Die approbierte Originalversion dieser Diplom-/Masterarbeit ist an der Hauptbibliothek der Technischen Universität Wien aufgestellt (http://www.ub.tuwien.ac.at).

The approved original version of this diploma or master thesis is available at the main library of the Vienna University of Technology (http://www.ub.tuwien.ac.at/englweb/).

TECHNISCHE UNIVERSITÄT WIEN

VIENNA UNIVERSITY OF TECHNOLOGY

Diplomarbeit

Quantifying of Operational Risk by an Advanced Measurement Approach - Model

Ausgeführt am Institut für
Statistik und Wahrscheinlichkeitstheorie
der Technischen Universität Wien

unter der Anleitung von
Ao. Univ. Prof. DI Dr. techn. Karl Grill und
DI Dr. techn. Andreas Weingessel

durch
Claudia Steiner
Frauenkirchner Str. 54
7141 Podersdorf am See

Datum

Unterschrift

Motivation

During my summer internship in one of the biggest Austrian banks my interest in operational risk was raised. Due to this fact I decided to focus my academic work on this field. Additionally, I was offered the possibility to write my masterthesis in collaboration with this bank.

Acknowledgement

I would like to thank Karl Grill for his support and ongoing supervision. Further, I would like to thank Ilinka Kajgana and Andreas Weingessel for their excellent mentoring throughout the entire mastersthesis. At last, I would like to thank my family and friends for their moral support. I really appreciate that great help in continuous motivation during difficult phases of writing.

Contents

List of abbreviations

AMA	Advanced Measurement Approach
AR	Administration Ratio
BIA	Basic Indicator Approach
BL	Business Line
EL	Expected Loss
EV	Event Type
EVT	Extreme Value Theory
KRI	Key Risk Indicator
KS-test	Kolmogoroff-Smirnoff test
LDA	Loss Distribution Approach
ORX	Operational Risk eXchange
Q-Q Plot	Quantile-Quantile Plot
RA	Risk Assessment
sb-AMA	scenario based - Advanced Measurement Appraoch
TSA	The Standardised Approach
UL	Unexpected Loss
VaR	Value at Risk

1 Introduction

The new requirements of the Basel Comittee on Banking Supervision -Basel II- force banks to incorporate operational risk into their risk management. Therefore operational risk became a popular and booming part of a bank‘s risk management. A few years ago only market risk and credit risk were seen as the main risk drivers, but nowadays, not only due to the new Basel Capital Accord, an increasing awareness of operational risks as substantial part of a bank‘s risk management appeals.

Additionally, Basel II requirements expect from an internationally acitive bank to use an Advanced Measurement Approach - AMA, the most sophisticated approach for modelling operational risks. This requires new analysis toward data consistency and completeness. The Basel II standards does not restrict the wide range of possibilities to model operational risk. Many different approaches and models, for implementing an AMA, have been developed recently. For example models based on extreme value theory - EVT, scenarios - sb-AMA or loss distributions. The model introduced in this work will be based on the Loss Distribution Approach - LDA, which on its part will be mainly driven by scenarios.

Scenarios as addendum to internal or external data sources are not backward-looking and provide a more risk sensitive way of managing and measuring operational risk. Scenarios will be generated by using various data sources, such as internal and/or external data and key risk indicators - KRIs.

The link between LDA and scenarios will be the following: Scenario analysis will be used to generate parameters for the loss severity distribution, which is one part of the compound loss distribution (the other part is the loss frequency distribution).

The next chapter gives an introduction to operational risks with a definition, proposed by Basel II. Some famous cases of operational risk, which are typical, are presented, too.

Chapter three describes the Basel II requirements concerning operational risk. The three approaches, Basic Indicator Approach - BIA, the Standardised Approach - TSA and the Advanced Measurement Approaches - AMA, for measuring and modelling operational risk are presented.

Chapter four gives the statistical framework of the Loss Distribution Approach LDA, which is the foundation of the model, which will be developed in the following chapters five and six, whereas a description of the different data sources with focus to scenarios will be given in chapter five and the final model will be developed in chapter six.

Finally, chapter seven presents future challenges and issues of operational risk, which are not taken under consideration so far.

2 Operational Risk

Globalization and the development of financial technologies forced international banking groups and the Banking Supervision to realize the importance of managing operational risk. Former market risk[1] and credit risk[2], which are still substantial, were seen as the main risk drivers, but nowadays the awareness raises, that operational risk plays an essential role in a bank‘s risk management. It is well known that market and credit risk tend to be isolated in specific areas of the business. On the other hand operational risks are inherent in all business processes [16].

Definition: *Operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk[3], but excludes strategic and reputational risk.*[25]

This definition is given by the Basel Committee on Banking Supervision also known as Basel II, which attempts to make the New Basel Capital Accord more risk sensitive. *Developing banking practices such as securitisation, outsourcing, specialised processing operations and reliance on rapidly evolving technology and complex financial products and strategies suggest that these other (remark different from credit and market risk) risks are increasingly important factors to be reflected in credible capital assessments by both supervisors and banks.*[23]

[1]Market risk is the risk that the value of an investment will decrease due to moves in market factors.

[2]Credit risk is risk due to uncertainty in a counterparty's (also called an obligor's or credit's) ability to meet its obligations. Because there are many types of counterparties - from individuals to sovereign governments - and many different types of obligations - from auto loans to derivative transactions - credit risk takes many forms .

[3]Legal risk includes, but is not limited to, exposure to fines, penalties, or punitive damages resulting from supervisory actions, as well as private settlements.

The four main operational risk categories are further clarified as follows:

People: Losses associated with intentional violation of internal policies by current or past employees and the risk extends to people who are working for the bank.

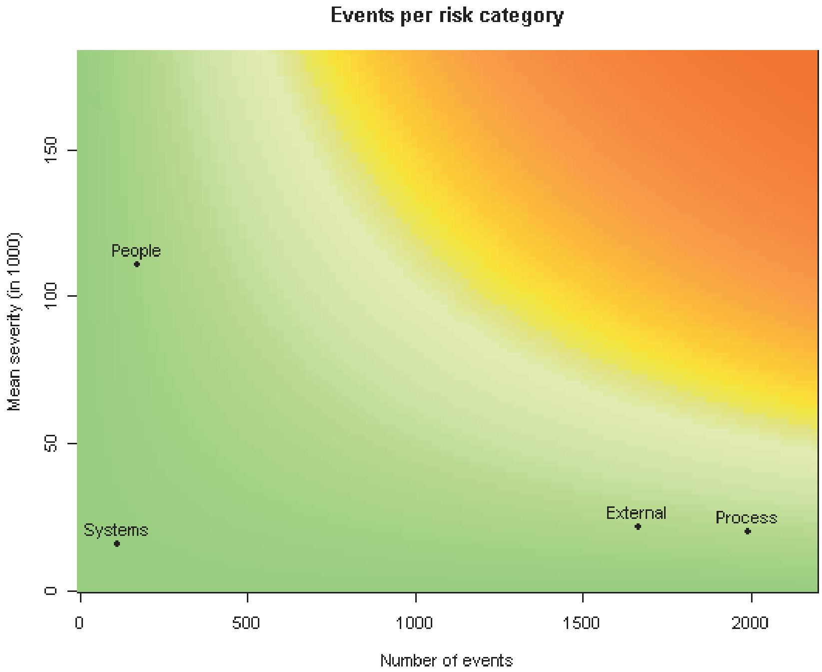
Process: These losses are unintentional. They occur because of deficient existing procedures or the absence of a procedure.

Systems: Losses caused by failures in the existing system such as breakdowns or a computer virus. This losses are unintentional, otherwise they rather should be categorised into External or People.

External: Losses caused by nature or man, or a direct result of a counterparty‘s action.

Fig. 2.1 presents an arrangement of this four categories by mean severity and frequency. As you can see the two risk categories External and Process have a high frequency, which means that many events are classified into these two categories, but show low severity. On the other hand the category People contains a lot of high severity events but with a much lower frequency. Just a few events are classified into the category Systems. These events have a rather low frequency, too.

Fig. 2.1 Typical arrangement of events per risk category



2.1 Examples of operational risk

This chapter gives a range of famous operational risk cases, that vary from the classical bank robbery to a flood disaster.[4] These cases are illustrating the wide range and diversity of operational risk events.

- Nick Leeson and Barings Bank - THE operational risk case

  Nicholas Leeson is a former derivatives trader whose unsupervised speculative trading caused the collapse of Barings Bank, the United Kingdom's oldest investment bank.
  He started working at Barings in the early 1990s; in a matter of a few years, he was appointed manager of a new operation in futures markets on the Singapore International Monetary Exchange (SIMEX).
  From 1992, Leeson made unauthorized speculative trades that at first made large profits for his employer, accounting for 10% of Barings' annual income. His luck quickly went sour, and he used a secret account to hide his losses. Leeson claims that this account was initially opened to hide a 20,000 trade of one of his subordinates that had been recorded incorrectly; however, Leeson used this account to cover future bad trades. He insists that he never used the account for his own gain, but in 1996 the New York Times quoted "British press reports" as claiming

[4]For detailed case studies see [5] and [21].

that investigators had located approximately 35 million dollars in various bank accounts tied to him. Management at Barings Bank also allowed Leeson to remain Chief Trader while being responsible for settling his trades, jobs that are usually done by two different people. This made it much simpler for him to hide his losses from the Bank. For detailed information see [14].

- Internal Fraud

  Subsequent case studies follow [5]:

  Peter Young, a fund manager at Morgan Grenfell, used money invested in the company's three largest European funds to purchase highly speculative stocks. These included a $ 30 million stake in Solv-Ex, a company with "a rather checkered past and nothing more tangible than ambitious plans for exploiting Canada's Athabasca tar sands for oil and minerals" (Barron's, 4 November 1996), whose stock Young bought not at a discount, but at a premium. In addition, Young established paired holding companies in order to circumvent Securities and Investment Board Regulation 5.14, which forbids a fund from owning more than 10% of any company, allowing him to increase the size of his risky stakes even further. After Morgan Grenfell became suspicious of the large quantity of unlisted shares in Young's portfolios, it shut down trading on the three funds and replaced the questionable investments with cash provided by parent company Deutsche Bank, an unusual move designed to maintain investors' confidence in the funds. When trading resumed, however, 30% of investors withdrew their money. The fiasco cost Morgan Grenfell an estimated 400 million pounds sterling, including compensation to some 80,000 investors for money lost due to trading irregularities. The damage to Morgan Grenfell's reputation was no doubt even more costly.

  On July 13, 1995, Daiwa Banks Toshihide Iguchi confessed, in a 30-page letter to the president of his bank in Japan, that he had lost around $1.1 billion while dealing in US Treasury bonds. The executive vice president of Daiwas New York branch had traded away the banks money over 11 years, an extraordinarily long period for such a fraud to run while using his position as head of the branchs securities custody department to cover up the loss by selling off securities owned by Daiwa and its customers.

- External Fraud

  According to a press release on 16.3.2001 in [21]: Abbas Gokal was head of the Gulf Group, a large shipping business based in Geneva with offices in more than forty countries. From the mid-1980's Gokal and his Gulf Group secretly received many millions of dollars from the Bank of Credit and Commerce International in London even though he knew, as did senior BCCI management, that his Gulf Group were insolvent. To cover up the fact that massive unsecured loans of more than $ 830 million were being made, Gokal and his fellow conspirators falsified documents on a vast scale and engineered an intricate money laundering operation. The relationship between Gokal's business and BCCI meant that fortunes were inextricably linked. BCCI's massive un-repaid loans to Gokal's Gulf Group inevitably played a significant part in the collapse of the bank in July 1991.

- Bank robberies

  Although Austria is blessed with a rather low quote of bank robberies, this cause of operational risk must not be neglected. There are some severe and sometimes also strange cases of bank robberies with losses above ten millions. One example is a bank robbery in Brasil.
  In 2005 a gang of burglars tunneled into the Banco Central in Fortaleza. They removed five containers of 50-real notes, with an estimated value of 164,755,150 realis (EUR 61.6 million) and weighting about 3,5 tons. The stolen money was uninsured. A bank spokeswoman stated that the risks were too small to justify the insurance premiums. The burglars managed to evade or disable the bank's internal alarms and sensors.
  Three months earlier, the gang of burglars had rented an empty property in the centre of the city and then tunneled 78 meters beneath two city blocks to a position beneath the bank. The gang had renovated a house and put up a sign indicating it was a landscaping company selling both natural and artificial grass as well as plants. Neighbours, who estimated that the gang consisted of between six and ten men, described how they had seen van-loads of soil being removed daily, but understood this to be a normal activity of the business.

- Settlements

  Swiss banks have agreed to pay $ 1.25 billion in restitution to survivors of the Holocaust. The deal settles a class-action lawsuit claiming banks failed to return funds to survivors and relatives after the second world war.

The next chapter describes the standards and qualifying criteria, which are required by Basel II for an operational risk management. Additionally, the three different approaches for measuring operational risk are introduced and explained.

3 Operational Risk and Basel II

One of the Basel II[1] requirements is to install an operational risk management. A bank must collect data about loss events, that fall into the category operational risk. This data must be monitored and controlled in a formalized and comprehensive way. This work will only consider the following statistical analysis of the operational risk data and the consequences for the mindest capital requirements[2].
There are three methodologies to measure operational risk:

1. the Basic Indicator Approach (BIA)
2. the Standardised Approach (TSA)
3. Advanced Measurement Approaches (AMA)

Each of these approaches should reflect the individual risk profile of each banking group. International Banking groups are expected to implement as soon as possible an AMA. Therefore the focus will be on this third approach.

---

[1]The Basel Committe on Banking Supervision provides a forum for regular cooperation on banking supervisory matters. Its objective is to enhance understanding of key supervisory issues and improve the quality of banking supervision worldwide. The Basel Committee of Banking Supervision (Basel II) members come from:

- Belgium
- Cananda
- France
- Germany
- Italy
- Japan
- Luxembourg
- Netherlands
- Spain
- Sweden
- Switzerland
- Great Britain
- USA

The countries are represented by their central bank. The present Chairman of the Committee is Mr. Nout Wellink, President of the Netherlands Bank. For more information about the Basel Committee see [22].

[2]For more information about creating an operational risk database or setting up an efficient risk management system see [29].

The following sections describe the three methodologies to measure operational risk. Definitions and different approaches are cited from "Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version" by the Basel Committee on Banking Supervision [25].

3.1 The Basic Indicator Approach - BIA

Banks are compelled to hold capital for operational risk to the average of the previous three years of a fixed percentage (denoted alpha) of positive annual gross income[3].

$$\mathbf{K_{BIA}} = [\sum(GI_{1,\ldots,n} * \alpha)]/n$$

Where:
K_{BIA}...the capital charge under the Basic Indicator Approach
GI...annual gross income, where positive, over the previous three years
n...number of the previous three years for which gross income is positive
α...15%, which is set by the Commitee, relating the industry wide level of required capital to the industry wide level of the indicator.

3.2 The Standardised Approach - TSA

To measure operational risk banks business activities are divided into eight categories, so called business lines: corporate finance, trading & sales, retail banking, commercial banking, payment & settlement, agency services, asset management and retail brokerage. For definitions in detail see Annex 1.

Within each business line, gross income is an indicator that serves as a proxy for the scale of business operations. Gross income is measured for each business line separately, instead of the whole institution. The capital charge for each business line is calculated by multiplying gross income by a factor (denoted beta) assigned to that business line. This factor is a proxy for the industry-wide relationship between the operational risk experience and the aggregate level of gross income.

[3]Gross income is defined as net interest income plus non-interest income.(Defined by national supervisors and/or national accounting standards.[25])

The capital charge is calculated as the three year average of the summation of the regulatory capital charges across each business line in each year.

$$\mathbf{K_{TSA}} = \{\sum_{years\ 1-3} \max[\sum(GI_{1-8} * \beta_{1-8}), 0]\}/3$$

Where:

K_{TSA}...the capital charge under the Standardised Approach
GI_{1-8}...annual gross income in a given year for each business line
β_{1-8}...a fixed percentage, set by the Basel Committee for Banking Supervision, relating the level of required capital to the level of the gross income for each of the eight business lines. Fig. 3.1 shows the different beta factors for each business line.

Fig. 3.1 beta factor for each business line

Business Line	**Beta**
Corporate finance (β_1)	18 %
Trading & sales (β_2)	18 %
Retail banking (β_3)	12 %
Commercial banking (β_4)	15 %
Payment and settlement (β_5)	18 %
Agency services (β_6)	15 %
Asset management (β_7)	12 %
Retail brokerage (β_8)	12 %

Qualifying criteria for the use of the TSA:

A bank must satisfy its supervisor that:

- Its board of directors and senior management, as appropriate, are actively involved in the oversight of the operational risk management framework;
- It has an operational risk management system that is conceptually sound and is implemented with integrity;
- It has sufficient resources in the use of this approach in the major business lines as well as the control and audit areas.

An internationally active bank using the Standardised Approach must meet some additional criteria. For detail see [25] p.141-142.

3.3 Advanced Measurement Approaches - AMA

Under the AMA, the regulatory capital requirement will equal the risk measure generated by the bank‘s internal operational risk measurement system using the quantitative and qualitative criteria for the AMA explained in the following sections. Use of the Advanced Measurement Approaches is subject to supervisory approval.

A further possibility will be to use the AMA for some parts of its operations and the Basic Indicator Approach or the Standardised Approach for the balance. This is called **partial use**. Here a bank must meet some criteria as well. For detail see [25] page 149.

3.3.1 Qualifying criteria

1. **General standards**

   The general standards are already mentioned in the section 3.2. The Standardised Approach.

2. **Qualitative Standards**[4]

   A bank using an advanced measurement approach must have an independent operational risk management function, that is responsible for design and implementation of the banks operational risk management framework, besides for the design and implementation of a risk-reporting system for operational risk, developing strategies to identify, measure, monitor and control/mitigate operational risk.

   The operational risk management system must be integrated into the day-to-day risk management processes of the bank. Further on there must be techniques to allocate operational risk capital to major business lines. Operational risk exposures and loss experience must be regularly reported to the senior management and to the board of directors. The management system must be well documented.

   The management processes and the measurement system must be regularly reviewed by external and internal auditors. Data flows must be transparent and accessible, whenever auditors want to judge it.

[4] For detailed listing of criteria see [25] page 142-148.

3. Quantitative Standards

The Committee is not specifiying the approach or distributional assumptions used to generate the operational risk measure for regulatory capital purposes. However, a bank must be able to demonstrate that its approach captures potentially severe tail loss events.

Any operational risk measurement system must include some certain key features: internal data, external data, scenario analysis and factors reflecting the business enviroment and internal control system. These features must be weighted in a credible, transparent, well-documented and verifiable way. Double counting of risk mitigants should be avoided.

- **Internal Data**

  Internal loss data can be used in different ways for the risk measurement system. One possibility is the use of internal data as foundation of empirical risk estimates, as a link between loss experience and risk manangement or it is used to validate the input and output of the risk measurement system.

  Internal data should be categorised into business lines and/or event types. For detailed information see Annex 1 and Annex 2. The bank should collect at least information about the date of the event, descriptive information about the cause and drivers and any recoveries of gross loss amount.

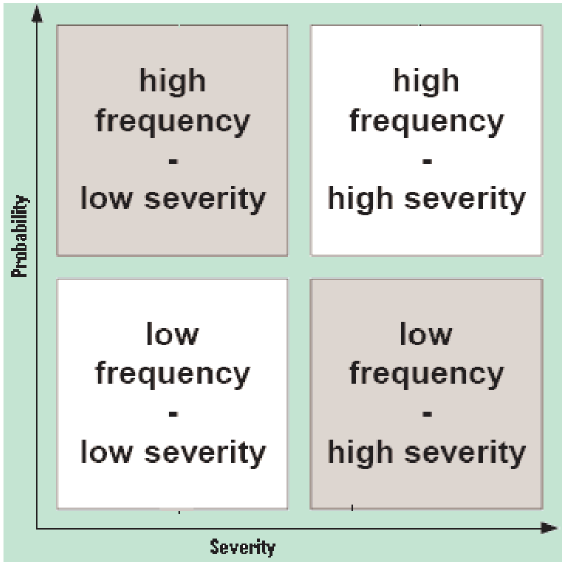
  Besides a five-year historical time series must be available except the bank for the first time moves to the AMA, then a three-year one is adequate.

  A further requirement is a de minimis gross threshold, i.e. EUR 1.000. This threshold may vary from business line to business line.

- **External Data**

  The operational risk measurement system must contain external data, either public data and/or pooled industry data. This data should complete the time series where internal data is rare. In particular there is often a lack of so called high severity - low frequency data (see fig. 3.2).
  External data should contain information about the date of the event, cause and drivers, of course, just like the internal database does.

Fig. 3.2 Database



The practice how external data is used must be documented and reviewed. Maybe it is necessary to scale or to make qualitative adjustments. This methodologies must be documented,too.

- **Scenario Analysis**

  A bank must use scenario analysis of expert opinion in conjunction with external data to evaluate its exposure to high-severity events. Experts assessments could be expressed as parameters of an assumed statistical loss distribution. Such analysis must be validated and re-assessed through comparison to actual loss experience.

- **Business enviroment and internal control factors**

  Business enviroment and internal control factors should make a bank's risk assessment more forward-looking. Such factors reflect the quality of the bank's control and operating enviroments more directly and recognise both improvements and deterioration in operational risk profile.

4. **Risk mitigation**

Under the AMA it will be allowed to take the risk mitigating impact of insurances under consideration for the measures, used for regulatory minimum capital requirements.

