	0.10000	10.0	0.03874	0.3874204890
2	10	8	2	3628800	40320	0.1	0.9	0.01000	45.0	0.00430	0.1937102445
3	10	7	6	3628800	5040	0.1	0.9	0.00100	120.0	0.00048	0.0573956280
4	10	6	24	3628800	720	0.1	0.9	0.00010	210.0	0.00005	0.0111602610
5	10	5	120	3628800	120	0.1	0.9	0.00001	252.0	0.00001	0.0014880348
6	10	4	720	3628800	24	0.1	0.9	0.00000	210.0	0.00000	0.0001377810
7	10	3	5040	3628800	6	0.1	0.9	0.00000	120.0	0.00000	0.0000087480
8	10	2	40320	3628800	2	0.1	0.9	0.00000	45.0	0.00000	0.0000003645
9	10	1	362880	3628800	1	0.1	0.9	0.00000	10.0	0.00000	0.0000000090
10	10	0	3628800	3628800	1	0.1	0.9	0.00000	1.0	0.00000	0.0000000001
										TOTAL	**1.00000**

As per table, probability that no loss event in happens in 10 events is 0.3486, probability that 1 loss event happens in 10 events is 0.3874 and so on. The sum of all probabilities in the sample size is 1 as per standard requirement. The mean of a Binomial Distribution is greater than its variance and the mean is calculated as n multiplied by p.

2. Modeling of Loss Severity Distribution

Severity refers to the financial impact of an operational risk event when it occurs. Severity modeling is quite a difficult task. One main reason is the lack of data. Internal loss data also has some inherent weaknesses as a foundation for risk exposure measurement, including:

- Loss data is technically backward-looking measure, and it does not capture changes to the risk and control environment of present.

- Loss data is not available in sufficient quantities in any financial institution to permit a reasonably accurate quantification of exposure, particularly in terms of quantifying the risk of extreme losses.

Internal loss data covering the last 5 to 7 years is usually not sufficient for calibrating tails of severity distributions. Tails of severity distributions represents loss events with extremely loss probability but extremely high severity. It is obvious that additional data sources like external loss data and scenarios are needed to improve the reliability of the model. However, inclusion of this type of information immediately leads to additional problems, e.g. scaling of external loss data, combining data from different sources, etc.

Even if all of the available data sources are used it is necessary to extrapolate beyond the highest relevant losses in the data base. The standard technique is to fit a parametric statistical distribution to the available data and to assume that its parametric shape will provide at least a near realistic model for potential losses beyond the current loss experience. The choice of the statistical distribution is a not an easy task and it usually has a significant impact on model results.

Sometimes it is not possible to identify a standard statistical distribution that provides reasonable fits to the loss data across the entire range. The only solution to this problem is to use different statistical distribution assumptions for the body and the tail of these severity distributions. However, this strategy adds yet another layer of complexity to severity modeling.

When internal data shows light-tailed behavior, the Beta, Chi-square, Exponential, Gamma, Inverse Gaussian, Log Normal, Normal, Weibull and Rayleigh distributions are considered for severity modeling. If internal data shows heavy-tailed behavior, the Burr, Cauchy, F-, Generalized Pareto, Generalized Extreme Value, Log Gamma, Log Logistic, Pareto and Student's t-distributions are used for severity modeling.

Once a standard statistical distribution is selected in line with data's tail behavior, various statistical tests are conducted to evaluate Goodness of Fit (GOF) to ascertain the appropriateness of selected statistical distribution. The most commonly used tests are Kolmogorov-Smirnov, Cramer von Mises, Anderson-Darling, Analysis of Fit Differences, Evaluation PP, Evaluation QQ, Chi-square Tests and Mean Square Error Estimates.

Apart from statistical tests, a number of graphical tests are also used to supplement the GOF tests. These include Probability-Differences Plots, Probability-Probability (PP) Plots and Quantile-Quantile (QQ) Plots. For QQ plots, Linear Scale QQ Plots, Logarithmic Scale QQ plots, Relative Error Plots and Absolute Error Plots.

The final decision is made for the selection of most suitable statistical distribution after all the graphical and non-graphical GOF measures. And finally, Loss Severity Distribution is generated as the result of combining of the actual distribution of the low severity distribution portion created by internal loss data, and the selected standard statistical loss distribution for the high severity distribution portion created by scenario data.

QUICK REVIEW OF STATISTICAL CONCEPTS - III

What is Goodness of Fit Test? Goodness of Fit (GOF) tests measures the compatibility of internal data distribution with selected standard statistical distribution. In simple words, these tests show how well the distribution you selected fits to your data.