The following chapter 4 gives the theoretically statistical framework for measuring operational risk and presents the statistical approach for computing aggregate loss distributions, the so-called Loss Distribution Approach - LDA.

4 Loss Distribution Approach - LDA

As already mentioned the Basel Committee on Banking Supervision does not make any concret requirements about statistical distributions or procedures to implement an AMA. The Loss distribution Approach (LDA) is a common and popular framework for computing the capital charge of a bank, using an AMA, for operational risk. It is a statistical approach for computing aggregate loss distributions. LDA can be considered as a framework for bottom-up economic capital allocation [11].

Under the **Loss Distribution Approach***, the bank estimates, for each business line/risk type (event type: see Annex II) cell, the probability distribution functions of the single event impact and the event frequency for the next (one) year using its internal data, and computes the probability distribution function of the cumulative operational loss.* See Annex 6 of [23].

Definition: *A* **loss distribution** *is the probability distribution of either the loss or the amount paid from a loss event or of the amount paid from a payment event. The distribution may or may not exclude payments of zero and may or may not include allocated loss adjustment expense, which is the amount of expense incurred directly as the result of a loss event[17].*

The modelling of the aggregate loss distribution is done in two steps.[1] First the single distributions of loss frequency and loss severity are computed and then these two distributions are compounded to the loss distribution.

$$S = X_1 + X_2 + ... + X_N$$

where:
X_i...single loss event
S...whole loss for one period (random variable)

• The distribution of the X_i $i = \{1, ..., N\}$ is independent of n, given N=n.
• Given N=n, the X_i $i = \{1, ..., N\}$ are independently and identically distributed.
• The distribution of N is independent of $X_1, ..., X_n$.

The random sum S has a distribution function

[1] Statistical concepts follow [17].

$$\begin{aligned} F_S(x) &= Pr(S \le x) \\ &= \sum_{n=0}^{\infty} p_n Pr(S \le x | N = n) \\ &= \sum_{n=0}^{\infty} p_n F_X^{*n}(x) \end{aligned}$$

where $F_X(x) = Pr(X \le x)$ is the common distribution function of the X_is and $p_n = Pr(N = n)$. $F_X^{*n}(x)$ is the "n-fold convolution" of the cummulative distribution function (cdf) of X. It can be obtained by

$$F_X^{*0}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \ge 0 \end{cases}$$

and

$$F_x^{*k}(x) = \int_{-\infty}^{\infty} F_X^{*(k-1)}(x-y) dF_X(y).^2$$

The density of the compound distribution can be written as

$$f_S(x) = \sum_{n=0}^{\infty} p_n f_X^{*n}(x).$$

Because of the independency of the distribution of N and the distribution of $X_1, ..., X_N$ expectation and variance are defined as follows:

$$\mathbb{E}S = \mathbb{E}N\mathbb{E}X$$
$$VarS = \mathbb{E}NVarX + VarN(\mathbb{E}X)^2$$

Computing the loss distribution often requires numerical methods, because in general there is no analytical expression of the compound distribution. One of these numerical methods is Monte Carlo simulation.

The **Monte Carlo simulation** approach can be explained in three steps:

1. You have to build a statistical model for S, which is dependent on certain random variables (here X and N). The distributions of the random variables and their dependencies are known.

2. Then pseudorandom variables x_i and n_i are generated for X and N. The next step is to compute s_i by using the model from above. This is done for i=1,...,k. Note that k should be large (i.e 10^6 or 10^7).

3. The distribution of S may be approximated by $F_{\hat{S}}(s)$the empirical distribution of the pseudorandom sample $s_1, ..., s_k$.

Other methods are the Panjer‘s recursive approach or Fast Fourier transform. For details see [17].

After computing the compound loss distribution, one possibilty to calculate the capital charge is using the Value-at-Risk (VaR) method.

[2] For more information see [17].

4.1 Value-at-Risk - VaR

It is suggested by Basel II that the capital charge is based on the simple sum of the operational risk VaR for each business line/risk type cell.[23]

The VaR is defined for a business line i and a risk type cell (event type) j as follows:

The expected loss $\mathbf{EL(i,j)}$ corresponds to the expected value of the random variable $S_{i,j}$. The expected loss can be computed for instance as the mean of Monte Carlo simulation results. The $EL(i,j)$ is expressed by:

$$EL(i,j) = \mathbb{E}[S_{i,j}] = \int_0^\infty s \; dF_{i,j}(s)$$

where:
$F_{i,j}$...specific distribution of S
$S_{i,j}$...random variable (loss for one BL/EV cell)

The unexpected loss **UL(i,j)** is determined by subtracting the EL(i,j) from the loss amount at a desired confidence level α. Using Monte Carlo simulation for aggregating the loss severity and the loss frequency distribution, this loss amount is the empirical α-quantile of the empirical distribution function $F_{\hat{S}}(s)$ of the pseudorandom sample $s_1, ..., s_k$ (see previous page).

$$UL(i,j,\alpha) = F_{i,j}^{-1}(\alpha) - EL(i,j) = \inf\{z|F_{i,j}(z) \geq \alpha\} - \int_0^\infty z \; dF_{i,j}(z)$$

The so defined unexpected loss is also called Value-at-Risk at confidence level α. For implementing an AMA a confidence level of $\alpha = 0.999$ (99.9%) is required[3].

The capital charge for one business line/event type cell:

$$\textbf{Capital charge (i,j,}\alpha\textbf{)=UL(i,j,}\alpha\textbf{)}$$

[3]In economic capital project, α is related to the rating target of a bank [11].

Rating target	BBB	A	AA	AAA
α	99.75%	99.9%	99.95%	99.97%

The capital charge for the whole institution is the sum of the capital charges across each of the business lines and event types:

$$\text{Capital charge } (\alpha) = \sum_{i=1}^{I} \sum_{j=1}^{J} \text{capital charge (i,j,}\alpha\text{)}$$

So far, it has not been reflected how to find the right distribution for the loss severity and the loss frequency. The next three sections try to answer this question. First of all some common distributions for the loss frequency are presented, then loss severity distributions are introduced and last but not least the procedure of goodness-of-fit tests is shown.

4.2 Frequency Distribution

To model the loss frequency, so called counting distributions are used. These distributions are discrete with support only on the non-negative integers (only defined on the points 0,1,2,...). Such distributions are for instance the Poisson, the binomial or the negative binomial distribution.

Definition: *The frequency is the number of losses or number of payments random variable. Its expected value is called the mean frequency. Unless indicated otherwise the frequency is for one exposure unit.*[17]

The probability function p_k denotes the probability that exactly k events occur.

$$p_k = \Pr(\mathrm{N}{=}k), \quad k{=}1{,}2{,}...,$$

where N is a random variable representing the number of events.

One important concept in probability theory is the probability generating function, which can be used to generate moments of a distribution.

Definition: The **probability generating function**[4] of a discrete random variable N with probabilty function p_k is

$$\mathrm{P}(z){=}\mathbb{E}(z^N){=}\sum_{k=0}^{\infty} p_k z^k.$$

The probabilty generating function has some nice properties:[5]

1. If X and Y are two independent random variables with probability generating funtion F and G, the probability generating function H of Z=X+Y is H=FG.

2. If $\mathbb{E}(X^k) < \infty$ and F is the probabilty generating function of X, then

$$F^{(k)}{=}\mathbb{E}(X(X-1)...(X-k+1)).$$

3. Let $X_1, ..., X_n$ be independently and identically distributed with probability generating function F. Y be independent of the sequence (X_n) with probability generating function G. The probability generating function of Z be H. Z is defined as follows:

$$\mathrm{Z}{=}\sum_{i=0}^{Y} X_i$$

   Then H can be written as $H(z) = G(F(z))$.

[4]In general: $A(z) = \sum_{n=0}^{\infty} a_n z^n$ is called **generating function** of the sequence $\{a_n\}$, where $\{a_n\}$ is a sequence of real numbers. Only defined, if the series converges.

[5]For proof see [7].

4.2.1 Poisson Distribution

Generally the number of events N occuring in one period is assumed to be Poisson distributed, i.e. $N \sim \mathcal{P}_\lambda$. The advantage of this distribution is that the mean and the variance of a Poisson random variable is the parameter λ. Therefore it is very easy to estimate this parameter. This property follows directly from the probabilty generating funtion. Fig. 4.1 shows the density function of a Poisson distribution with different parameters λ. Because λ is both mean and variance the density function is the flatter the bigger λ.

The probabilty function of the Poisson distribution is $p_k = \frac{e^{-\lambda}\lambda^k}{k!}$, k=0,1,2,... . The probability generating function follows as

$$P(z) = e^{\lambda(z-1)}, \quad \lambda > 0 \; .$$

$$\mathbb{E}(N) = P^{'}(1) = \lambda$$

$$\mathbb{E}[N(N-1)] = P^{''}(1) = \lambda^2$$

$$\begin{aligned} Var(N) &= \mathbb{E}[N(N-1)] + \mathbb{E}(N) - [\mathbb{E}(N)]^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

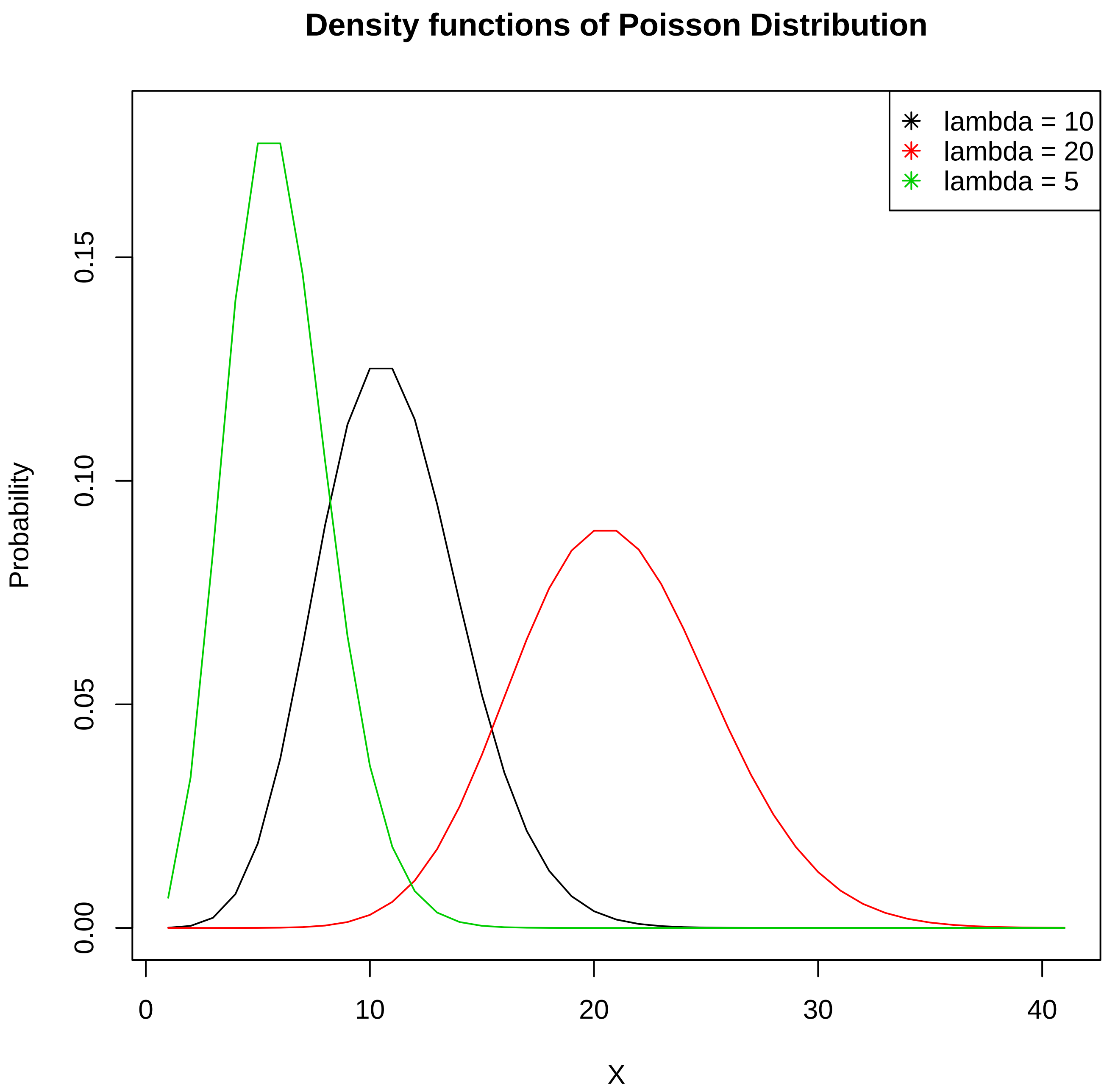
As already mentioned banks activities are divided into business lines or/and event types. The capital charge is the sum of the VaR of each business line and/or event type. The next two properties are very useful for aggregating the loss distribution.

Theorem 4.1: Let $N_1, ..., N_n$ be independent Poisson variables with parameters $\lambda_1, ..., \lambda_n$. Then $N = N_1 + ... + N_n$ has a Poisson distribution with parameter $\lambda = \lambda_1, ..., \lambda_n$.[6]

Theorem 4.2: Suppose that the number of events N is a Poisson random variable with mean λ. Further suppose that each event can be classified into one of m types with probability $p_1, ..., p_m$ independent of all other events. Then the number of events $N_1, ..., N_m$ corresponding to event types $1, ..., m$ respectively, are mutually independent Poisson random variables with means $\lambda p_1, ..., \lambda p_m$ respectively.[7]

[6]For proof see [17].
[7]For proof see [17].

Fig. 4.1 Poisson Distribution with different λ



Estimation

The parameter λ of the Poisson distribution can be estimated by the method of the moments and the method of maximum likelihood.

1. Method of the moments

   The empirical estimate of the **mean** is defined by[8]

   $$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

   Empirical estimates of other moments:

   Definition: The k^{th} **raw moment** is

   $$\mu_k^{'} = \mathbb{E}(X^k).$$

   The k^{th} **central moment** is

   $$\mu_k = \mathbb{E}[(X-\mu)^k].$$

   The **variance** is $\sigma^2 = \mu_2 = \mu_2^{'} - \mu^2$.

2. Maximum-Likelihood estimation

   The likelihood function for a set of n independent observations is

   $$L(\theta; x) = \prod_{j=1}^{n} L_j(\theta; x_j) = \prod_{j=1}^{n} f(x_j; \theta),$$

   where: $L_j(\theta; x)$...contribution of the j^{th} observation to the likelihood;

   Definition: The **Maximum-Likelihood estimator** is a measurable function $\hat{\theta}(X)$, with

   $$L(\hat{\theta}(X), X) = \sup_{\Theta} L(\theta, X).$$

   This estimator may not be unique. To calculate the estimator it is easier to compute the log-likelihood. By setting the derivative of the log-likelihood (with respect to θ) zero the maximum likelihood estimator can be achieved.

[8] Definition by [31].

The log-likelihood:

$$l(\theta; x) = \log L(\theta; x) = \sum_{j=1}^{n} \log L_j(\theta; x)$$

<u>The results for the Poisson distribution:</u>
The expected frequency (sample mean) is

$$\hat{\lambda} = \frac{\sum_{k=0}^{\infty} k n_k}{\sum_{k=0}^{\infty} n_k}$$

where:
n_k...number of years in which a frequency of exactly k loss events occurs, n_k represents the number of observed values at frequency k.

The Maximum-Likelihood estimation results in following:

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}$$
$$\log p_k = -\lambda + k \log \lambda - logk!$$

The log-likelihood is

$$l = -\lambda n + \sum_{k=0}^{\infty} k n_k \log \lambda - \sum_{k=0}^{\infty} n_k \log k!.$$

By setting the derivative of the log-likelihood zero, the maximum likelihood estimator is:

$$\hat{\lambda} = \frac{\sum_{k=0}^{\infty} k n_k}{n}.$$

For the Poisson distribution the maximum likelihood estimator and the method of moments estimator are identical. The estimator has mean λ and variance λ/n.

4.2.2 Negative Binomial Distribution

The negative binomial distribution has positive probabilties on the non-negative integers just like the Poisson distribution. One big difference to the Poisson distribution is the fact, that the negative binomial distribution has two parameters, which provides more flexibility in shape. You say $N \sim NegBin_{r,\beta}$ if:

$$Pr(N = k) = p_k = \binom{k+r-1}{k} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^k, \quad k = 0, 1, ..., \quad r > 0, \beta > 0.$$

The probability generating funciton is:

$$P(z) = [1 - \beta(z-1)]^{-r}$$

Therefore mean and variance of the negative binomial distribution are:

$$\mathbb{E}(N) = r\beta \qquad \text{and} \qquad Var(N) = r\beta(1+\beta).$$

Because $\beta > 0$ the variance exceeds the mean. This is contrary to the Poisson distribution, where the mean is equal to the variance. If the observed variance is larger than the observed mean, the negative binomial distribution will be a better choice for the loss frequency distribution than the Poisson distribution. This case is also called overdispersion.

The probability function of the negative binomial distribution can also be written as:

$$Pr(N = k) = \frac{\Gamma(k+r)}{\Gamma(r)k!} p^r (1-p)^k,$$

where:
$p ... 1/(1+\beta), \;\; 0 < p \leq 1.$

The method of moments estimators for the two parameters of the negativie binomial distribution are:

$$r\beta = \frac{\sum_{k=0}^{\infty} k n_k}{n} \qquad \text{and}$$

$$r\beta(1+\beta) = \frac{\sum_{k=0}^{\infty} k^2 n_k}{n} - \left(\frac{\sum_{k=0}^{\infty} k n_k}{n}\right)^2.$$

The following two figures 4.2 and 4.3 show the influence of different values for p and r to the shape of the density function of the negative binomial distribution. The smaller p the flatter is the density distribution. For the parameter r this behaviour is vice versa.

Fig 4.2 Negative Binomial Distribution with different p

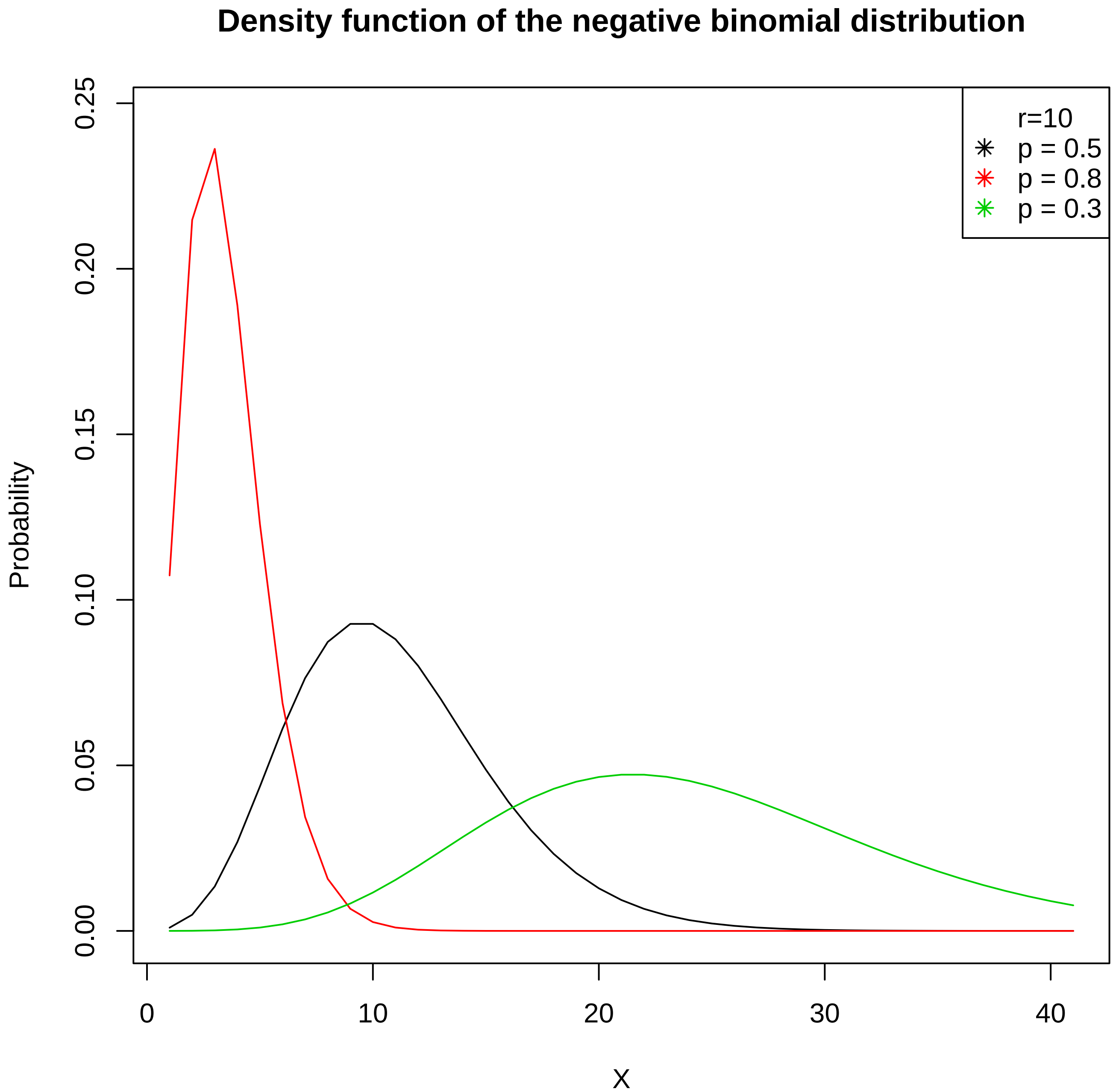
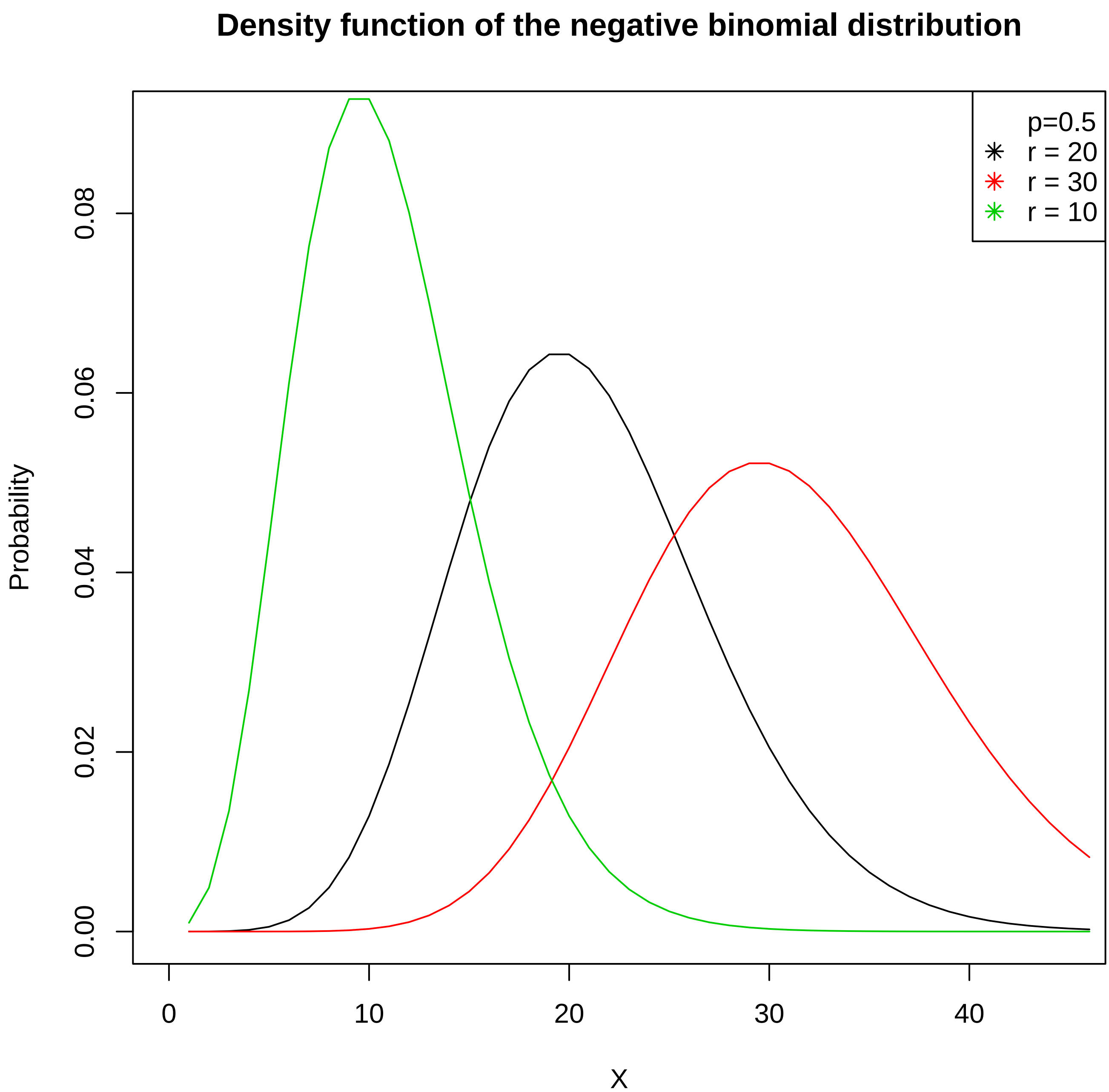


Fig 4.3 Negative Binomial Distribution with different r



4.2.3 Binomial Distribution

The binomial distribution is a counting distribution, too. Unlike the negative binomial distribution the variance of the binomial distribution is smaller than the mean. Therefore this distribution should be used for data samples with smaller empirical variance than empirical mean. This is called underdispersion.

The probability function represent the probability that exactly k claims occur.

$$p_k = Pr(N = k) = \binom{m}{k} q^k (1-q)^{m-k}, \quad k = 0, 1, ..., m \quad \text{and} \quad 0 < q < 1$$

The probabilty generating function is given by:

$$P(z) = [1 + q(z-1)]^m, \quad 0 < q < 1.$$

Mean and the variance of the binomial distribution are defined as follows:

$$\mathbb{E}(N) = mq \quad \text{and} \quad Var(N) = mq(1-q).$$

In general the parameter m is known and fixed. Then the parameter q is estimated as

$$\hat{q} = \frac{\text{Number of observed events}}{\text{Maximum number of possible events}}$$

This is the method of moments estimator when m is known.
The log likelihood is

$$l = \sum_{k=0}^{\infty} n_k \log p_k.$$

Setting the derivative of the log likelihood zero leads to

$$\hat{q} = \frac{1}{m} \frac{\sum_{k=0}^{\infty} k n_k}{\sum_{k=0}^{\infty} n_k}.$$

For fixed m the maximum likelihood estimator is equal to the method of moments estimator.

Fig. 4.4 illustrates the influence of different parameters q and fig. 4.5 shows the density functions of the binomial distribution for different m. Growing q results in a shifted density function to the right and a steeper density function follows by a decreasing m .

Fig 4.4 Binomial distribution with different q

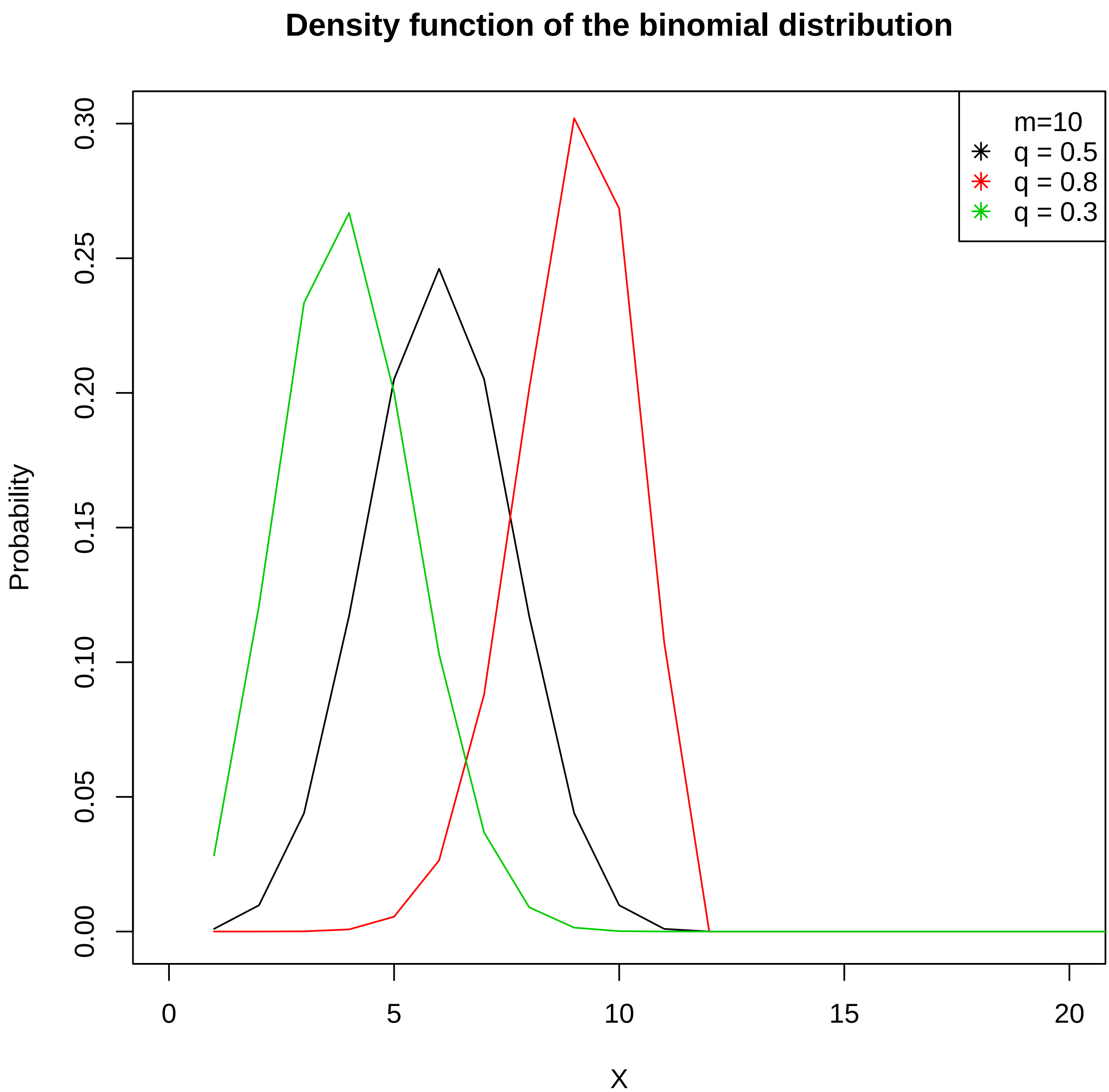
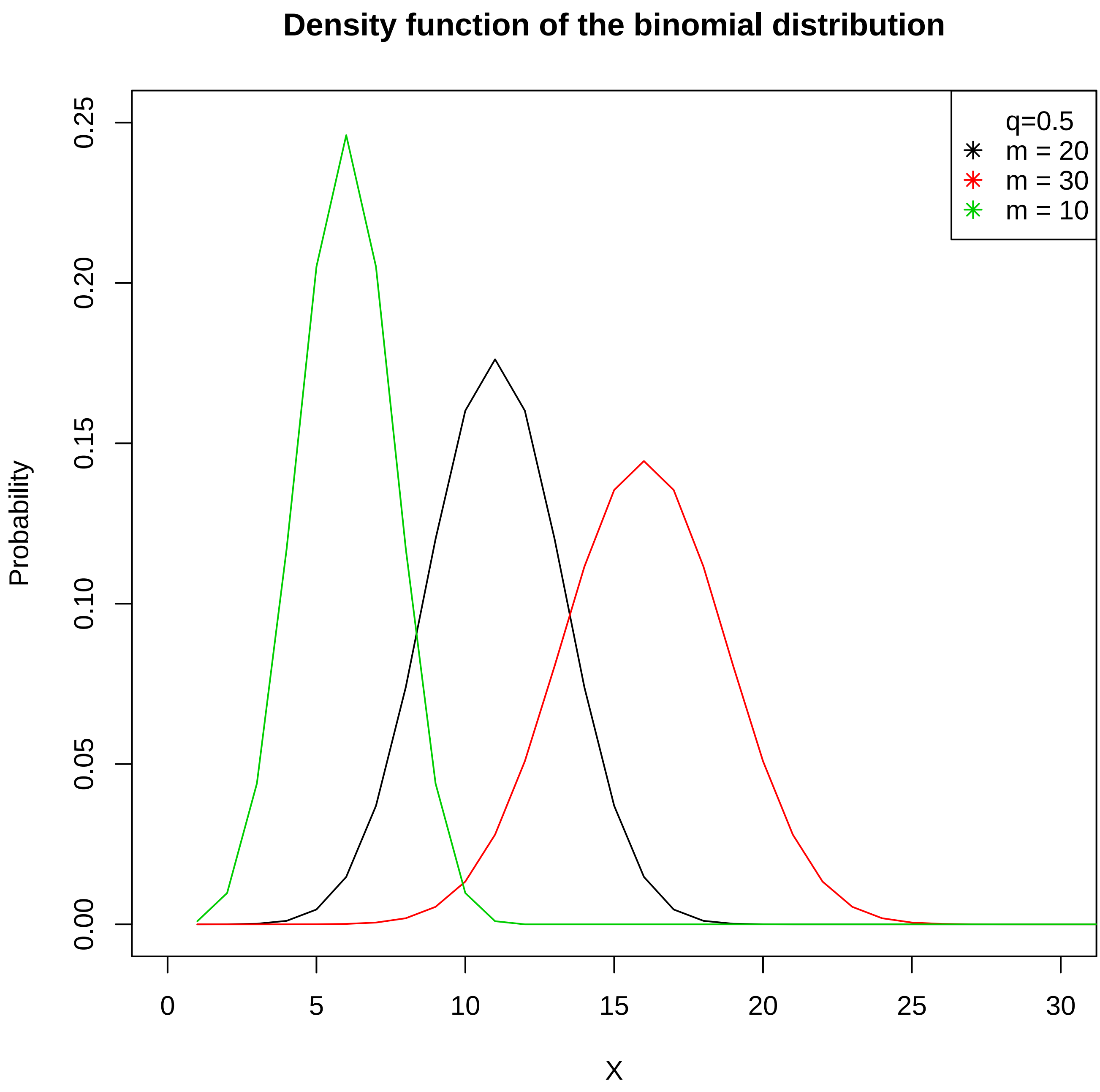


Fig 4.5 Binomial distribution with different m



4.3 Severity Distribution

The severity distribution should fulfill some characteristics:

1. It should be smooth
2. and place probability on all non-negative real numbers.

Distributions, that meet these criteria, are the log-normal distribution, the Generalized Pareto distribution or the Weibull distribution. The methods used to estimate the parameters of these distributions are the method of moments and the method of maximum likelihood.

Definition: *The severity can be either the loss or amount paid random variable. Its expected value is called the mean severity.*[17]

4.3.1 Log-normal distribution

A positive random number X is called to be log-normal distributed i.e. $X \sim LN(\mu, \sigma^2)$, if

$$\log(X) \sim N(\mu, \sigma^2).$$

This means, that the logarithm of X follows a normal distribution with same parameters μ and σ^2. Fig. 4.6 and fig. 4.7 give a survey of the shape of the density function with different values for the parameters. Growing μ with constant σ causes a flatter density function and growing σ with constant μ a steeper one.
The density function of the log-normal distribution is given by

$$f(x) = \begin{cases} \frac{1}{\sigma x \sqrt{2\pi}} \exp[-\frac{(lnx-\mu)^2}{2\sigma^2}] & \text{for} \quad x \geq 0 \\ 0 & \text{else} \end{cases} \qquad \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+.$$

Where $\mu = \mathbb{E}(lnX)$ and $\sigma^2 = Var(lnX)$. Therefore it is simple to estimate the parameters by the method of the moments, where the parameters μ and σ^2 are just the empirical mean and the empirical variance.
Note, that the empirical mean is the maximum likelihood estimator of μ, too.

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} lnX_i$$

$$\hat{\sigma^2} = \frac{1}{n-1} \sum_{i=1}^{n} (lnX_i - \hat{\mu})^2$$

Fig 4.6 Log-Normal distribution with different variances

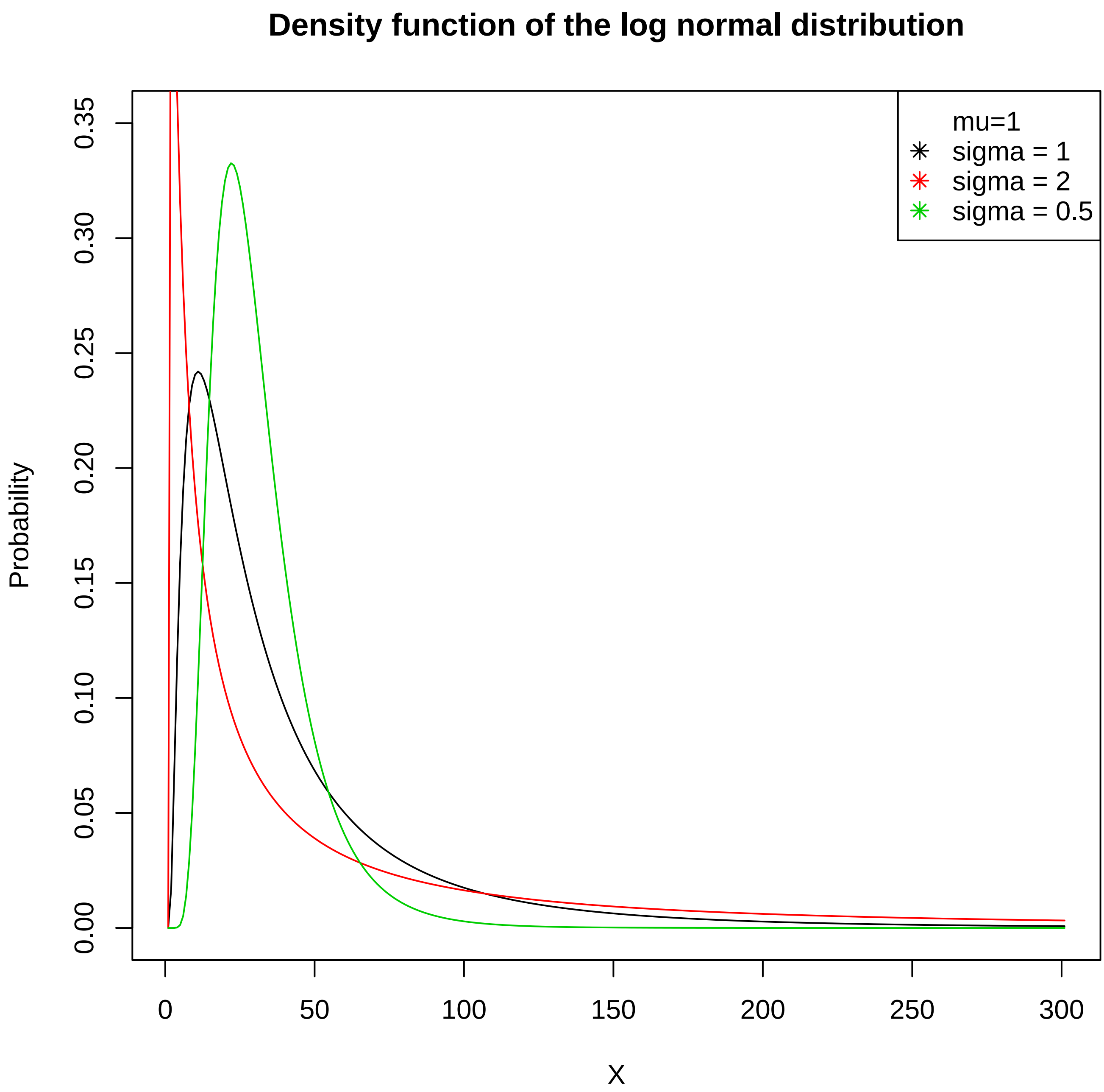
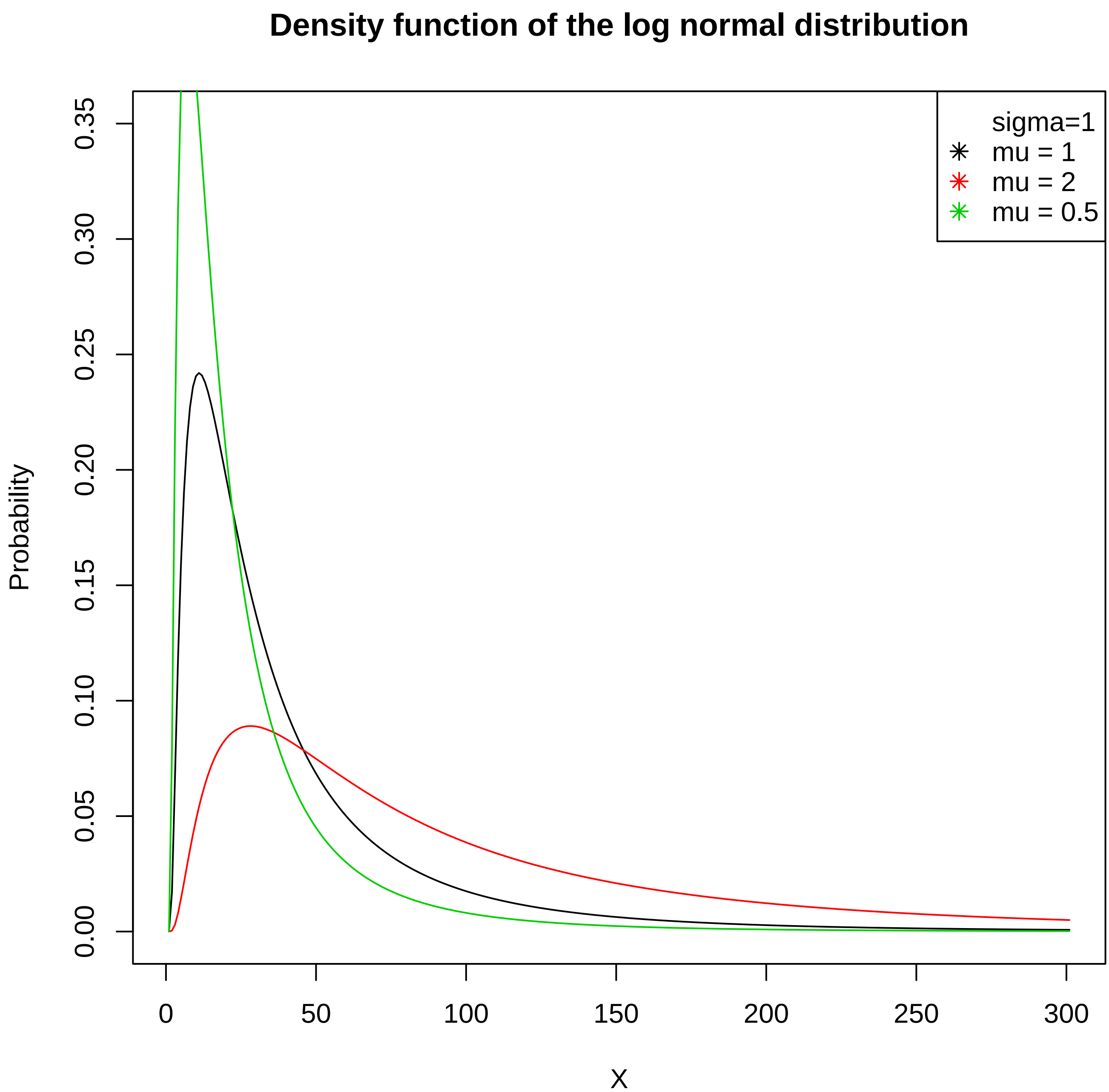


Fig 4.7 Log-Normal distribution with different means



4.3.2 Generalized Pareto distribution

The distribution function of the Generalized Pareto distribution has three parameters and is given by

$$F(z) = 1 - [1 + \tfrac{\xi(z-\mu)}{\sigma}]^{-\frac{1}{\xi}} \quad , \text{for } z > \mu, \sigma > 0 \quad \text{and} \quad 1 + \tfrac{\xi(z-\mu)}{\sigma} > 0$$

where:
μ...location parameter
σ... scale parameter
ξ... shape parameter

Fig.4.8 to fig.4.10 illustrate the changing of the density function by varying the parameter values.