What is Chi-Square Test? Chi-Square is a goodness of fit test for discrete distributions and is used to test validity of selected frequency distribution for modeling loss frequency distribution. Chi-Square test is applicable to any discrete value distribution, thus it is ideal for loss frequency distributions. The formula for Chi-square calculation is as follows:

Chi-Square = ∑ Square of (Observed Frequency – Expected Frequency) / Expected Frequency

Where,
Expected Frequency = Frequency of internal data
Observed Frequency = Frequency of selected distribution

The idea is that if the observed frequency is really close to the expected frequency, then the square of the deviations will be small. If the sum of these weighted squared deviations is small, the observed frequencies are close to the expected frequencies. A difference of 10 may be very significant if 12 was the expected frequency, but a difference of 10 isn't very significant at all if the expected frequency was 1000.

What is Kolmogorov-Smirnov Test? Kolmogorov Smirnov test is a goodness of fit test for continuous distributions and is used to test validity of selected probability distribution for modeling loss severity distribution.

In the first step of KS test, cumulative probability of internal data is calculated. In second step cumulative probability of selected distribution is calculated. Then the difference between these cumulative probabilities is measured and the greatest difference is called D-statistic. A high D-statistic value means both data sets are compatible. Given below is the statistical formula:

$$K = D\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

And,

$$D = |cp1_i - cp2_i|_{max}$$

Where,
Cp1 = Cumulative probability of internal data distribution
Cp2 = Cumulative probability of selected distribution

The Kolmogorov-Smirnov one-sample test is similar to the chi-square test of goodness of fit. The main difference is though that in a Kolmogorov-Smirnov test the observed and expected distributions are expressed as proportions rather than frequency counts and are converted to cumulative distributions before comparisons are made.

a. Normal Distribution
The normal distribution is defined by the following equation which is called Normal Equation:

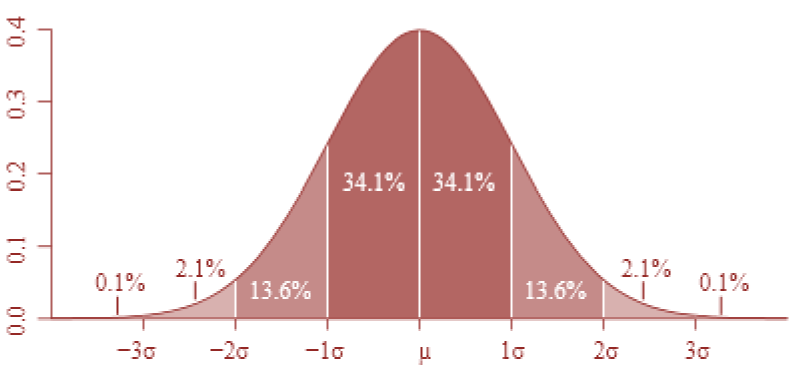
$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

Where,
x = Normal random variable
μ = Mean
σ = Standard deviation
π = 3.14159 (Constant)
e = 2.71828 (Constant)

The random variable X in the normal equation is called the normal random variable. The normal equation is the probability density function for the normal distribution. The graph of the normal distribution depends on two factors - the mean and the standard deviation. The mean of the distribution determines the location of the center of the graph, and the standard deviation determines the height and width of the graph. When the standard deviation is large, the curve is short and wide; when the standard deviation is small, the curve is tall and narrow. All normal distributions look like symmetric and bell-shaped curves.

Every normal distribution curve conforms to the certain rules regardless of its mean or standard deviation. These rules are collectively known as the empirical rule or the 68-95-99.7 rule.

- About 68% of the area under the curve falls within 1 standard deviation of the mean.
- About 95% of the area under the curve falls within 2 standard deviations of the mean.
- About 99.7% of the area under the curve falls within 3 standard deviations of the mean.



b. How to Calculate Probability of a Value under Normal Distribution?
Assuming that loss severity data has a mean of $100 and standard deviation of 9, what is the probability of a $120 loss? In order to find the answer, a simple calculation of Z-score is conducted.

Z-score = (Expected Loss – Mean Loss) / Standard Deviation
Z-score = (120 – 100)/9
Z-score = 2.22

Standard Z-score table is consulted to find the probability of $120 loss which has Z-score of 2.22. Z-score table is a matrix where whole numbers are given on Y axis and decimals are given on X axis. The table gives the probability of $120 loss as 0.9783 or 97.83%