Fig 4.8 Generalized Pareto distribution with different location parameters

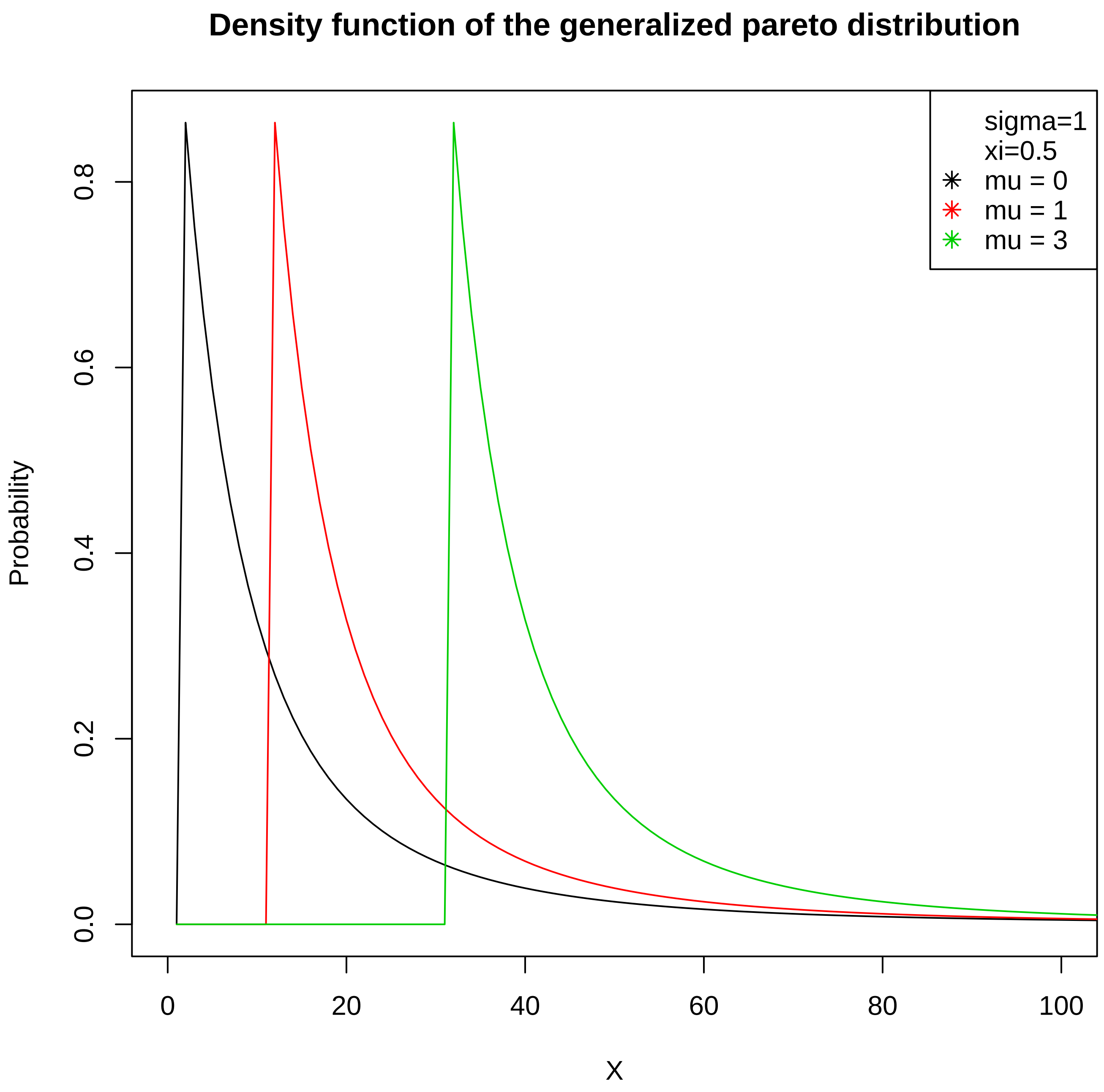


Fig 4.9 Generalized Pareto distribution with different scale parameters

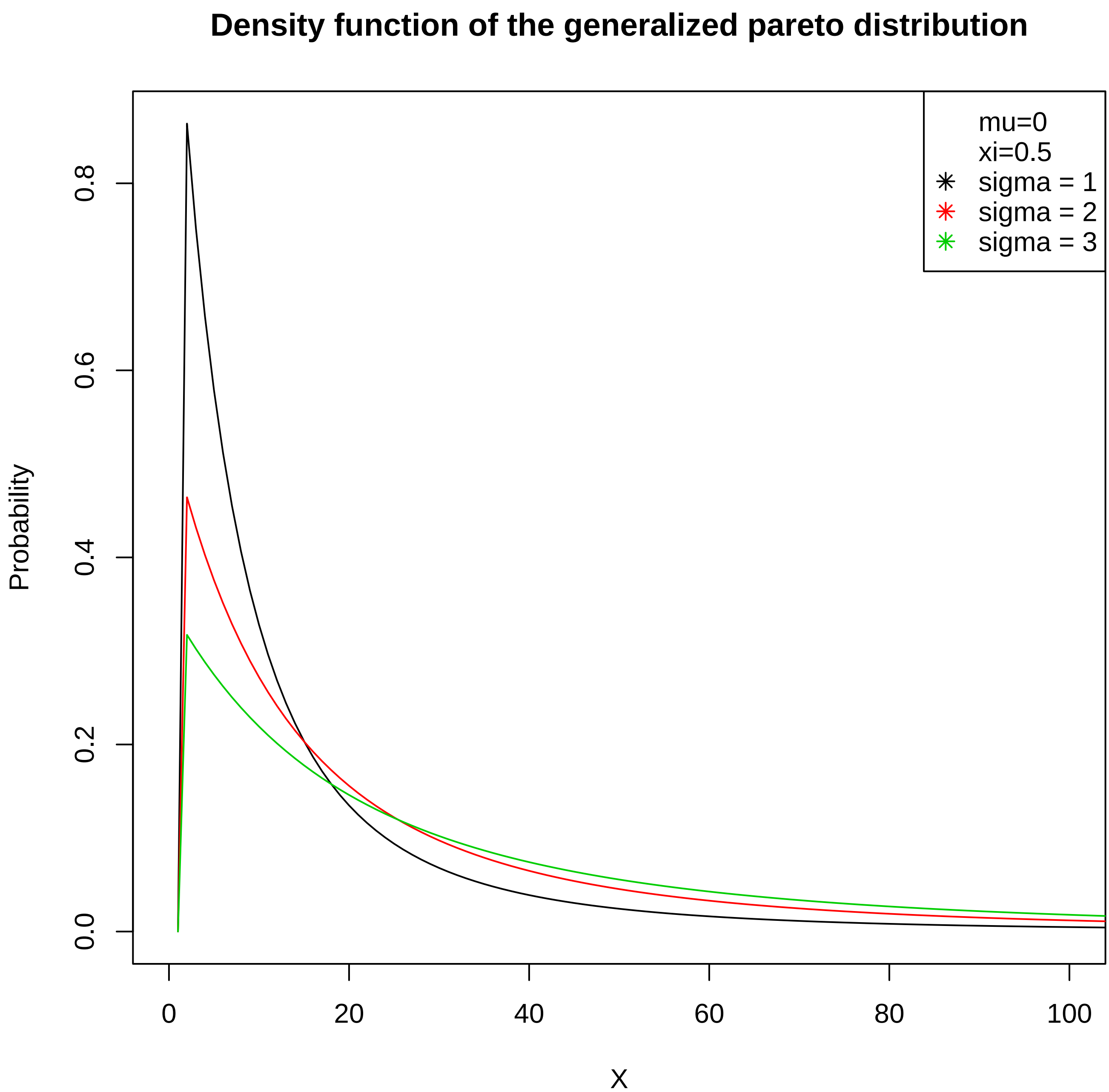
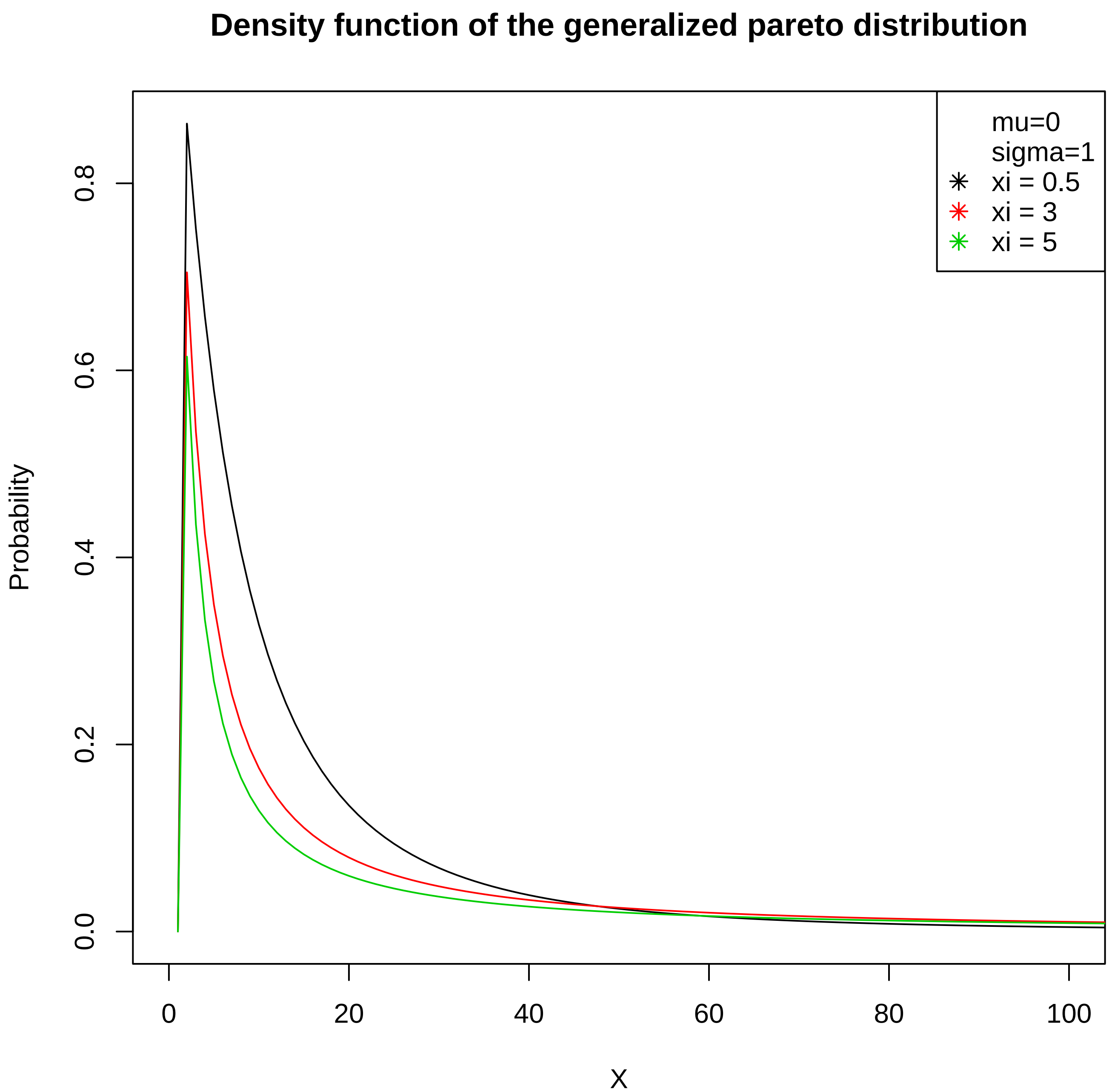


Fig 4.10 Generalized Pareto distribution with different shape parameters



4.3.3 Weibull distribution

The Weibull distribution has two parameters a and b, where a represents the scale and b represents the shape of the distribution. $X \sim We(a,b)$ if

$$f(x) = \begin{cases} \frac{a}{b}(\frac{x}{b})^{a-1}\exp[-(\frac{x}{b})^c] & \text{for} \quad x \geq a \qquad a,b \in \mathbb{R}^+ \\ 0 & \text{else} \end{cases}$$

$$\mathbb{E}(X) = b\Gamma(1+\tfrac{1}{a}) \quad \text{and} \quad Var(X) = b^2(\Gamma(1+\tfrac{2}{a}) - \Gamma(1+\tfrac{1}{a})^2)$$

The maximum likelihood estimators for a and b are the solutions of

$$\hat{a} = [\tfrac{1}{n}\textstyle\sum_{i=1}^{n}(X_i^{\hat{b}})]^{\frac{1}{\hat{b}}} \qquad \text{and}$$

$$\hat{b} = \frac{n}{\frac{1}{a}^{\hat{b}}\sum(X_i)^{\hat{b}}ln(X_i) - \sum ln(X_i)}$$

Fig. 4.11 and fig. 4.12 show the density functions of the Weibull distribution with different parameter values for a and b. The parameter b influences the slope and a the height of the density function.

Fig 4.11 Weibull distribution with different a

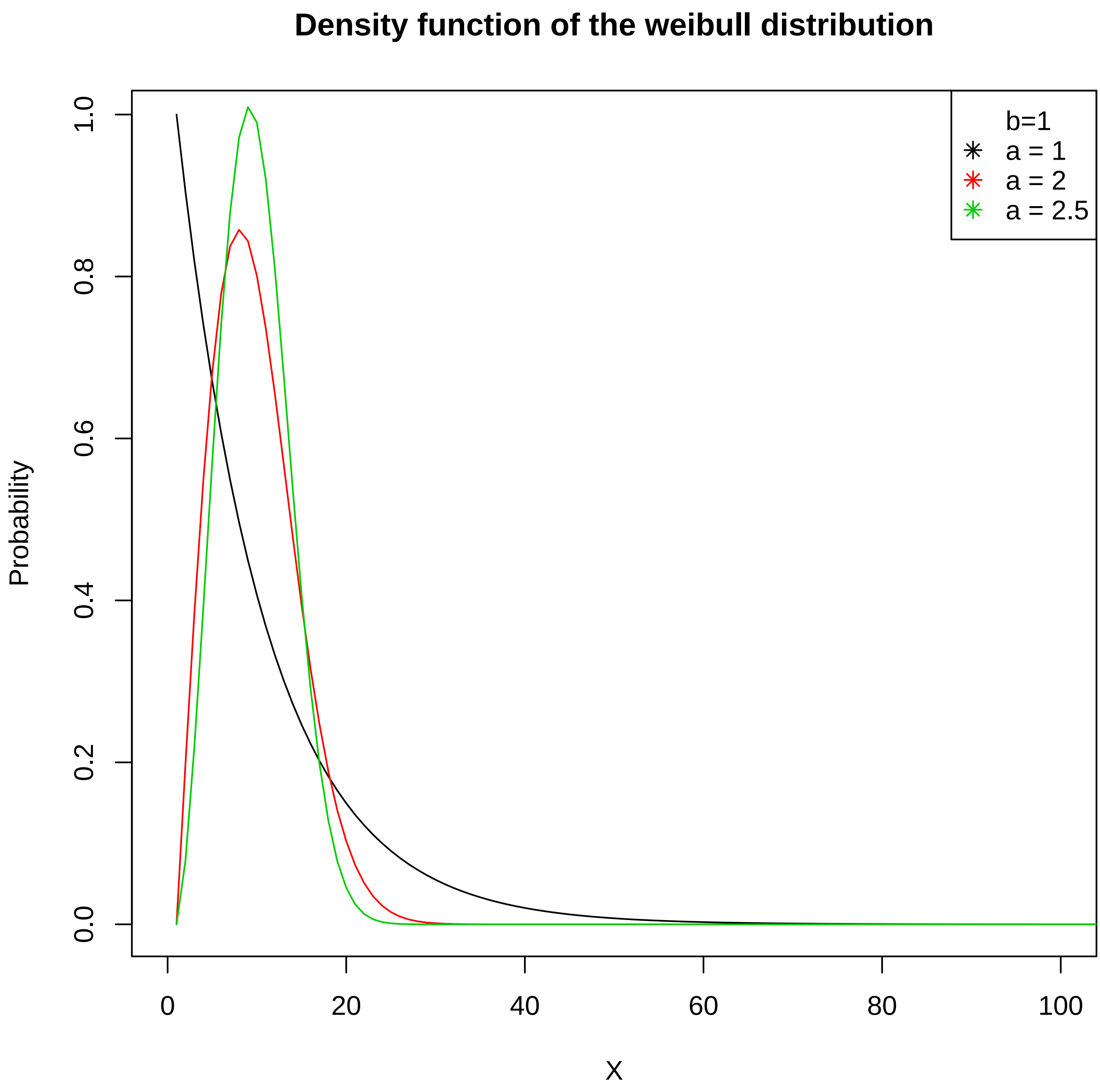
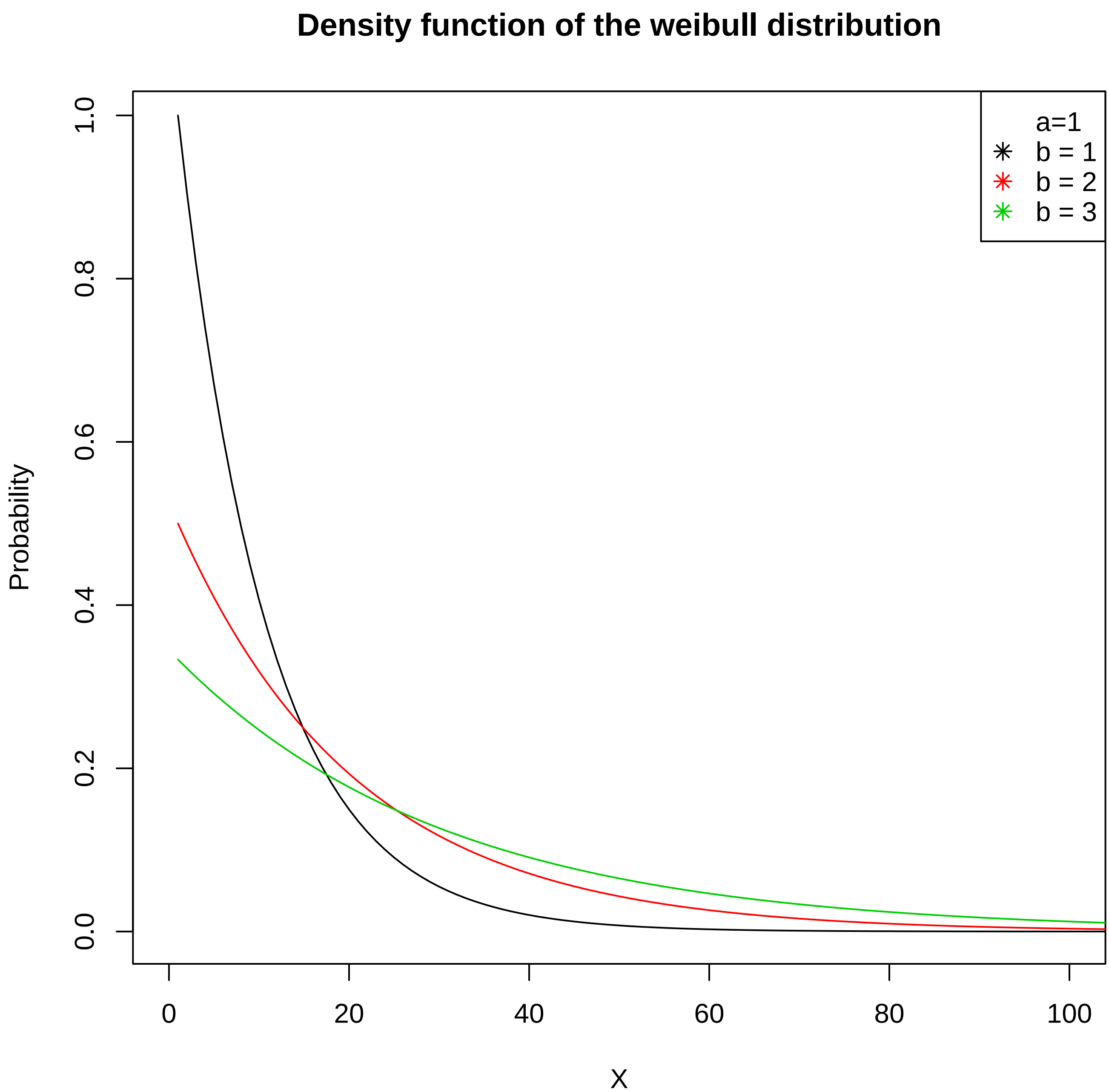


Fig 4.12 Weibull distribution with different b



4.4 Piecewise-defined loss severity distribution

The loss severity distribution may be divided into m parts, here $m = 2$. Losses, smaller than a certain threshold $(= t_1)$[9] are modelled by their empirical distribution function. For losses above this threshold a loss severity distribution is fitted (The choice for the model, which will be developed in chapter 6, is the log-normal distribution.). An illustration of this concept is presented in fig. 6.7 and fig. 6.8 of chapter 6.

The combined loss severity distribution $F(x)$ is given by

$$\overline{F}(x) = \begin{cases} \overline{F}_1(x) & \text{for} \quad x < t_1 \\ \overline{F}_2(x) \cdot \overline{F}_1(t_1) & \text{for} \quad x \geq t_1 \end{cases},$$

where $\overline{F}(x)$ denotes $1 - F(x)$ [2]. $F_1(x)$ is the empirical distribution function and $F_2(x)$ is the distribution function of the loss severity distribution.

To estimate parameters of $F_2(x)$ some theory about truncated data and distribution must be introduced.

Definition: *Data are said to be* **truncated** *when observations that fall in a given set are excluded. Data are said to be* **truncated from below** *when the set is all numbers less than a specific value. Data are said to be* **truncated from above** *when the set is all numbers greater than a specific value.*[17]

In the case of operational risk data the second form occurs. Operational risk data are not collected from EUR 0. The treshold for collecting operational risk data can be variable (EUR 1.000, EUR 10.000 etc.). This truncation is called truncation from above and negliable.

A truncation, which cannot be neglected, occurs when it comes to the fitting of the second part of the distribution. If X represents the loss, then the loss severity distribution is fitted for the variable Y defined by:

$$Y = \begin{cases} \text{undefined} & \text{for} \quad X \leq t_1 \\ X - t_1 & \text{for} \quad X > t_1 \end{cases}$$

The distribution function of Y is given by:

$$F_Y(y) = \begin{cases} 0 & \text{for} \quad y = 0 \\ \frac{F_X(y+t_1) - F_X(t_1)}{1 - F_X(t_1)} & \text{for} \quad y > 0 \end{cases}$$

[9] For instance this threshold can be EUR 100.000 or EUR 1 million.

4.5 Goodness-of-fit Tests

Two non-parametric[10] statistical methods for checking the goodness-of-fit of a loss frequency or a loss severity distribution choice are the Kolmogoroff-Smirnoff-Test and the Chi-Squared-Test.

To explain the test procedure of the KS-test some characteristics of the empirical distribution function are presented.

Definition: The **empirical distribution function** $\hat{F}_n(x)$ is defined by:

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(x_i),$$

where $I_i(x_i)$ denotes the Indicator function.

Theorem 3.3: The Glivenko - Cantelli Theorem[11]
For the empirical distribution function $\hat{F}_n(x)$ of the random sample $X_1, ..., X_n$ with $X_i \sim F(\cdot)$ the distance

$$D_n := \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \rightarrow 0 \text{ almost surely.}$$

I.e. $\Pr(\lim_{n\rightarrow\infty} D_n = 0) = 1$. D_n is called the Kolmogoroff-Smirnoff statistic.

If $F(\cdot)$ is continuous, the distribution of D_n is independent of $F(\cdot)$. $\sqrt{n}D_n \rightarrow D$ in distribution, the distribution of D is called Kolmogoroff-Smirnoff distribution.

4.5.1 Kolmogoroff-Smirnoff Test

This test can only be used for continuous distributions. Therefore it is only applicable to the loss severity distributions, because each of the presented loss frequency distributions is discrete.

The hypothesis $H_0 : F = F_0$ is tested by calculating the maximum distance of the empirical distribution function $\hat{F}_n$ and F. This maximum distance is the Kolmogoroff-Smirnoff statistic D_n, mentioned above.

The test rule is the following:
The hypothesis $H_0 : F = F_0$ is rejected, if $D_n > d_{1-\alpha}$, where $d_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the Kolmogoroff-Smirnoff distribution D. α is the confidence level, in general $\alpha = 0.05$ or $\alpha = 0.01$.

[10] A model is called to be **non-parametric**, if the range of the parameter is not finite. It does not contain numbers, but functions like density or distribution functions.

[11] For proof see [15] p.190.

4.5.2 Chi-Squared-Test

This test can be used for all kind of datasets[12], but for continuous data a discretisation is made. The data is divided into k disjount classes.

The teststatistic is

$$\chi^2 = \sum_{i=1}^{k} \frac{n_i - np_i}{np_i},$$

where:
n_i...number of data points in class i
p_i...probability, that one data point falls in class i

Under the $H_0 : F = F_0$ the teststatistic χ^2 is asymptotically χ^2_{k-1} distributed.

The H_0 is rejected, if $\chi^2 > \chi^2_{k-1;1-\alpha}$, where $\chi^2_{k-1;1-\alpha}$ denotes the $(1-\alpha)$-quantile of the χ^2_{k-1} distribution.

4.5.3 Q-Q Plot

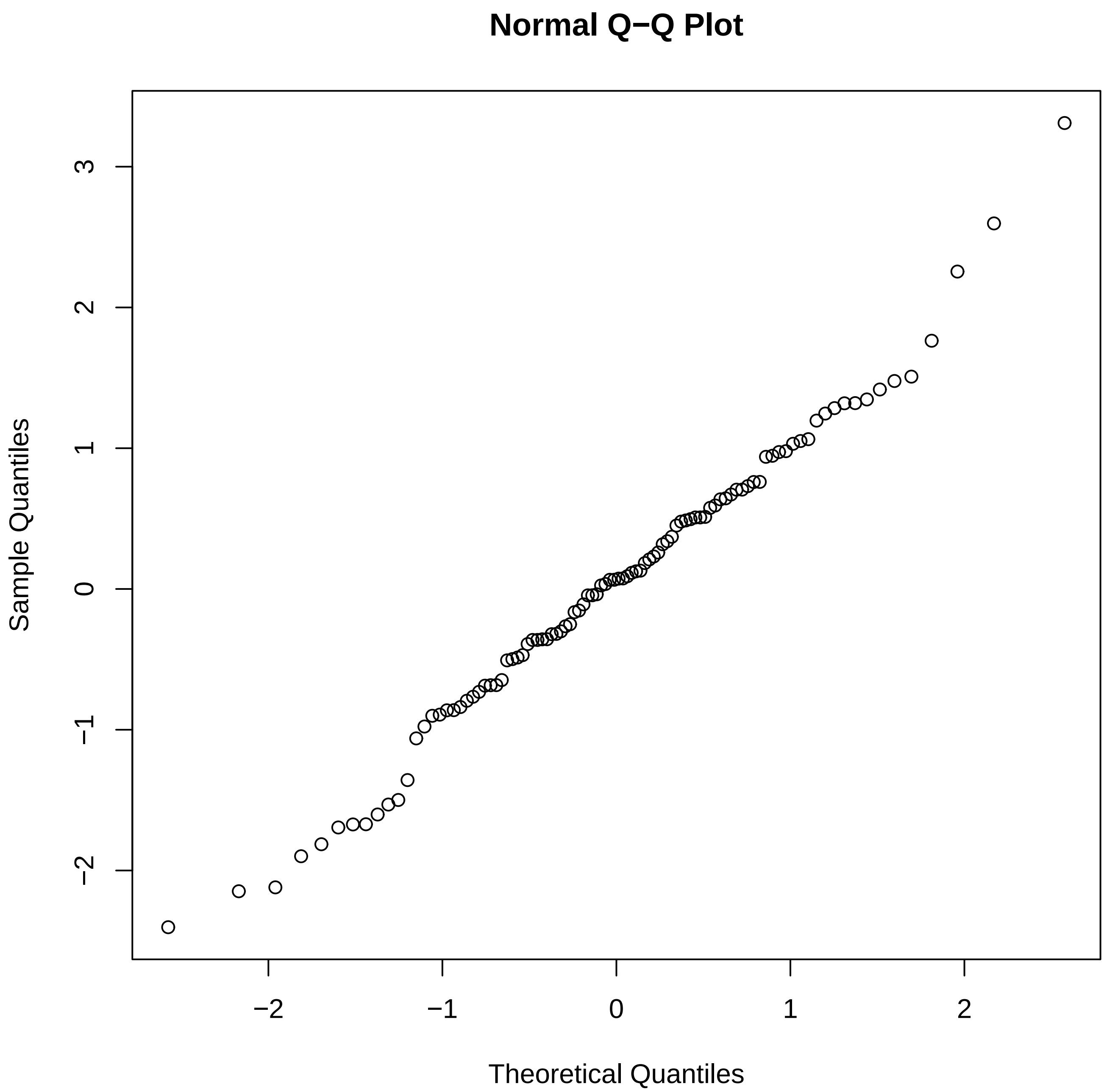
In order to examine if the log normal distribution is an appropriate loss severity distribution, a Q-Q Plot, which stands for Quantile-Quantile Plot, can be used.

If the distribution of X, $X \sim F$, is identical to a distribution G, then the empirical α-quantiles x_α should be identical with the theoretical α-quantiles g_α of the distribution G, too. If the empirical α-quantiles of $\hat{F}_n$ are plotted on the abscissae and the theoretical α-quantiles of the distribution G are plotted on the ordinate, then the points (x_α, g_α) should lie on the straigth line $x = g$.

The Q-Q Plot is often used for testing, wether a sample is normally distributed or not. Fig. 4.13 shows a Q-Q Plot of a normally distributed random sample. As you can see the points lie almost on a straight line just as expected.

[12] A rule of thumb for applicability of this test is that $n * p_i$ should be greater than 5.

Fig. 4.13 Q-Q Plot of a normally distributed random sample



Subsequent chapter five describes the four components, internal, external and scenario data and business enviroment and control factors, of an AMA in detail and highlights the differences and peculiarities of each data source. Focus will be on scenarios, which will be the main data source for the model (especially for the loss severity distribution).

5 Internal, External and Scenario Data

As already mentioned in chapter 3 the Basel Committee on Banking Supervision demands from a bank, using an AMA, to implement internal, external and scenario data. The question arises how to pool this data properly. Internal and external data have to be treated differently. External data, both public data and consortium data, tend to be skewed towards large losses, which leads to unacceptable high capital charges. But why is external data used? The reason is simple. Internal data is rather rare, especially in the range of high severity/low frequency data, and calibration only based on internal data may not suffice for computing an accurate capital charge. It is well known that high severity/low frequency (see fig. 3.2) events are those events which contribute the most to the operational risk capital charge.[4]

The next sections describe the differences and characteristics of the different data sources.

5.1 Internal data

Internal data is the most important data source for measuring operational risk, because its reliability is well known and this data reflects the specific loss profile of a bank most accurately. Internal data is the foundation of the capital charge computation. At least a five-year time series[1] of internal data is required.

Internal data is categorized into eight business lines and seven event types (see Annex 1 and Annex2), and herewith a data matrix with 56 cells can be developed. Each cell of the matrix contains the number of losses n_{ij}, which are categorized by business line i and event type j. This matrix is illustrated in fig. 5.1.

[1]For a bank, using an AMA the first time, a three-years time series is adequate.

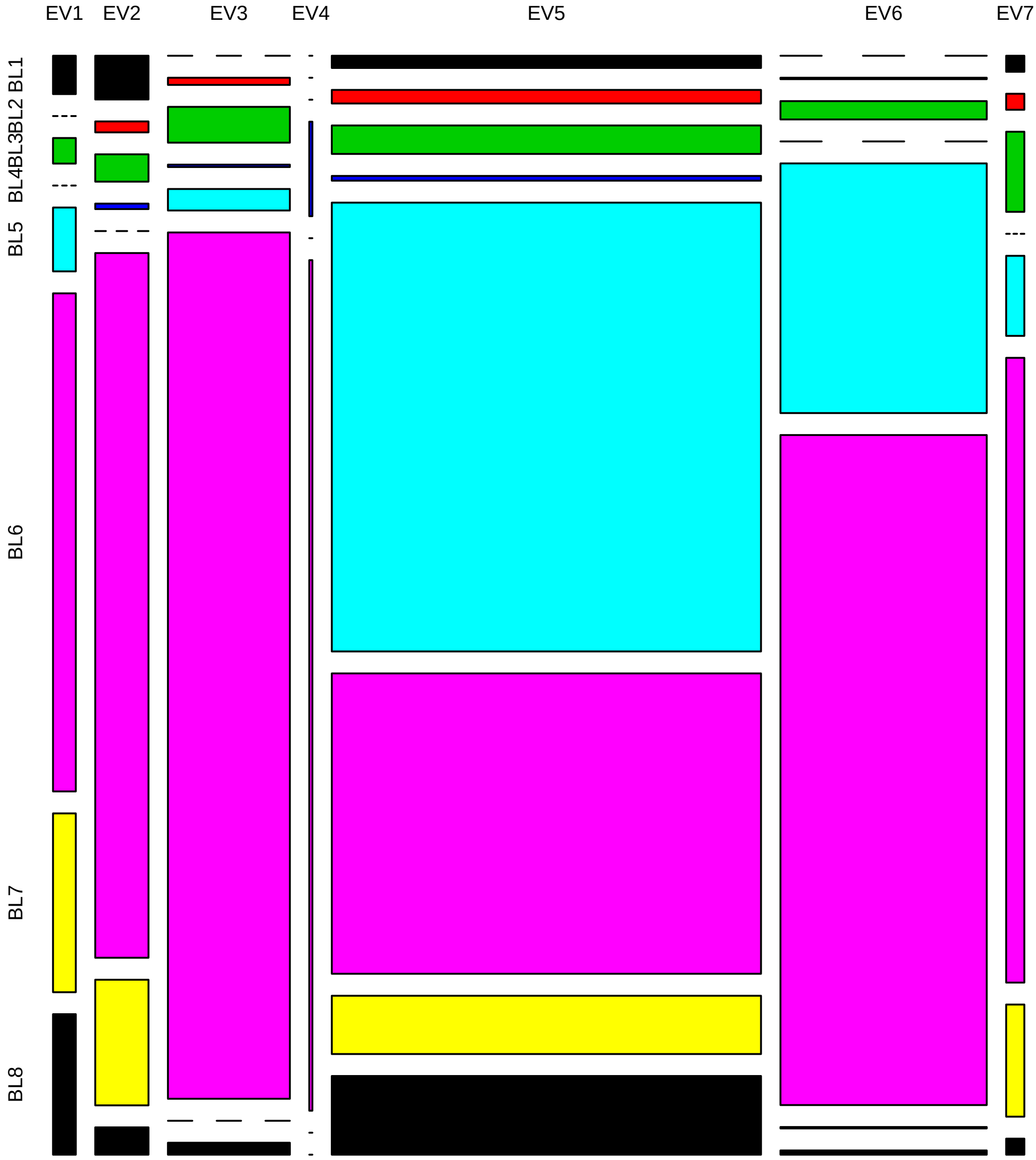
Fig. 5.1 Business Line/Event Type matrix

<table>
<tr><td></td><td></td><td colspan="7">Event Types</td></tr>
<tr><td></td><td></td><td>Internal Fraud</td><td>External Fraud</td><td>Employment and Workplace Safety</td><td>Clients, Products & Business Practices</td><td>Damage to Physical Assets</td><td>Business disruption and System Failures</td><td>Execution, Delivery & Process Management</td></tr>
<tr><td rowspan="8">Business Lines</td><td>Corporate Finance</td><td rowspan="8" colspan="7">n_{ij}</td></tr>
<tr><td>Trading & Sales</td></tr>
<tr><td>Retail Banking</td></tr>
<tr><td>Commercial Banking</td></tr>
<tr><td>Payment and Settlement</td></tr>
<tr><td>Agency Services</td></tr>
<tr><td>Asset Management</td></tr>
<tr><td>Retail Brokerage</td></tr>
</table>

Fig. 5.2 shows a mosaic plot of the number of entries per business lines/event types. Abbreviation for event type is EV, for business line BL. Most events fall into BL6, or if you classify into event types, into EV5.

The basic idea of the Advanced Measurement Approaches is that the LDA (see previous chapter) is implemented for every cell. Unfortunately not every cell has entries. Some n_{ij} are zero or are very small and therefore not suitable for modelling operational risk.

Fig. 5.2: Mosaic-plot of business lines/event types



For that reason some cells are grouped together. However one should have in mind that these cells should be homogeneous, and one has to decide which kind of clustering will be appropriate. For this purpose some cluster analysis (hierachical analysis technique: agglomerative technique)[2] was made by F. Piacenza et al., who came to the conclusion that the choice of event types as risk classes is preferable above that of business lines[3] [27]. Event types as risk classes are also used by [2] and [30].

Fig. 5.3 and fig. 5.4 show density plots of internal data of a specific bank first divided into business lines then into event types[4]. As you can see density plots of data divided into event types are more homogenous than those of data divided into business lines.

[2]Each class belongs to another, larger class. For more information see Annex 4.

[3]The exercise is based on UniCredit‘s data losses with accounting date between 01/01/2002 and 31/05/2006 and a loss exceeding EUR 5.000.

[4]Note that internal data is scaled, so confidential principles are not violated.