POSITIVE Z SCORE VALUES FOR NORMAL DISTRIBUTION										
Z	**0**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0	0.50000	0.50400	0.50800	0.51200	0.51600	0.51990	0.52390	0.52790	0.53190	0.53590
0.1	0.53980	0.54380	0.54780	0.55170	0.55570	0.55960	0.56360	0.56750	0.57140	0.57530
0.2	0.57930	0.58320	0.58710	0.59100	0.59480	0.59870	0.60260	0.60640	0.61030	0.61410
0.3	0.61790	0.62170	0.62550	0.62930	0.63310	0.63680	0.64060	0.64430	0.64800	0.65170
0.4	0.65540	0.65910	0.66280	0.66640	0.67000	0.67360	0.67720	0.68080	0.68440	0.68790
0.5	0.69150	0.69500	0.69850	0.70190	0.70540	0.70880	0.71230	0.71570	0.71900	0.72240
0.6	0.72570	0.72910	0.73240	0.73570	0.73890	0.74220	0.74540	0.74860	0.75170	0.75490
0.7	0.75800	0.76110	0.76420	0.76730	0.77040	0.77340	0.77640	0.77940	0.78230	0.78520
0.8	0.78810	0.79100	0.79390	0.79670	0.79950	0.80230	0.80510	0.80780	0.81060	0.81330
0.9	0.81590	0.81860	0.82120	0.82380	0.82640	0.82890	0.83150	0.83400	0.83650	0.83890
1	0.84130	0.84380	0.84610	0.84850	0.85080	0.85310	0.85540	0.85770	0.85990	0.86210
1.1	0.86430	0.86650	0.86860	0.87080	0.87290	0.87490	0.87700	0.87900	0.88100	0.88300
1.2	0.88490	0.88690	0.88880	0.89070	0.89250	0.89440	0.89620	0.89800	0.89970	0.90150
1.3	0.90320	0.90490	0.90660	0.90820	0.90990	0.91150	0.91310	0.91470	0.91620	0.91770
1.4	0.91920	0.92070	0.92220	0.92360	0.92510	0.92650	0.92790	0.92920	0.93060	0.93190
1.5	0.93320	0.93450	0.93570	0.93700	0.93820	0.93940	0.94060	0.94180	0.94290	0.94410
1.6	0.94520	0.94630	0.94740	0.94840	0.94950	0.95050	0.95150	0.95250	0.95350	0.95450
1.7	0.95540	0.95640	0.95730	0.95820	0.95910	0.95990	0.96080	0.96160	0.96250	0.96330
1.8	0.96410	0.96490	0.96560	0.96640	0.96710	0.96780	0.96860	0.96930	0.96990	0.97060
1.9	0.97130	0.97190	0.97260	0.97320	0.97380	0.97440	0.97500	0.97560	0.97610	0.97670
2	0.97720	0.97780	0.97830	0.97880	0.97930	0.97980	0.98030	0.98080	0.98120	0.98170
2.1	0.98210	0.98260	0.98300	0.98340	0.98380	0.98420	0.98460	0.98500	0.98540	0.98570
2.2	0.98610	0.98640	0.98680	0.98710	0.98750	0.98780	0.98810	0.98840	0.98870	0.98900
2.3	0.98930	0.98960	0.98980	0.99010	0.99040	0.99060	0.99090	0.99110	0.99130	0.99160
2.4	0.99180	0.99200	0.99220	0.99250	0.99270	0.99290	0.99310	0.99320	0.99340	0.99360
2.5	0.99380	0.99400	0.99410	0.99430	0.99450	0.99460	0.99480	0.99490	0.99510	0.99520
2.6	0.99530	0.99550	0.99560	0.99570	0.99590	0.99600	0.99610	0.99620	0.99630	0.99640
2.7	0.99650	0.99660	0.99670	0.99680	0.99690	0.99700	0.99710	0.99720	0.99730	0.99740
2.8	0.99740	0.99750	0.99760	0.99770	0.99770	0.99780	0.99790	0.99790	0.99800	0.99810
2.9	0.99810	0.99820	0.99820	0.99830	0.99840	0.99840	0.99850	0.99850	0.99860	0.99860
3	0.99870	0.99870	0.99870	0.99880	0.99880	0.99890	0.99890	0.99890	0.99900	0.99900
3.1	0.99900	0.99910	0.99910	0.99910	0.99920	0.99920	0.99920	0.99920	0.99930	0.99930
3.2	0.99930	0.99930	0.99940	0.99940	0.99940	0.99940	0.99940	0.99950	0.99950	0.99950
3.3	0.99950	0.99950	0.99950	0.99960	0.99960	0.99960	0.99960	0.99960	0.99960	0.99970
3.4	0.99970	0.99970	0.99970	0.99970	0.99970	0.99970	0.99970	0.99970	0.99970	0.99980
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997
4	0.99997	0.99997	0.99997	0.99997	0.99997	0.99997	0.99998	0.99998	0.99998	0.99998

For Z-score value of 2.225:

- Find 2 in first column.
- Find 0.2 in first row.
- Find the intersection of 2 and 0.2.
- This intersection will show the probability of the loss value with z-score of 2.225.