Fig. 5.3 Density plots of internal data divided into business lines

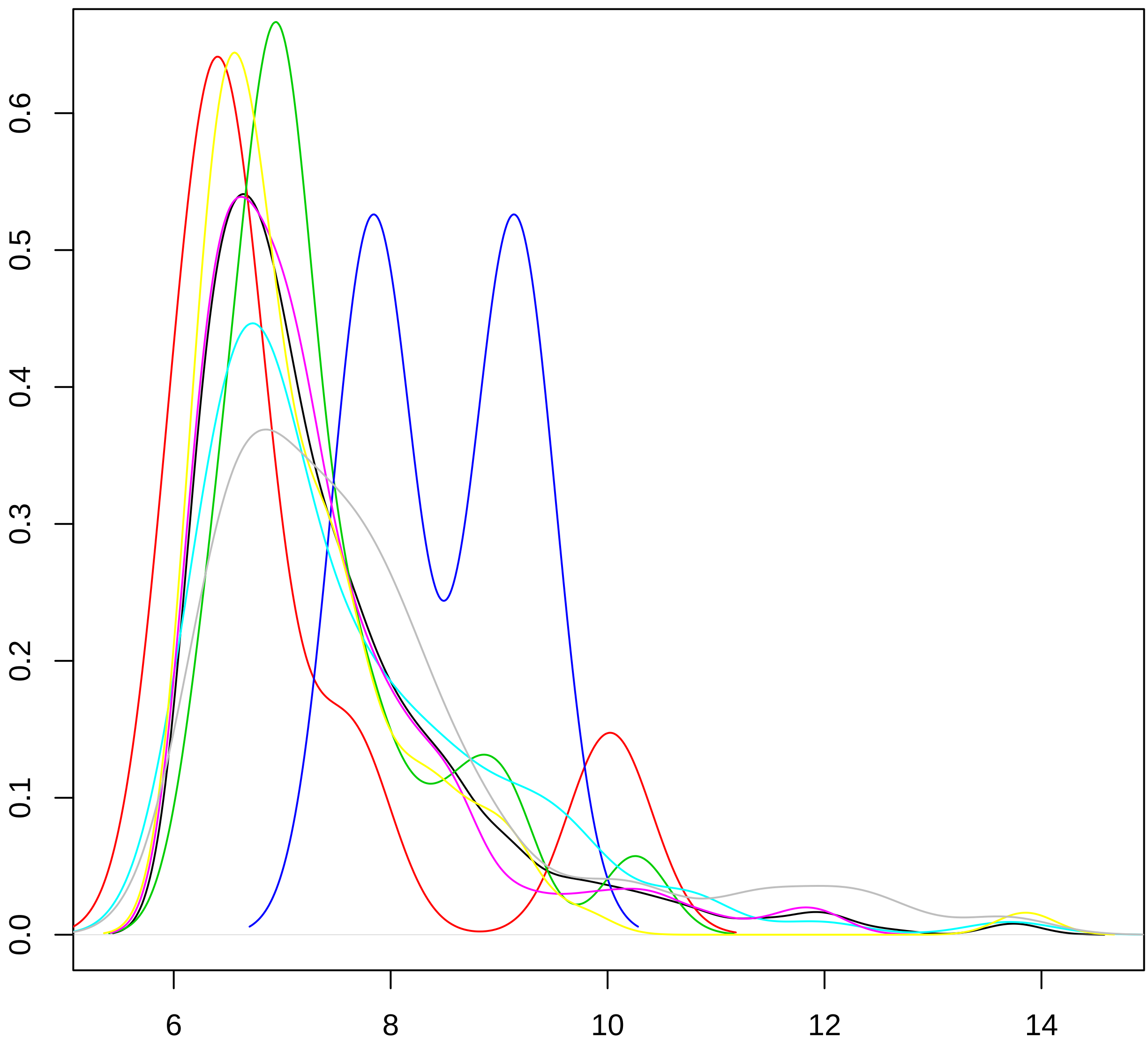
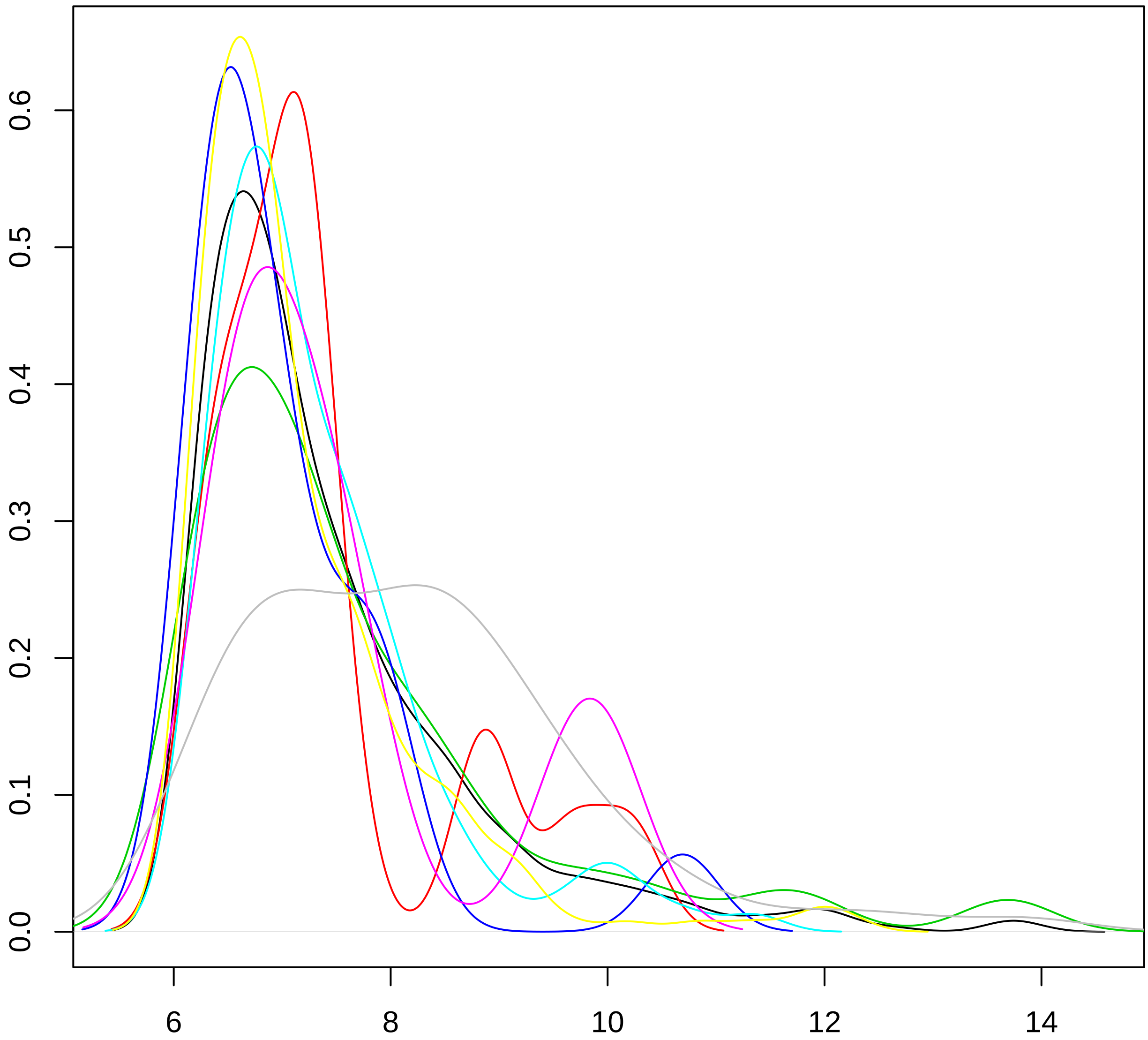


Fig. 5.4 Density plots of internal data divided into event types



Most banks concentrate on the categorization with event types. The advantage is that one specific event type gathers all losses with the same origin of cause, which is certainly a more consistent way of gathering data into consistent cells than the division into business lines.

That applies that seven LDA models are built and aggregated for the capital charge.

Nevertheless this categorization is not mandatory and some similar event types such as Employment Practices and Workplace Safety and Clients, Products & Business Practices for example can be pooled together for meaningful modelling purposes.

5.2 External data

Extremely high losses in every bank are so rare that no credible distribution in the tail can be obtained exclusively from internal data.
In general, usage of external data is a well accepted methodology in order to fill the gaps of an internal data base. Additionally this process enlarges understanding of operational risk exposure by benchmarking.

There are two different types of external data.

1. The first type of database records public losses, which are far too large and important, that they could be hidden from public eyes.

2. The second type is a consortium database with a systematical collection of losses over certain thresholds. The losses are made anonymous, so confidential principles are not violated. One important example of a consortium based database is ORX-Operational Riskdata eXchange.

The main difference between these two databases is the way losses are supposed to be truncated[5]. It is generally known that in the first case, as only publicly-released losses are recorded, the truncation threshold is expected to be much higher than in the consortium-based data.[3] (ORX records all losses above EUR 20.000.) Nevertheless also external data from a consortium of banks are pushed towards high losses, whereas internal data is biased towards small losses.

5.2.1 ORX - Operational Riskdata eXchange

Referending to the official webside of ORX, this consortium of banks is the world‘s leading operational risk loss data consortium for the financial services industry. ORX was founded in 2002 with the primary objective of creating a platform for the secure and anonymised exchange of high-quality risk loss data. [6]

[5] A defintion of truncation is given in section 4.4.

ORX is incorporated in Zurich, Switzerland and is owned by its members. ORX has currently 36 members (originally 12 member banks) all over the world:

- ABN Amro
- Banco Bilbao Vizcaya Argentaria
- Banc Sabadell
- Bank Austria - Creditanstalt
- Bank of America
- Barclays Bank
- Bank of Nova Scotia
- Banco Portugues de Negocios
- BMO Financial Group
- BNP Paribas
- Cajamar
- Caja Laboral
- Caixa Catalunya
- Commerzbank AG
- Credit Agricole
- Danske Bank A/S
- Deutsche Bank AG
- Dresdner Bank AG
- Euroclear Bank
- Erste Bank
- Fortis
- Grupo Banesto
- Grupo Banco Popular
- Grupo Santander
- HBOS plc
- ING
- Intesa SanPaolo
- JPMorganChase
- Lloyds TSB
- RBC Financial Group
- Skandinaviska Enskilda Banken AB
- TD Bank Financial Group
- US Bancorp
- Wachovia Corporation
- Washingtion Mutual
- West LB

Data provided by ORX are also divided into business lines and event types. The classification is not exactly the same as the one of Basel II. A Mapping of event types is necessary. See Annex 3.

To analyse external data density plots are made, just like for internal data. Here it is even more evident, that a classification by event types is much more satisfying. This is illustrated in fig. 5.5 and fig. 5.6[6].

[6]Note that ORX data is scaled, so confidential principles are not violated.

Fig. 5.5 Density plots of ORX data divided into business lines

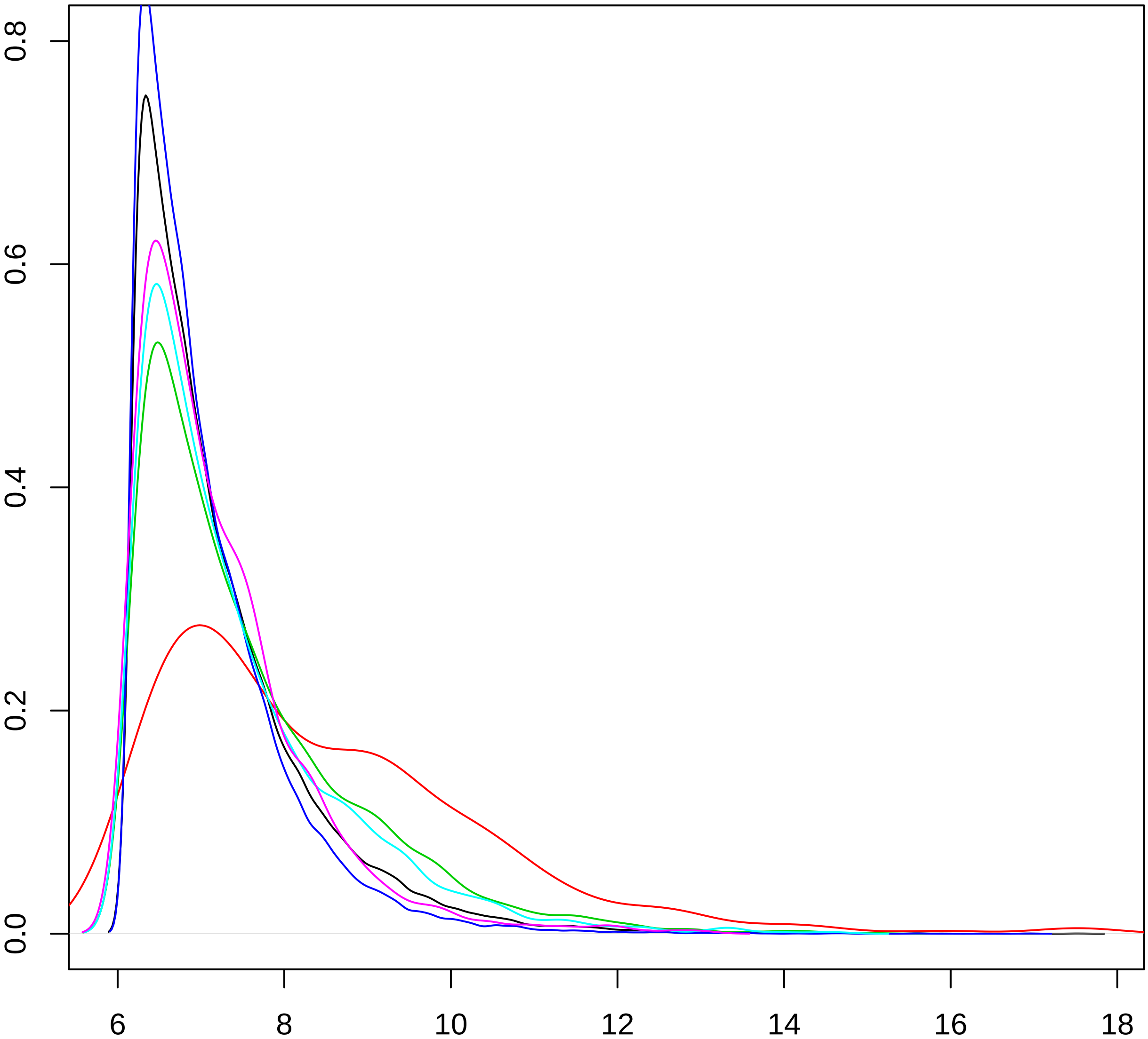
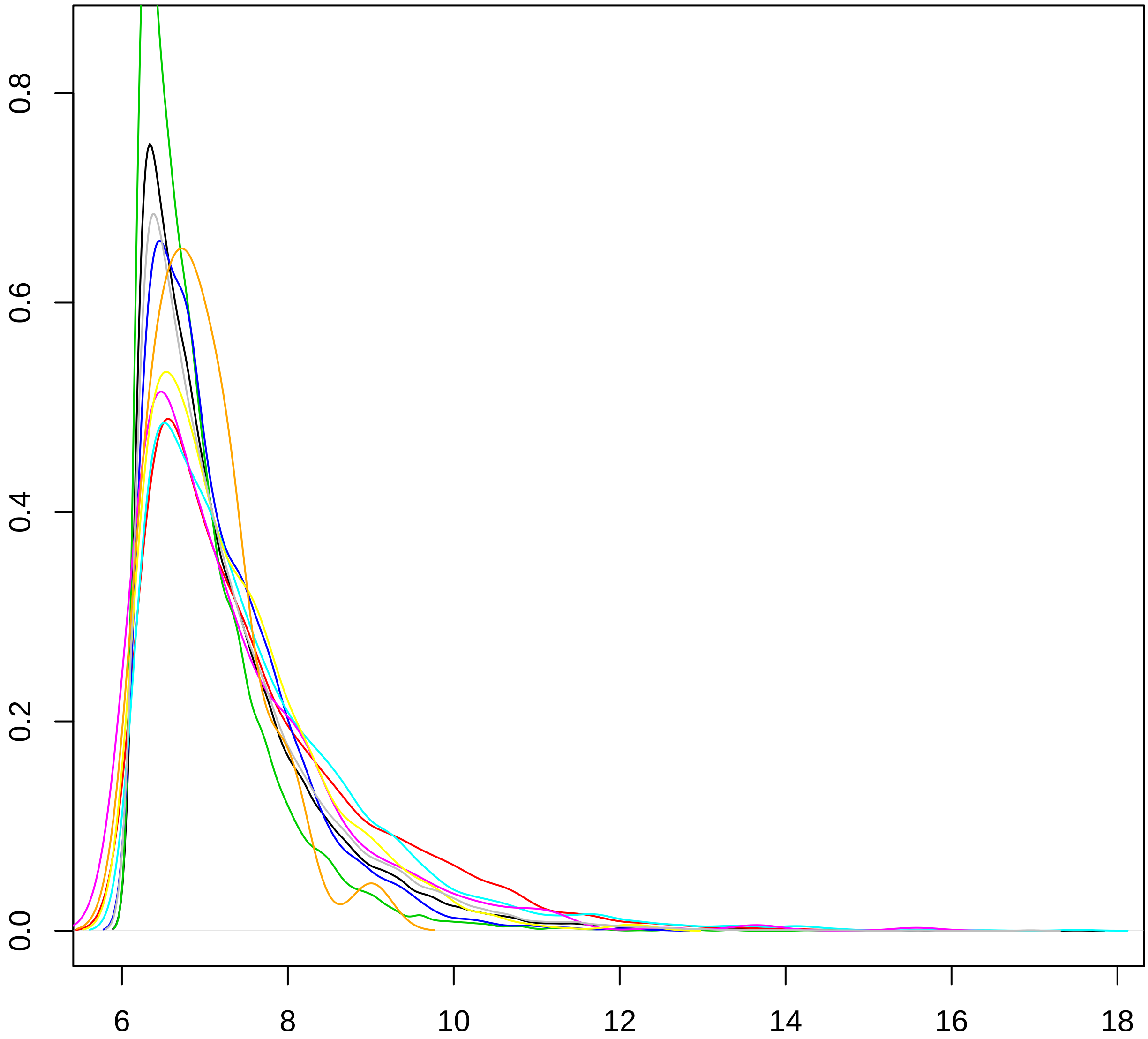


Fig. 5.6 Density plots of ORX data divided into event types



5.3 Scenario data

Internal and external loss data are "backward-looking", which imply they do not immediatly capture changes to the risk and control environment.[2]

This fact led banks to develop scenario analysis as addendum to internal and external data sources in order to model operational risk more accurately. Scenario data provides more accurate and future oriented estimations for a bank and its business units.

Defintion: *Although there is no standardized and mandatory definiton of* **scenario**, *scenario can be defined as sequence of possible events including description of possible developments leading to those events.*[9]

A scenario is something tangible that might happen in the future (i.e. a potential event). It can fill in the gaps in the internal data base by analysing other data sources or it might describe hazard incidents (very infrequent but severe) designed in discussions with experts. It is however clear that a set of representative scenarios should represent all risk factors for all risk types.

Generally, in order to answer the question what-if one has to focus on the future. *Focusing on future events, scenario analysis is an important addendum to loss data bases which are solely based on past events and do not regard future events.*[9]

According to the European Union Directives, [20] p. L177-178, a credit institution shall use scenario analysis of expert opinion in conjunction with external data to evaluate its exposure to high severity events. Over time such assessments need to be validated and reassessed through comparison to actual loss experience to ensure their reasonability.

Scenario analysis allows to understand the types and magnitudes of operational risk losses that have crucial impact on loss distributions and finally on capital charge requirement. It is necessary to develop clear procedures which determine a representative set of scenarios which takes into account all relevant risk drivers.

Developed scenarios are clearly determined by event type classes for easy implementation into a model, because internal and external data are categorized in event types, too. In order to answer the questions if certain events are going to happen or not, structured interviews take into account diverse information such as historic internal data, expert experience, external data and results of self assessments.
One of the objectives in structured scenario discussions is to analyze business environment and control factors, which are a natural sequence of risk assessment results, in order to determine an overall risk profile.

This approach in scenario analysis should ensure that modelling output is going to be stable over time. It must be assured that changes in the capital requirement are dependant on risk profile changes and that they are not driven by variations in the model.

5.3.1 Value of scenario analysis

Scenario analysis:

- promotes current understanding of OpRisk control
- uses all available data as input in structured scenario discussions with high-level experts
- is sensitive to internal and external changes of environment and collected data
- uses established statistical methodology for implementation of acquired estimations in the model
- is inherently forward looking
- generates potential losses that have not occurred in the past

This sensitivity of the scenario analysis is well suited to respond to any changes in business environment and internal control factors and actively supports risk management.

Scenario analysis and generation can be very valuable for the management of a bank. The results and outcomes can be used to reduce risk by improving and evaluating the quality of specific risk factors or controls. [28]

Any growth of a Bank can be reflected in increased frequency and severity of the estimates. On the other hand all improvement of controls or procedures in the bank can result in reduction of the frequency or severity estimates.

To increase transparency of the process and to raise awareness, a close involvement of experts from different organisational parts is necessary. So the flexibility of the whole process to adjust to the paricular needs of the bank is guaranteed. [28]

Being embedded in this way; scenario analysis contributes to meet Basel II test requirements.

5.3.2 Purpose of scenario analysis

Developed scenarios should capture all material sources of operational risk in a comprehensive way and they should cover all bank business activities and differences due to geographical locations.

Knowledge and experience of business managers and risk management experts have to be used to full capacity in order to derive reasoned assessments of plausible severe losses on one hand and on the other hand business managers should exchange their experience in structured discussions and maybe become even more aware of operational risk.

Applied procedures to generate scenarios have to be consistent across the bank and they have to be independently reviewed and validated.

Purpose of scenarios can be outlined as follows:

- scenarios should determine future potential losses not yet experienced in internal environment
- considering comparison with external losses what would be internal loss
- synthetic losses- fill in the gaps
- generate severity functions which relay on high-level expert opinions and structured discussions
- structured scenario discussions might have an additional role of bringing up experts together and consequently improving business environment and internal control systems
- discussions should include analysis of all possible sources of data and their possible outcome → listing of major risks

5.3.3 Scenario methodology

Scenario developing process includes a few steps which will be explained as following:

1. The first step is to analyse existing data and define input for creating certain scenarios. The challenge is to determine scenario methodology that can be applied across all organisational parts of the bank. Therefore, all scenario classes are categorized by event type. This approach allows a manageable amount of possible scenarios.

2. The next step is to review data quality and identify gaps in order to design scenarios. This try of identifying scenarios is done by operational risk experts who analyse all data sources and make suggestions for structured scenario discussions. This is achieved using blend of information such as: internal historical losses, relevant external banking industry experience, relevant risk factors and the control environment reported during risk assessments.

   It is well known that historic data can be successfully used for determining frequency of events [28]. In order to determine the scope of the scenario it is very important to answer "what - if" questions which is soon after going to lead us to the severity and frequency of potential events.

3. Quantitative operational risk experts analyse and prepare severity and frequency estimations after having analysed all data sources.

4. All data and parameters are presented in the structured scenario discussions with high - level experts and give output for modelling.
In those structured and guided discussions high level experts are asked to give their knowledge and experience and estimate the potential loss frequency and severity for a specific scenario. In principle, they are asked to adjust estimated numbers prepared and analysed by operational risk experts.
This comprehensive and consistent process assures quality of obtained scenarios and subsequently quality of the risk profile required for modelling risk capital requirements.

5. Output of scenario discussions are incorporated into model .
If it is assured that data resulting from scenario discussions have a good quality, determined parameters can be fed into model.

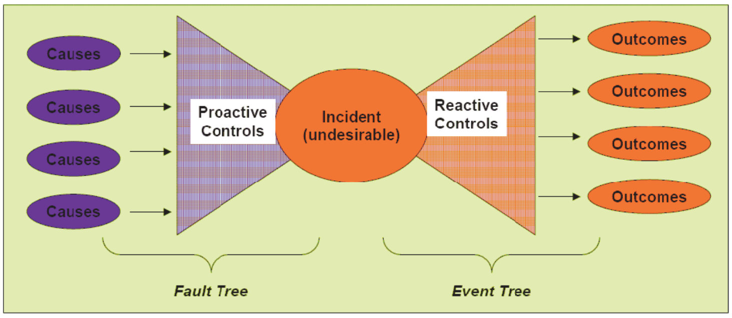
5.3.4 Bow-Tie Diagramm

A graphical analysis of scenarios is the so-called **"Bow-Tie diagramm"**[18]:

- pre-events (also called causes): potential events leading to an undesirable incident
- proactive or preventive controls: actions taken to reduce the likelihood of an undesirable incident occurring
- incident: an event which might cause an undesirable outcome
- reactive (mitigating) controls: actions taken to reduce the impact of an undesirable incident
- outcomes: potential result of an undesirable Incident

Fig. 5.7 illustrates the components of a Bow-Tie diagramm.

Fig. 5.7 Components of a Bow-Tie diagramm (source [18])



According to [18] the bow-tie diagram is a helpful tool for making risk management assumptions, analysis and conclusions explicit.
The main objective of the diagram is to analyse not only severity and frequency of potential losses that can be suffered, but also business environment and control factors that are reflected in the proactive and reactive controls, as well as causes that lead to the event itself.

The left hand part of the diagram is called fault tree and describes the events or faults which might give rise to an undesirable incident. The right hand side is called an event tree which analyses consequences of undesirable incidents. [18]

5.3.5 Validation of scenarios

The validation process should assure that the scenario template is reviewed on annual basis and that all changes in the business environment are considered. These changes might be reflected in frequency of certain scenarios, i.e. if some improvements of proactive control or similar changes took place or if there have been some less advantageous events that might increase frequency of some scenarios.

Procedures for validating scenarios:

- Comparison of severe losses internal vs. scenario
  It is of great importance to validate if the scenario frequency projections match the internal annualized loss experience particularly in the tail.
  According to [30] it is expected that scenarios over a certain amount limit (i.e. EUR 10 mio.) are greater than actual loss experience.

- Distribution curve
  The next step in validating scenarios is to analyse the distributions of losses and to look if they match actual loss experience.
  The expectation by [30] is that distribution curves of actual losses have a more volatile profile.

- Maximum loss
  The validation contains analysis of the influence of maximum loss data, either internal or external, on scenario models.
  It is important that those data should influence, but not dictate scenario input models. [30]

- Geographical, demographical, economical and other regional specialities
  It is generally known by experts that the coverage of all local specialities must be fulfilled.[1]

5.4 Business enviroment and control factors

LDA models mainly rely on loss data and are inherently backward looking. It is therefore very important to incorporate components that reflect changes in the business and control environment in a timely manner.

There have been intensive discussions in the banking industry regarding different strategies that might fulfill this demand, however it came out that key risk indicators (KRI) and risk self-assessments would be the common method to accomplish them.

These qualitative adjustments do not occur directly in the model, but through adjustments of the parameters of the loss severity distribution.[7] Direct application would be difficult to justify with statistical means.

The following chapter provides one possibility to mix the different data sources for the Advanced Measurement Approach mentioned above. This mixing is the main challenge when it comes to operational risk measuring.

[7]For detail see section 6.3.

6 Mixing Internal, External and Scenario Data

The AMA is the most sophisticated option to quantify the capital charge for operational risk. The model explained in this section is mainly driven by scenario analysis. Data used in scenario analysis are blended of internal and external data and risk assessment.

6.1 Model principles

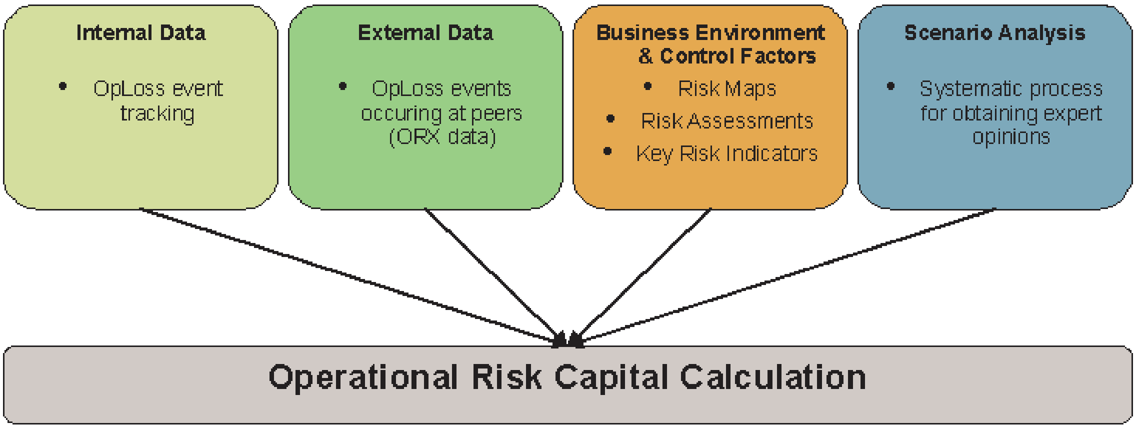
The underlying model:

- reflects the operational risk profile of the specific bank
- is able to evaluate the risk at any given confidence level
- calculates expected and unexpected loss
- is responsive to the changes in the business environment
- is sensitive to the large losses
- is stable to statistical outliers
- is understandable for the management
- is able to allocate the risk to subentities

The following sections describe the process of implementing an AMA step by step from mixing different data sources to generating the parameters for the loss distribution and the subsequent computation of the capital charge.

6.2 Mixing of data

Fig. 6.1 Components of the operational risk capital charge calculation



In order to mix all different data sources, shown in fig. 6.1, some statistical analysis has been made to compare internal and external data and in which way these sources can be implemented in scenario analysis.
Such statistical analysis contains of the comparison of density functions of several event types EV and business lines from both external and internal data, see previous chapter. Here regard must be paid to different thresholds of each data source.
To achieve significant outcomes all data must be on the same scale and above the same certain threshold. If ORX data is used, this threshold will be EUR 20.000, because in general banks collect data above a smaller threshold.

The analysis comes to the conclusion, that external data and internal data cannot be treated the same way. External data, mainly based on ORX data, consists of data from some very large banks. The loss profile of an average bank will completely differ from the loss profile of one of these large banking groups. And as already mentioned external data is strongly biased toward high-severity events.

According to [12], frequency and severity data must be treated differently as they raise rather different issues.
A mixture of internal and external data is only used for the loss severity distribution above a certain threshold t_1. Therefore the following computations are only made with data above this threshold. First of all the scaling method and the scenario generation will be introduced.

6.3 Scaling of internal and external data

One possiblity to mix internal and external data is scaling. One scaling factor, which turned out to be appropriate, is gross income GI[1]. Gross income is an indicator of a bank's risk exposure. Large banks with a large gross income tend to be more exposed to risks than smaller banks with less banking activities.

As already mentioned only data above a certain threshold t_1 is used for the mixing, which will be the input for the parameter generation of the loss severity distribution.

Method of scaling with gross income as indicator:

1. Gross income of ORX is computed per business line. The same is made for the specific bank for which the capital charge should be calculated. Unfortunately there is no information about the gross income per event type, which would be more appropriate, because the capital charge computation and aggregation is made by event type.

2. The next step is to calculate the number of losses per business line and event type for ORX data and internal data. (See matrix defined in chapter 5 p.46.) So two matrices are developed. Note that, to compare these two data sources, a mapping of event types and business lines is necessary. See Annex 3.

3. The actual scaling process for each event type is the following:

$$\frac{\text{\# of losses of ORX per business line i}}{\text{GI}_i \text{ of ORX}} = \frac{\text{X}_i}{\text{GI}_i \text{ of the bank}}$$

   The table in fig 6.2 tries to illustrate this concept:

Fig. 6.2 Table of scaling for one event type

	# losses ORX	GI ORX	# losses bank	GI bank	# scaled by ORX
BL 1					X_1
BL 2					X_2
BL 3					X_3
BL 4					X_4
BL 5					X_5
BL 6					X_6
BL 7					X_7
BL 8					X_8
Sum					$\sum_{i=1}^{8} X_i$

[1]Defined in chapter 3 p.10.

4. The $\sum_{i=1}^{8} X_i$ is the new number of losses in this event type above the certain threshold t_1. This number is splitted into k loss amount buckets. The splitting factors p_l , $l = 1, ..., k$, are given by ORX data. The splitting factor for the first bucket is the percentage of ORX data, whose loss amount falls into this first bucket and so on. A table of the splitting factors is presented in fig. 6.3.

   Bucket boundaries may be t_1, 1 million, 5 millions and 10 millions (here $k = 4$).Granularity of buckets is also possible on the higher level. The bucket boundaries depend on the specific internal data and risk profile of each bank.
   This splitting is done for each event type j , $j = 1, ..., 7$.[2] This concept is illustrated in fig. 6.4.

Fig. 6.3 Table of percentage of ORX data in each bucket

	t_1 - 1 mio.	1 - 5 mio.	5 - 10 mio.	>10 mio.	Sum
EV 1	p_{11}	..	..	p_{14}	
EV 2	.			.	
EV 3	.			.	$\sum_{i=1}^{4} p_{jl} =$
EV 4	.			.	100%
EV 5	.			.	for $j = 1, ..., 7$
EV 6	.			.	
EV 7	p_{71}	..	..	p_{74}	

Fig. 6.4 Splitting into buckets for one event type j

	t_1 - 1 mio.	1 - 5 mio.	5 - 10 mio.	>10 mio.	Sum
# scaled by ORX	$[\sum_{i=1}^{8} X_i] \cdot p_{jl}$				$\sum_{i=1}^{8} X_i$
# bank	[$\sum$ of # losses bank] $\cdot p_{jl}$				Sum of # losses bank

   where p_{jl} , $l = 1, ..., 4$, is the percentage of ORX data in each bucket for one event type j.

5. This table, given by fig. 6.4, is the basis for a discussion with experts and management, who decide wether scaled ORX data, internal data or a mixture of both describes reliably the current risk situation. As support for making this decision some special cases of loss events (severe losses, which illustrate the potential problem) with detailed descpription and cause, results and outcomes of risk assessments and KRIs are presented to experts and management.

[2]Several EVs may be merged. Number of risk cells depends on data availability.

After validation and correction through experts of operational risk, the number of losses in these buckets n_{jl} , $j = 1, ..., 7$ and, $l = 1, ..., k$ (here k=4), are the foundation for parameter estimation of the loss severity distribution. The detailed technique is presented in section 6.4.

6.4 Generating parameters for the loss severity distribution

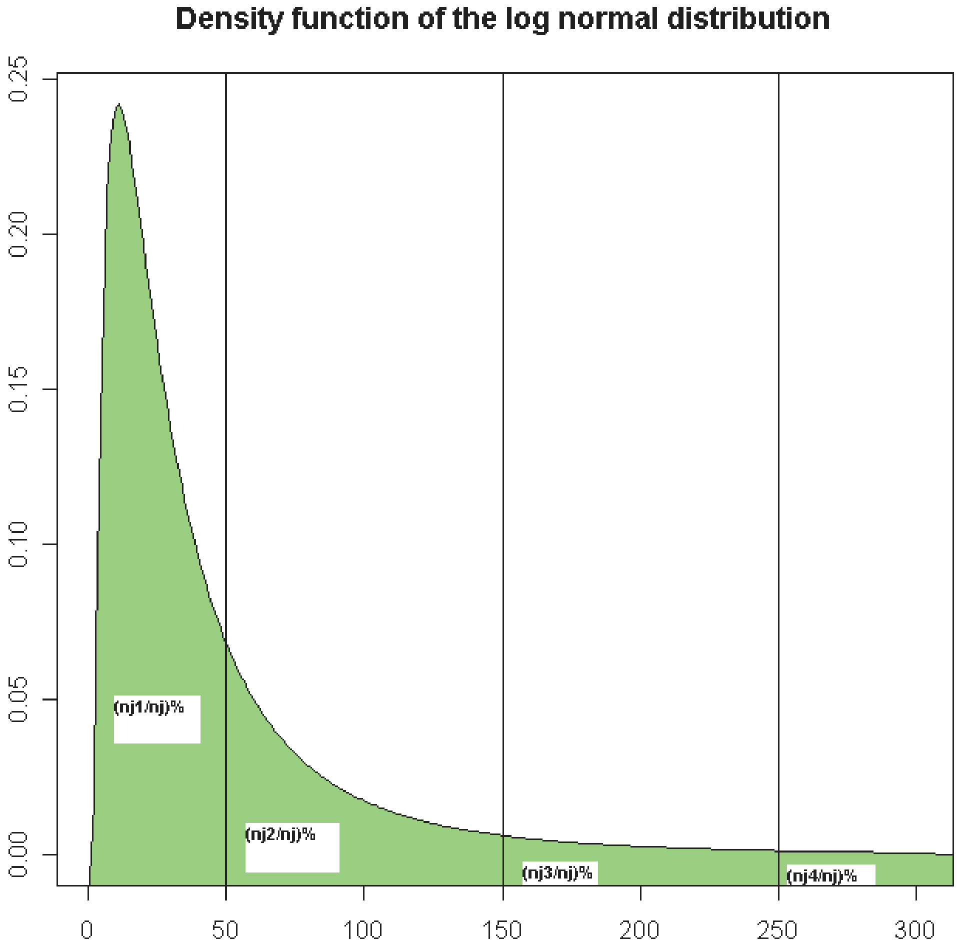
After validation through experts and senior management following numbers
n_{jl} , $j = 1, ..., 7$ and, $l = 1, ..., k$ (here k=4), are used for generating scenario data parameters for the loss severity distribution, respectively see fig. 6.5. As already mentioned in general the loss severity distribution is not fitted from zero to infinity ($t_1 > 0$ alternatively $t_1 >$ collecting threshold). A piece-wise defined loss severity distribution is fitted. See section 4.4. For the range from zero (from the collecting threshold respectively) to t_1 the empirical distribution of loss severity is used.
This fitting is made by event type. This means, that the empirical distribution function is only estimated by the specific data for the corresponding event type. The same holds for the log normal distribution. Parameter estimation is made for each event type. For illustration of this concept see fig. 6.7 and fig. 6.8.

Fig. 6.5 Table of number of losses per bucket and event type

	t_1 - 1 mio.	1 - 5 mio.	5 - 10 mio.	>10 mio.	Sum
EV 1	n_{11}	..	..	n_{14}	
EV 2	.			.	
EV 3	.			.	$n_j = \sum_{l=1}^{k} n_{lj}$
EV 4	.			.	for $j = 1, ..., 7$
EV 5	.			.	
EV 6	.			.	
EV 7	n_{71}	..	..	n_{74}	

The main idea of the parameter estimation is to fit a distribution that fulfills the splitting best. This means that (n_{lj}/n_j)% of the area beneath the density of the loss severity distribution of one event type j should lie in the bucket l, here $l = 1, ..., 4$. In other words the probability that a loss event causes an loss amount of y, where for instance $t_1 \leq y <$ 1mio., should be n_{j1}/n_j for one event type j, because y falls into the first bucket. Fig. 6.6 illustrates this concept.

Fig. 6.6 Density function with four buckets for one event type



This fitting is made by optimization. The method of choice is BFGS, which stands for Broyden-Fletcher-Goldfarb-Shanno, a quasi-Newton method. This method is used to solve an unconstrained nonlinear optimization problem.[3]

For the loss severity distribution the log-normal distribution was chosen. This choice is a result of the Kolmogoroff-Smirnoff tests. Computations were made for the Generalized Pareto distribution, too, but for this distribution unreasonable high capital charges were calculated. (A bank cannot loose more money than the worth of the whole institution.)

Therefore the parameters, which are to be estimated are mean and variance of each event type j , $j = 1, ..., 7$, of the log-normal distribution. The method for measuring the "difference" of (n_{jl}/n_j) and the theoretical probabilty of a bucket (intervall), according

[3]For information about optimization see [8] or [19].

to the fitted log-normal distribution, is a least-squares estimator. Following function has to be minimized:

$$(\mu^*, \sigma^{2*}) = \arg\ \min\{\textstyle\sum_{l=1}^{k}[(F(X_{l+1}) - F(X_l)) - \frac{n_{jl}}{n_j}]^2\},$$

where $F(\cdot)$ denotes the distribution function of the log-normal distribution with parameters μ and σ^2 and, X_l , $l = 1, ..., k$, are the bucket boundaries mentioned above.[4]

Starting points for μ and σ^2 for the optimization are the empirical mean and the empirical variance of all internal loss data on a logarithmic scale.

[4]Optimization is made with the free statistical program R [10].

Fig. 6.7 Piece-wise defined loss severity distribution for two pooled event types

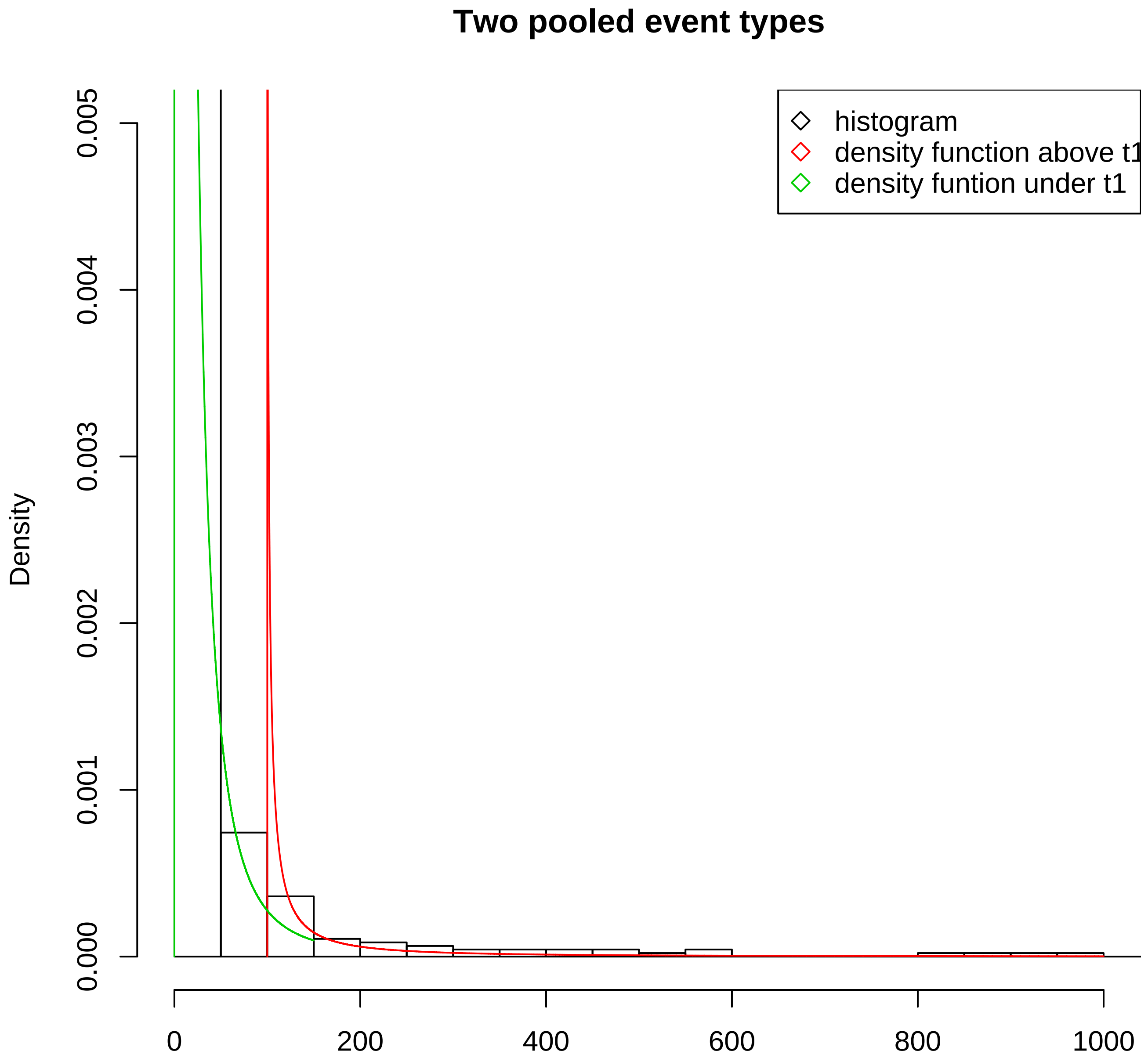
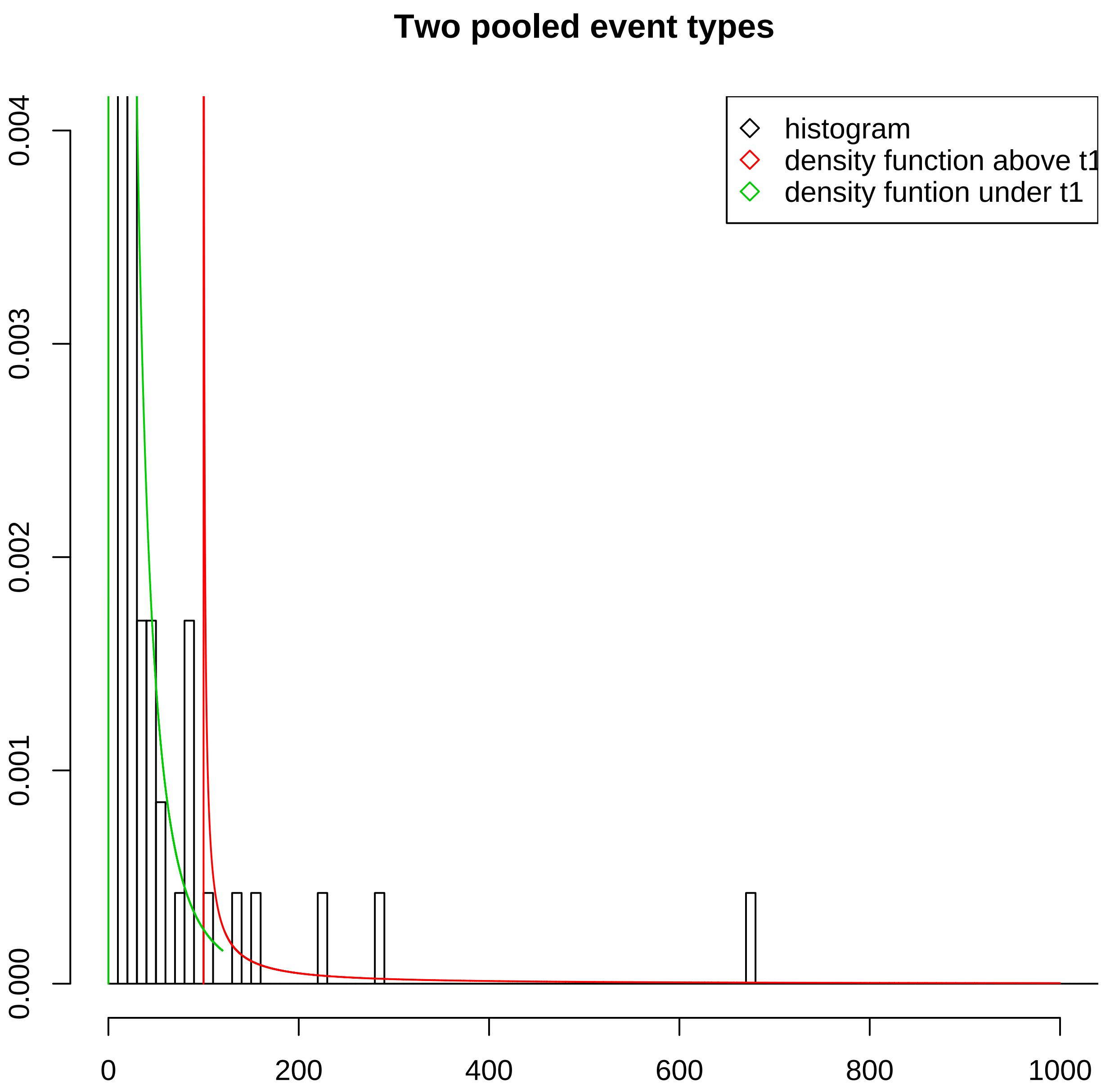


Fig. 6.8 Piece-wise defined loss severity distribution for two pooled event types



This graphical analyse is a very helpful tool for finding the right value of t_1, too. As one can see here $t_1 = 100.000$[5]. However another range between empirical and log-normal loss distribution function would be more appropriate. Here for example $t_1 = 120.000$ or $t_1 = 150.000$. One possibility is to fit the log-normal distribution function from $t_1 = 100.000$ all the time, but to achieve a smoother transition the empirical distribution

[5]Data is divided by 1000.

function is used for calculation till $t_1 + x^6$ and the log-normal distribution function is used from $(t_1 + x) - \infty^7$.