NEGATIVE Z SCORE VALUES FOR NORMAL DISTRIBUTION										
Z	**0**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
-4	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00103	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2	0.02275	0.02222	0.02169	0.02118	0.02067	0.02018	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.7	0.04456	0.04363	0.04272	0.04181	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
-1.5	0.06681	0.06552	0.06425	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
-1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07214	0.07078	0.06944	0.06811
-1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08691	0.08534	0.08379	0.08226
-1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09852
-1.1	0.13566	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
-1	0.15865	0.15625	0.15386	0.15150	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
-0.9	0.18406	0.18141	0.17878	0.17618	0.17361	0.17105	0.16853	0.16602	0.16354	0.16109
-0.8	0.21185	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
-0.7	0.24196	0.23885	0.23576	0.23269	0.22965	0.22663	0.22363	0.22065	0.21769	0.21476
-0.6	0.27425	0.27093	0.26763	0.26434	0.26108	0.25784	0.25462	0.25143	0.24825	0.24509
-0.5	0.30853	0.30502	0.30153	0.29805	0.29460	0.29116	0.28774	0.28434	0.28095	0.27759
-0.4	0.34457	0.34090	0.33724	0.33359	0.32997	0.32635	0.32276	0.31917	0.31561	0.31206
-0.3	0.38209	0.37828	0.37448	0.37070	0.36692	0.36317	0.35942	0.35569	0.35197	0.34826
-0.2	0.42074	0.41683	0.41293	0.40904	0.40516	0.40129	0.39743	0.39358	0.38974	0.38590
-0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43250	0.42857	0.42465
0	0.50000	0.49601	0.49202	0.48803	0.48404	0.48006	0.47607	0.47209	0.46811	0.46414

Z-score basically shows the percentage of data between 0 and Z-score value. For example a Z-score of 0.45 has a value of 0.1736 which is equal to 17.36%. Therefore, 17.36% of the data population is between 0 and 0.45 standard deviations from the mean. Because the curve is symmetrical, so if Z-score has a negative value of 0.45, it also means that 17.36% of the data population is between 0 and - 0.45 standard deviations from the mean.

c. How to Generate Values under Normal Distribution?
Normal distribution of severity values can be generated by using a number of algorithms. Simplest way to model normal distribution of severity is to use Excel built-in function or applying a formula. Parameters that are required to generate normal distribution of severity values include the following:

- Mean: This refers to the mean of internal severity data.
- Standard Deviation: This refers to standard deviation of internal severity data.
- Z Maximum: This refers to maximum value of Z which is taken as 4.
- Z Minimum: This refers to minimum value of Z which is taken as -4.
- Data Points: This refers to number of values to be generated under normal distribution.

Descriptive and numeric formulae for normal distribution in Excel are under:

Descriptive Formula: = NORM.DIST (X, mean, standard deviation, 0)

Where,
NORM.DIST = Excel built-in function for normal distribution.
X = Severity value for which probability is to be calculated.
Mean = Mean of the internal severity data.
Standard Deviation = Standard deviation of internal severity data.
0 = In Excel function, 0 refers to fact that non-cumulative probability is being calculated. In case of cumulative probability 0 is replaced by 1 in the formula.

Numeric Formula: = NORM.DIST(X,μ,σ,0)

Where,
NORM.DIST = Excel built-in function for normal distribution.
X = Severity value.
μ = Mean.
σ = Standard Deviation.
0 = Non-cumulative probability calculation.

The equivalent of above built-in function is the following manual formula which will generate the same results:

Manual Formula: = EXP(-1/2*X^2)/SQRT(2*PI())

Assuming that internal severity data has 2 parameters i.e. mean of 1000, and standard deviation of 100, a normal distribution can be generated by using above Excel function. The number of values to be generated should be enough to reach 99.99% cumulative probability.

This will require a table of 4 columns.

- **Column 1 – No:** This column shows the count of number of values.
- **Column 2 – Z:** This column shows the values of Z. First value is the Z-Minimum which -4. Other values are calculated as: Z = (Z maximum value – Z minimum value) / (Total numbers of values to be generated x Standard deviation).
- **Column 3 – X:** This column shows the calculated value of severity under normal distribution. It is calculated as: X = Mean + (Z value in column 2 x Standard deviation).
- **Column 4 – Probability:** This column shows the probability of the loss calculated in column 3 using Excel built-in function. It is calculated as: P = NORM.DIST(X,μ,σ,0).

Using above assumed data, given below is the table that shows a cutout of normal distribution of loss severity values where total number of severity losses are assumed 80. This is obviously not enough to generate a cumulative probability of 99.99% as 80 losses are giving a cumulative probability of 9.99% only.

SEVERITY NORMAL DISTRIBUTION				
No	Z	X	Probability	Cumululative
1	-4	600	0.00000	0.00000
2	-3.9	610	0.00000	0.00000
3	-3.8	620	0.00000	0.00001
4	-3.7	630	0.00000	0.00001
5	-3.6	640	0.00001	0.00002
6	-3.5	650	0.00001	0.00003
7	-3.4	660	0.00001	0.00004
8	-3.3	670	0.00002	0.00005
9	-3.2	680	0.00002	0.00008
10	-3.1	690	0.00003	0.00011
11	-3	700	0.00004	0.00016
12	-2.9	710	0.00006	0.00022
13	-2.8	720	0.00008	0.00029
14	-2.7	730	0.00010	0.00040
15	-2.6	740	0.00014	0.00053
16	-2.5	750	0.00018	0.00071
17	-2.4	760	0.00022	0.00093
18	-2.3	770	0.00028	0.00122
19	-2.2	780	0.00035	0.00157
20	-2.1	790	0.00044	0.00201
21	-2	800	0.00054	0.00255
22	-1.9	810	0.00066	0.00321
23	-1.8	820	0.00079	0.00400
24	-1.7	830	0.00094	0.00494
25	-1.6	840	0.00111	0.00605
26	-1.5	850	0.00130	0.00734
27	-1.4	860	0.00150	0.00884
28	-1.3	870	0.00171	0.01055
29	-1.2	880	0.00194	0.01249
30	-1.1	890	0.00218	0.01467
31	-1	900	0.00242	0.01709
32	-0.9	910	0.00266	0.01975
33	-0.8	920	0.00290	0.02265
34	-0.7	930	0.00312	0.02577
35	-0.6	940	0.00333	0.02911

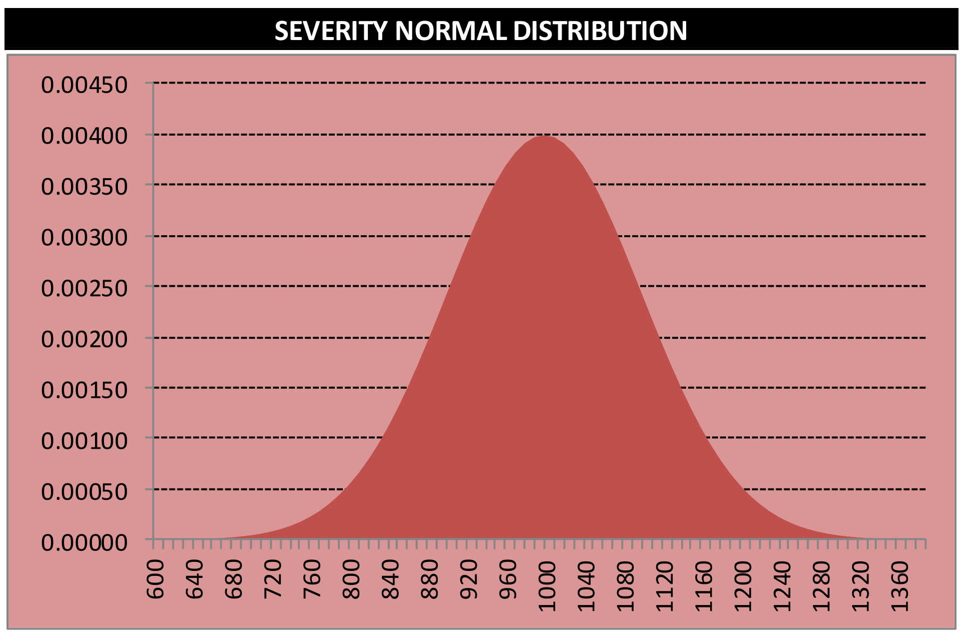
Excel table needs to be stretched up to the level where cumulative probability reaches 99.99%.

d. How to Make a Bell Curve for Normal Distribution in Excel?

Mapping of above results on a graph will result in a bell shape curve showing probability on y-axis and loss severity values on x-axis. The bell curve on Excel is made by following below mentioned simple steps:

- Go to Insert on main menu of Excel.
- Click on Area graph type.
- Click on Select Data.
- Click Add on Legend Entries.
- Give range of probability column in Series Value.
- Click Ok.
- Click Edit on Horizontal Axis Labels.
- Give range of X column showing loss severity values.
- Click Ok.

The following graph will be made, showing loss severity values on x-axis and probability of values of y-axis:



3. Modeling of Aggregate Loss Distribution

Once frequency and severity distributions are modeled, the next step is to model aggregate loss distribution. Aggregate loss is estimated by combining frequency and severity distributions. As frequency is a discrete distribution while severity is a continuous distribution, therefore frequency is converted into continuous probability during the process. Event categories are assumed to be independent of each other; therefore, one simulation per risk category for each business unit needs to be calculated. Therefore, this process is done for every risk category within each business line.

In order to gauge the soundness of this process, each modeled risk is reviewed and analyzed for its reasonableness in terms of matching average loss of aggregated distribution with actual data and comparing 99.9% confidence levels with worst historic cases for similar businesses and risk event types. There are two commonly used ways to convolute/combine frequency and severity distributions i.e. simulation method and tabulation method.

a. Simulation Method

The process of combining frequency and severity through simulation methodology has following steps:

- A random draw is taken from frequency distribution which represents n number of loss events per year.
- Random draws are taken from the severity distribution. Number of random draws is equal to n number of loss events in the first step. These losses are represented by L1, L2, …Ln.
- Sum of n number of losses is taken to obtain an annual loss X which is equal to L1 + L2 + …+ Ln.
- These3 steps are repeated several thousand times in order to obtain X1, …..XN where N is a very large number.
- A histogram of X1, …..XN is formed which represents the annual loss distribution.

This process takes a lot of time if done manually, therefore standard simulation models e.g. Monte Carlo are used to form loss distributions based on above steps.

b. Monte Carlo Simulation

Monte Carlo simulation is the most common method which is used to generate an aggregate loss distribution from the frequency and severity distributions. To start, the MC simulation randomly chooses an annual number of events from the frequency distribution. The most likely choice will always be equal to the mean, and the further a number is away from the mean, the less likely it is that the MC process will chose this number. This randomly selected number is the frequency for that iteration. The frequency is then used as the number of draws that the MC simulation selects from the severity distribution. Each of these draws from the severity distribution represents a loss event. All these drawn loss amounts are summed to create the aggregate annual loss amount. This process is repeated until the desired number of iterations is run. The aggregate loss amounts from each iteration are sorted from low to high. The average of all the results is the mean of the aggregate loss distribution.

Once the parameters for all the different risk categories are calculated, the combined Monte Carlo simulation is used to generate a total aggregate loss distribution for the business unit. During this process an aggregate loss distribution is calculated for each of the event categories, using the single Monte Carlo simulation. During the simulation process, the loss amounts generated by the iterations are added together to create the amount of the combined distribution. Once all the iterations are complete, the mean of the combined distribution can be calculated by taking the average of the total amounts from the iterations. The amounts at different percentiles are determined using the same method used by the single Monte Carlo simulation process.

c. Tabulation Method

LOSS DATA

Frequency	Probability	Severity	Probability
0	0.6	1,000	0.5
1	0.3	10,000	0.3
2	0.1	100,000	0.2

LOSS TABULATION

No. of Losses	1st Loss	2nd Loss	Total Loss	Probability
0	0	0	0	0.6
1	1,000	0	1,000	0.15
1	10,000	0	10,000	0.09
1	100,000	0	100,000	0.06
2	1,000	1,000	2,000	0.025
2	1,000	10,000	11,000	0.015
2	1,000	100,000	101,000	0.010
2	10,000	1,000	11,000	0.015
2	10,000	10,000	20,000	0.009
2	10,000	100,000	110,000	0.006
2	100,000	1,000	101,000	0.010
2	100,000	10,000	110,000	0.006
2	100,000	100,000	200,000	0.004
Total				**1.00**

LOSS AGGREGATION

Total Loss	Cumulative Probability
0	0.600
1,000	0.750
10,000	0.840
100,000	0.900
2,000	0.925
11,000	0.940
101,000	0.950
11,000	0.965
20,000	0.974
110,000	0.980
101,000	0.990
110,000	0.996
200,000	1.000

4. Calculation of Expected and Unexpected Losses

5. Estimation of Capital Charge

Data used and classification
Internal loss data and risk scenarios are the two inputs into the capital calculation model. Loss data are historically suffered losses, while risk scenarios are prospective risk exposure estimates. Each risk scenario is quantified by experts who specify loss amounts at specific probability (or frequency) levels. Loss data and risk scenarios are classified in a matrix, where the vertical axis is typically business lines, while risk classes are depicted on the horizontal axis.

Data and extreme value theory analysis
Internal loss data exploration and analysis are an essential step in the overall modeling process and need to be performed before analytical modeling of available internal loss data can be performed.

Tabular and graphical data analysis provides the modeler with an indication of data completeness, spread, classification, patterns, breaks and possible compatibility with certain analytical model families. Typical tools that are utilized are summary tables, regulatory data matrices, multidimensional histograms and empirical distribution representations.

More formalized statistical tests are used to determine which family of distributions may be a possible fit for the data over various logical segments (specific reference is made here to light-tailed and fat-tailed theoretical distributions). These tests also help to determine the most appropriate truncation points and thresholds for modeling data in a single cell.

Some of the graphical plots that are used to determine the applicability of using extreme value theory (and light-tailed versus heavy-tailed distributions in general) are mean excess plots, Hill estimator plots, HKKP-Hill plots, DEdH plots, tail plots and stability parameter plots. These plots help to determine whether the data show light-tailed or heavy-tailed behavior or both (in different segments), whether certain data segments can be modeled using the empirical distribution and what the possible thresholds for modeling might be, and therefore whether one dataset or cell needs to be divided into and modeled across multiple segments.

The basic data and extreme value theory analysis also assists in determining the point at which risk scenarios should be incorporated into the models. This is typically done at a point where observations are very scarce and business areas are exposed to high severity events.

It is important to note that scenarios are incorporated at a threshold that corresponds to an identified modeling segment for a specific cell, or from an additional threshold created specifically to facilitate the incorporation of risk scenarios into the capital model.
Once the thresholds have been determined, as well as the type of distributions that may be applicable, analytical modeling of the underlying loss data and scenarios can be performed.

Modeling of internal loss data
In segments where light-tailed behavior is observed, the beta, chi-square, exponential, gamma, inverse Gaussian, log normal, normal, Weibull and Rayleigh distributions are usually considered for severity modeling. In segments where heavy-tailed behavior is observed, the Burr, Cauchy, F-,

generalized Pareto, generalized extreme value, log gamma, log logistic, Pareto and Student's t-distributions are tested for severity modeling.

Any of five methods of distribution fitting can be used, and in many cases more than one method is applied for a specific distribution, since they may yield different results. The methods used include the maximum likelihood estimation, least squares method, probability weighted least squares method, robust least squares method and the method of moments (for frequency models only).

Once a series of fits have been performed, various non-graphical goodness of fit (GOF) measures are used to evaluate the accuracy and appropriateness of each fit. The most commonly used tests are Kolmogorov-Smirnov, Cramer von Mises, Anderson-Darling, analysis of fit differences, evaluation PP, evaluation QQ, chi-square tests and mean square error estimates.

A number of graphical representations are also used to supplement the GOF measures. These include probability-differences plots, probability-probability (PP) plots and quantile-quantile (QQ) plots. For QQ plots, linear scale QQ plots, logarithmic scale QQ plots, relative error plots (for specific quantiles) and absolute error plots (for specific quantiles) are evaluated.

Based on all the graphical and non-graphical GOF measures, a decision is made on the most suitable severity distribution for the data segment under consideration for a specific cell.

When performing frequency modeling for a segment where a corresponding severity model exists, tests are performed for the geometric, negative binomial and Poisson distributions.

The same graphical and non-graphical GOF measures are evaluated for frequency distributions as for severity distributions in order to find the most appropriate and accurate frequency fit for a specific segment in a particular data cell under consideration. However, as a general assumption, the Poisson distribution is used for frequency modeling. While this assumption is well supported by research and literature, the Poisson distribution is also chosen to ensure consistency across all cells and segments, and to enable the integration of internal data and scenario models.

Modeling of risk scenarios

Each individual risk scenario should be quantified (loss estimates) at various probability/frequency levels. In addition, experts also provide an annual loss frequency for each scenario. This information is used to construct an empirical severity cumulative distribution function for each scenario, which consequently can be modeled with an analytical distribution.

For frequency modeling, the annual frequency estimate is assigned as the mean parameter of the Poisson distribution. As discussed, each risk scenario is modeled individually.

Scenarios are consequently aggregated per cell in the classification matrix using Monte Carlo simulation with a high number of iterations. The result is an empirical dataset that contains all possible annual permutations and combinations of scenario realizations. For each cell, this empirical distribution is incorporated into the model from a specified threshold.

Since each point in the empirical distribution represents a combination of losses from various scenarios (annualized), a frequency distribution of Poisson (1) is associated with each empirical severity set. This mean parameter may be adjusted for threshold values. In cases where internal data are also present in the specific cell segment where scenarios will be incorporated, the internal data frequency distribution is set equal to the scenario frequency distribution. This is to ensure stability during simulation.

This frequency setting or equalization determines the value of the threshold from where scenarios are incorporated. The threshold is chosen so that the annual frequency of internal data above the threshold is equal to 1 (or a slightly smaller parameter should adjustment for threshold value be necessary).

Independent simulation and aggregation

Before starting the simulation process, a decision needs to be made on the weights that will be assigned to internal loss data models and scenario models, respectively, during the simulation process for each segment in each cell. These weights determine the percentage of random values that are drawn from loss data models and risk scenario models. The weights are individually specified for each segment in each cell where both an internal loss data model and a risk scenario model were constructed. The weighting of the two input data type models is subjective and is determined by a predefined list of factors.

Multiple segment severity distributions can be used to introduce scenario analysis into the simulation above specific thresholds. Several thresholds, and hence segments in a specific cell, can be defined in order to specify the weight of internal loss models and scenario models per segment, as described earlier. In many cases it makes sense to assign a higher weight in the simulation to scenarios in higher value segments (tails) where internal data are scarcer or less reliable.

The process followed for simulation with multiple segments containing internal loss models and scenario models is the same as when only internal losses are used, except for the added complexity of mixing internal loss distributions and scenario distributions.

Monte Carlo simulation is performed simultaneously across all segments and distributions within a specific cell. For each simulation iteration the total losses across all segments are added up to arrive at an annual aggregate loss for the specific iteration. A large number of iterations are performed to construct a dense annual aggregate loss distribution for each cell.

Value-at-risk (VAR) at the 99.9th percentile is calculated for each cell to arrive at the regulatory capital charge for a specific cell. For the calculation of the Group's (and each business line's) capital charge, all applicable cells' 99.9th percentile values-at-risk are added together. This equates to assuming full dependence between all cells and business lines.

Simulation taking correlation into account

Since the data are classified in a matrix, it is possible that inherent correlations are present between the individual internal loss datasets. These correlations can be taken into account during aggregation (simulation) to derive a diversified economic capital charge under Pillar 2 of the Basel II Framework.

Correlation is estimated based on internal data only. Consequently, the calculated correlation is applied to the whole cell and it is then implicitly assumed that the scenarios also pertaining to these cells have the same correlation characteristics and structures.

The copula calculation and simulation are performed in two steps. First, the aggregate loss distribution for each cell is generated in an independent process with several segments, including internal loss data models and scenario-based models. Second, the empirical distributions resulting from the simulation process are provided with the desired dependencies, tail properties and other distributional properties using copulas.

Copulas are used to model correlation structures. Gaussian and Student's t copulas require a correlation matrix for the simulation process and a tail parameter for the Student's t copula, in order to define the inter-cell dependencies and other distributional properties. A rank correlation matrix is calculated using event dates of the fitted data; therefore, it is only possible to calculate correlation parameters for the cells populated with empirical data (observed internal loss data).

The process followed to construct copulas and create multivariate distributions with marginal distributions correlated via copulas can be summarized in a few steps. The Gaussian copula is used as an example.

1. Generate/construct an empirical aggregate loss distribution for each cell utilizing an independent Monte Carlo simulation procedure.
2. Generate independent normal random numbers (X), which are correlated through the rank correlation matrix, obtaining X*.

Calculate the normal cumulative probabilities φ(X*) in order to recover the arguments of Cρ Ga (υ); and

4. Finally, determine the xi (i.e. the loss of division/loss event type i) by inverting the marginal distributions Fi: xi = Fi^-1(ui).

By iterating this process and summing up the xi losses each time, we trace the whole integrated distribution for each cell and for the Group. Sampling is performed simultaneously across all cells, taking correlation structures into account. The Group's annual aggregate loss distribution will therefore inherently contain all underlying dependencies and correlation structures.

Calculating the VAR at the 99.9th percentile of the annual loss distribution will therefore yield a diversified operational risk capital charge number for the Group, where correlation has been taken into account.

Using insurance as mitigation agent
Insurance can be used as a mitigation instrument when calculating operational risk capital requirements. Insurance is applied to losses generated during the Monte Carlo simulation process. In the case of independent simulation, insurance is applied to losses as they are generated from various distributions. When correlation is taken into account, insurance is applied to losses generated after dependence structures are modeled. However, the principles of insurance application for independent and correlated losses are the same.

In order to incorporate the effect of insurance, available insurance policies and coverage clauses need to be mapped to the modeling structure, i.e. business lines and Basel II loss event type combinations. This is quite a large project that needs to be conducted before insurance data are in a format usable for the capital model.

In addition to the policy and clause mapping to each cell, various insurance properties need to be parameterized for each cell, including maximum coverage, deductible and an indication whether cover is global or per event. It is also important for information on all applicable policies' compliance with Basel II minimum standards to be available. This includes parameters that will be used in haircut (discount) parameter calculations.

All of the above-mentioned parameterized insurance characteristics are consequently applied during the simulation process to each simulated loss in order to arrive at a mitigated aggregate loss distribution where insurance has been taken into account.

Capital allocation

After calculating the Group's total capital charge, the extent to which each business line and loss-event type combination contributes to the overall operational risk profile is estimated. This information will enable risk managers to focus efforts on and prioritize the mitigation of operational risk. In the case of independent simulation, each business line's capital charge is simply the sum of the VAR numbers at the 99.9th percentile across all loss event types.

Where an annual aggregate loss distribution has been constructed for the Group taking correlation structures into account, total capital is allocated based on the marginal contribution of each division/loss-event type combination's unexpected loss (UL) to the Group's unexpected loss. This is done on the principle of marginal UL contribution to the overall risk profile. The normalized contribution of each cell's UL to the Group's UL is also referred to as residual operational risk.

A very important property of this capital allocation methodology is that the sum of the allocated capital numbers equals the total calculated Group capital. Statistical (theoretical) allocations are only made down to business line level in the Group – allocations to lower levels in the organization are done utilizing subjectively compiled risk scorecards.

Bibliography

Basel Committee on Banking Supervision (2002a). Sound Practices for the Management and Supervision of Operational Risk. Bank for International Settlements, July 2002. *Available at www.bis.org.*

Basel Committee on Banking Supervision (2002b). Operational Risk Rules Language. Bank for International Settlements, August 2002. *Available at www.bis.org.*

Basel Committee on Banking Supervision (2006). International Convergence of Capital Measurement and Capital Standards: A Revised Framework – Comprehensive Version. Bank for International Settlements, June 2006. Available at www.bis.org.

Artzner, P., Delbaen F., Eber J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* 9/3, 203-228.