6.5 Loss Frequency distribution

For the loss frequency distribution internal data is the only data source, because this data reflects the number of loss events occuring during one year for a specific bank most accurately. It is well known that considering one specific bank, its internal frequency data likely convey information on its specific riskiness and the soundness of its risk management practices.[12]

After some analysis, with focus on the right choice of the loss frequency distribution, the Poisson distribution was chosen. For this analysis the number of events were categorized into years and months, which leads to a sample of numbers, that denotes the number of events per month and/or year.
Note that december data should be excluded, because in december more events usually occur than in other months. This peculiarity of operational risk data became evident during the analysis. This happens not by reason of more fraudulent people in december, but of the way information is monitored. In december many events are reported, but actually these events happened earlier that year (late reporting).

With this sample of numbers a χ^2-test is made to test the hypothesis if this sample is Poisson or negative binomial distributed. The binomial distribution is excluded, because the mean of numbers per month always exceeds the variances of numbers per month.

The conclusion of these tests was, that the hypothesis that the poisson distribution would be a good choice, cannot be rejected. Besides it turned out that it is almoust irrelevant for the computation of the capital charge if poisson or negative binomial distribution is used [2].

The choice of severity distributions for modelling operational risk data by using LDA usually has a much more severe impact on the capital charge than the choice of the frequency distribution. [2]

Now the Poisson distribution is fitted separately for each event type. The parameters λ_j , $j = 1, ..., 7$, of the Poisson distributions correspond with the number of events categorized into the corresponding event type divided by the time horizon.

[6] x maybe 20.000 or 50.000.

[7] Empirical tests have shown that it does not make any difference for the capital charge computation wether t_1 or $t_1 + x$ is used as boundary between empirical and log-normal distribution function.

$$\lambda_j = \frac{\text{number of events in event type j}}{\text{horizon}},$$

where the horizon is the length of the time-series (at least 3 years).

Because of the piece-wise defined loss severity distribution, the loss frequency distribution must be fitted (for subsequent Monte-Carlo simulations) piece-wise, too. This fact requires two lambdas, one estimated for "small" losses (loss amount from 0 or collecting threshold - t_1) - λ_{1j} - and the other one estimated for "large" losses (loss amount from $t_1 - \infty$) - λ_{2j}.
The adaption of the above formula for λ_j is quite easy. The number of events per event type j must be divided into "small" and "large" losses and then the calculation can be done as described before.

6.6 Compound loss distribution

The actual loss distribution is achieved by Monte-Carlo simulation, introduced in chapter 4.

Practical implementation of Monte-Carlo simulation:[8]

1. After generating all parameters for the log normal and the Poisson distribution a random number r_1 of the Poisson distribution with parameter λ_{1j} is drawn.

2. Then r_1 losses of the empirical loss severity distribution are taken. This r_1 losses are the first part of the loss sample r.

3. r_2 is generated from the Poisson distribution with parameter λ_{2j}.

4. r_2 losses are generated from the log normal distribution with parameters (μ^*, σ^{2*}). This is the second part of the loss sample r.

5. These four steps are repeated $k = 10^6$ or $k = 10^7$ times and sample mean and α-quantiles of all simulations r_k are calculated. This sample mean is the expected loss. To get the unexpected loss of a certain level α the sample mean is subtracted from the certain α-quantile.

6. For computing the deviation of expected and unexpected loss the method of bootstrapping is used. The generated samples r_k are merged together and resampled m times (at least 10.000 times).[9] At all times mean and α-quantiles are calculated. So m means and m α-quantiles are computed. The deviation of the m means is the deviation of the expected loss and the deviation of the α-quantile is the deviation of the unexpected loss.

[8] Simulations are made with the free statistical software R.[10]
[9] Computation time may rise very fast with increasing m.

This simulation is made for each event type j. The capital charges of each event type are accumulated. This sum is the capital charge for the specific bank institute.

Further challenges and issues of operational risk, which are not already covered in the model presented and developed above, will be introduced in the next chapter.

7 Future challenges

Since operational risk is a rather new field in the bank‘s risk management, developments and challenges are growing constantly in this sector.

7.1 Correlation Problem

One of these challenges is to model diversification effects in operational risk, which is a very crucial task to do. Generally, operational risk events might be, at least partially, decorrelated. According to [13], perfect correlation across risk types or business lines, which means, that capital charges by risk types or business lines are summed, may lead to a much higher capital charge than in the Standardised Approach proposed by Basel II.

Basel II statement:

Risk measures for different operational risk estimates must be added for purposes of calculating the regulatory minimum capital requirement. However, the bank may be permitted to use internally determined correlations in operational risk losses across individual operational risk estimates, provided it can demonstrate to a high degree of confidence and to the satisfaction of the national supervisor that its systems for determining correlations are sound, implemented with integrity, and take into account the uncertainty surrounding any such correlation estimates (particularly in periods of stress). The bank must validate its correlation assumptions. [24] p.126

Let L_1 and L_2 be two aggregated losses and L the global aggregate loss:

$$\begin{aligned} L &= L_1 + L_2 \\ &= \textstyle\sum_{i=1}^{N_1} X_i + \sum_{j=1}^{N_2} Y_j \end{aligned}$$

There are two types of correlation:

1. **Frequency correlation:** Two annual frequencies N_1 and N_2 are said to be correlated, if those two variables are not independent. For instance when the number of internal fraud events is high, then the number of external fraud events is high, too.

2. **Severity correlation:** Loss X_i, randomly drawn from the first class of events, and loss Y_j, randomly drawn from the second class of events, are not independent

with one another. This could occur if internal fraud loss amounts are high when external fraud loss amounts are high. This could be empirically shown: If the mean loss amounts are correlated over time, this is an indicator for severity correlation. The difficult matter of severity correlation is the assumption of simultaneously severity independence within each risk type and severity correlation between two risk types.

The next step is to implicate those two types of correlation into the extisting model. Adding frequency correlation to the LDA-model is an easy task and does not destroy the derived model. In fact it does not change the capital charge computation. Modelling severity correlation needs a whole new family of models. Such an extension is not reasonably feasible, but developments concerning the correlation problem must be kept in view.[1]

7.2 Insurances

A further challenge would be to take insurances and their risk mitigating impact for the capital charge under consideration.
Given a typical operational risk model (such as the one developed in the previous chapter), which samples loss frequency and loss severity (via a LDA) independently, the main difficulty in incorporating existing insurance policies into the model is to determine the amount of losses that are actually covered under the insurance premium.
From this point on, Monte Carlo simulation and aggregation of operational VaR becomes fairly straightforward.
A typically used approach is to assume a fixed percentage of incurred losses to be insurance claims. The critical part, of course, is estimating the correct percentage.

Relying on expert opinion may not be practical in this case as precisely unforeseeable events have to be handled estimating the probability that such an unforeseen event not only occurs but is also covered by any broad-coverage insurance policy is close to impossible. Relying on historical data only might also not be functional as it is unlikely that very large losses have already occurred and assuming that the breadth of coverage by the insurance transfers over perfectly from small losses to very big ones may be false.

A more sophisticated premium-based approach would be the following cited from [26]:
The basis for this approach is the assumption that the negotiated insurance premium is a price for a commodity traded on a more or less free market and thus must roughly represent a fair price.
Of course there are only a limited number of players in the insurance market. Nonetheless it is fair to assume that there is enough competition to guarantee that premiums

[1]For concept in detail see [13].

negotiated by as large a player as the EB AG are not inflated above market average.

This approach has several advantages:

- Calculations do not rely solely on historical data.
- Calculations can be performed for all past, active or future insurance contracts.
- Results can be verified with expert opinion and historical data.

As stated above, the starting point for a premium-based calculation of insurance coverage is the assumption that the negotiated insurance premium represents a fair market price. In other words: The negotiated insurance premium will cover the insurers expected loss, as well as the proportionate amount of administration and overhead costs.

Given the distribution of loss severity (in the developed model, this would be the log-normal distribution),the insurers expected loss is calculated as follows:

$$EL = \int_0^\infty v(x)\lambda_{disb} dF(x) \ ,$$

Where λ_{disb} denotes the number of disbursed claims the insurer expects, $v(x)$ is the amount paid by the insurer in each of these cases and $F(x)$ is the assumed distribution of loss severity (log-normal).

The premium covers not only the expected loss but also overhead and administration costs.

The typical ratio of expected loss to received premiums in non-life insurance is roughly 2/3 and is fairly constant over time. Competition ensures that the ratio of expected loss to the so-called administration ratio - AR remains stable.[2]
Using this relationship the following implicit relationship between the insurance premium and the number of disbursed claims that the insurer may expect can be derived by:

$$EL = \int_0^\infty v(x)\lambda_{disb} dF(x)$$

$$EL = p * (1 - AR)$$

$$p = \frac{1}{1-AR} \int_0^\infty v(x)\lambda_{disb} dF(x)$$

While this approach may not be perfect, it is clear to see that it provides more accurate results than the aforementioned other estimates, since it relies not on the information the insured has access to but on the implicit expertise of the insurers model, which it is safe to assume contains all possible information about the insurance contract.
In other words, the insurance premium contains implicit information about the frequency of insurance claims, which is exactly the information needed to incorporate insurance

[2]This was estimated with data from Statistical Office of the European Communities.

coverage into a LDA model.

These results from above are incorporated into a Monte Carlo simulation as follows: Insured and uninsured losses are generated separately, according to their respective frequencies. After a simulation of the loss severity, the net loss for insured losses is adjusted according to the insurance policies deductible and limit (information provided by v(x)), if applicable. Afterwards net losses are pooled together and aggregation of the global loss distribution as well as calculation of operational VaR may be performed as described in chapter 6.

8 Conclusion

This work attempts to provide the statistical framework for modelling operational risk using an AMA. Additionally, it developed a model for computing the capital charge of a specific bank step by step from the pure database to the calculation of the operational risk capital charge. The main question of this LDA-based approach was how to mix all different data sources, internal and external data, scenarios and KRIs, properly and how to ensure stability of the outcome.

A further question was which kind of splitting of data would be appropriate. Data cells should be homogenous to receive stable, robust and reasonable results, but on the other hand cells must contain a certain amount of events, because otherwise generated parameters and results may vary enormously. This would be contrary to the assumption of robustness and stability. The conclusion raises, that splitting into event types would be the most appropriate solution. Almost each event type has enough data entries to perform statistical analysis. Those event types with less events are pooled together with respect to a certain consistency of data. For example it is suggested, that the event type Employment Practices and Workplace Safety could be pooled with Client, Products & Business Practices, because all events have cause in legal issues.

The choice of the right loss severity distribution and loss frequency distribution was also a mentioned task. Whereas this question was easy to answer when it comes to the loss frequency distribution (Poisson distribution was chosen, according to empirical analysis and literature.[2])[1], the choice of the loss severity distribution is a more complex challenge. With respect to the trade-off between parsimony in the parameterization of the loss severity distribution of operational losses, and the accuracy of the resulting fit, the log-normal distribution was chosen. Distributions, like the Generalized Pareto or the Weibull distribution, whose fitting would need a lot of additional work, are not worthwhile.

The main problem, this work attended to, was the generating of scenarios. The model presented in the previous sections and chapters is mainly driven by scenarios. Scenarios were developed by a mixture of internal and external data, risk assessments and KRIs. Because of a lack of internal data in the tail of the loss severity distribution, external data was taken into account to complete and round off internal data. This mixing was made by scaling, with the gross income of the specific bank and the consortium of banks

[1]The loss frequency distribution does not influece the capital charge as much as the loss severity distribution. The impact is almost negliable.

(ORX) as scaling factor. At last results of risk assessments and KRIs fulfill the whole picture, which is the foundation for the parameter generation of the loss severity distribution.

After generating the parameters by optimization, the log normal distribution was fitted. The loss severity distribution is defined piece-wise, because it is significant that internal data fits the range beneath a certain threshold[2] (t_1) best. Internal data reflects the specific risk profile of a bank the most accurate way and in general enough data is available beneath this threshold. Therefore the empirical loss severity distribution is used from zero to t_1.(A collection threshold of EUR 1.000 etc. must eventually be taken under consideration.)

The aggregation of the loss severity and the loss frequency distribution was made by Monte-Carlo simulation, because there is no analytical expression of the compound loss distribution.

The final conclusion may be that an AMA is the most sophisticated approach to model operational risk. An AMA provides a lot of possibilities to fit the model perfectly to the specific risk profile of a bank. Theoretical and practical risk sensitivity or exposure will match in a better way than using BIA or TSA. Although the presented model complies with all required components of Basel II, further developments and research is still necessary, especially correlations and insurances.

[2]For instance $t_1 = 100.000$.

Annex 1

Business Lines

<table>
<tr><th>Level 1</th><th>Level 2</th><th>Activity Groups</th></tr>
<tr><td rowspan="4">Corporate Finance</td><td>Corporate Finance</td><td rowspan="4">Mergers and acquisitions, underwriting, privatisations, securitisation, research, debt (government, high yield), equity, syndications, IPO, secondary private placements</td></tr>
<tr><td>Municipal/Government Finance</td></tr>
<tr><td>Merchant Banking</td></tr>
<tr><td>Advisory Services</td></tr>
<tr><td rowspan="4">Trading & Sales</td><td>Sales</td><td rowspan="4">Fixed income, equity, foreign exchanges, commodities, credit, funding, own position securities, lending and repos, brokerage, debt, prime brokerage</td></tr>
<tr><td>Market Making</td></tr>
<tr><td>Proprietary Positions</td></tr>
<tr><td>Treasury</td></tr>
<tr><td rowspan="3">Retail Banking</td><td>Retail Banking</td><td>Retail lending and deposits, banking services, trust and estates</td></tr>
<tr><td>Private Banking</td><td>Private lending and deposits, banking services, trust and estates, investment advice</td></tr>
<tr><td>Card Services</td><td>Merchant/commercial/corporate cards private labels and retail</td></tr>
<tr><td>Commercial Banking</td><td>Commercial Banking</td><td>Project finance, real estate, export finance, trade finance, factoring, leasing, lending, guarantees,bills of exchange</td></tr>
<tr><td>Payment and Settlement</td><td>External Clients</td><td>Payments and collections, funds transfer, clearing and settlement</td></tr>
<tr><td rowspan="3">Agency Services</td><td>Custody</td><td>Escrow, depository receipts, securities lending(customers) corporate actions</td></tr>
<tr><td>Corporate Agency</td><td>Issuer and paying agents</td></tr>
<tr><td>Corporate Trust</td><td></td></tr>
<tr><td rowspan="2">Asset Management</td><td>Discretionary Fund Management</td><td>Pooled, segregated, retail, institutional, closed, open, private equity</td></tr>
<tr><td>Non-Discretionary Fund Management</td><td>Pooled, segregated, retail, institutional, closed, open</td></tr>
<tr><td>Retail Brokerage</td><td>Retail Brokerage</td><td>Execution and full services</td></tr>
</table>

Annex 2

Event Type Classification

Level 1	**Level 2**
Internal Fraud	Unauthorised Activity
	Theft and Fraud
External Fraud	Theft and Fraud
	Systems Security
Employment Practices and Workplase Safety	Employee Relations
	Safe Enviroment
	Diversity & Discrimination
Clients, Products & Business Practices	Suitability, Disclosure & Fiduciary
	Improper Business or Market Practices
	Product Flaws
	Selection, Sponsorship & Exposure
	Advisory Activities
Damage to Physical Assets	Disasters and other events
Business disruption and system failures	Systems
Execution, Delivery & Process Management	Transaction Capture, Execution & Maintenance
	Monitoring and Reporting
	Customer Intake and Documentation
	Customer/Client Account Management
	Trade Counterparties
	Vendors & Suppliers

Annex 3

Mapping of Event Types ORX - Basel II

Event Type - Level 1 ORX	**Event Type - Level 1 Basel II**
Internal Fraud	Internal Fraud
External Fraud	External Fraud
Malicious Damage	
Employment Practices and Workplace Safety	Employment Practices and Workplace Safety
Clients, Products and Business Practices	Clients, Products and Business Practices
Disasters and Public Safety	Damage to Physical Assets
Technology and Infrastructure Failures	Business Disruption and System Failures
Execution, Delivery and Process Management	Execution, Delivery and Process Management

Annex 4

Hierachical Clustering - Agglomerative Technique

The following technique is based on the cluster analysis introduced in [27].

Starting point: every unit is a single cluster. After $n-1$ subsequent agglomerations, a unique cluster, which contains all units, is achieved.

Measure for the distance (differences) between two units:

$$d(i,j) = 1 - pv(i,j),$$

where $pv(i,j)$ represents the p-value of the Kolmogoroff-Smirnoff test (see chapter 3) applied to the clusters i and j.

In each of the $n-1$ iterations the distances $d(i,j)$ between all clusters are calculated and those two clusters, which present the minimum value of $d(i,j)$ are merged together.

List of figures

Bibliography

[1] Takashi Arai. Key points of scenario analysis. Technical report, Operational Risk Scenario Workshop - Bank of Japan, 2006. http://www.boj.or.jp/en/type/release/zuiji_new/fsc0608a_add.htm.

[2] Falko Aue and Michael Kalkbrener. Lda at work, November 2006.

[3] Nicolas Baud, Antoine Frachot, and Thierry Roncalli. Internal data, external data and consortium data for operational risk measurement: How to pool data properly?, 2002. http://www.gro.creditlyonnais.fr/content/rd/home_ro.htm.

[4] Nicolas Baud, Antoine Frachot, and Thierry Roncalli. An internal model for operational risk computation, 2002. http://www.gro.creditlyonnais.fr/content/rd/home_ro.htm.

[5] BancWare ERisk. http://www.erisk.com/Learning/CaseStudies.asp.

[6] Operational Riskdata eXchange ORX. http://www.orx.org.

[7] Dorian Feldman and Martin Fox. *Probabilty-The Mathematics of Uncertainty.* Marcel Dekker Inc, New York, 1999.

[8] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients, 1964. The Computer Journal 1964 7.

[9] NB (2005) Leitfaden: Management des operationellen Risikos FMA. http://fma.cms.apa.at/cms/basel2//attachments/1/0/0/CH0337/CMS1143027369577/lfopris1.pdf.

[10] The R Project for Statistical Computing. http://www.r-project.org/.

[11] Antoine Frachot, P. Georges, and Thierry Roncalli. Loss distribution approach for operational risk, April 2001. http://www.gro.creditlyonnais.fr/content/rd/homero.htm.

[12] Antoine Frachot and Thierry Roncalli. Mixing internal and external data for managing operational risk, 2002. http://www.gro.creditlyonnais.fr/content/rd/home_ro.htm.

[13] Antoine Frachot, Thierry Roncalli, and Eric Salomon. The correlation problem in operational risk, 2004. http://www.gro.creditlyonnais.fr/content/rd/home_ro.htm.

[14] Risk Glossary. riskglossary.com/link/baringdebacle.htm.

[15] B.W. Gnedenko. *Lehrbuch der Wahrscheinlichkeitstheorie.* Verlag Harry Deutsch, 1997.

[16] Dr. Fotios C. Harmantzis. Operational risk management in financial services and the new basel accord.

[17] Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot. *Loss Models. From Data to Decision.* John Wiley and Sons, 2004.

[18] Patrick Mc Connell and Martin Davies. Safety first - scenario analysis under basel ii, 2006.

[19] John C. Nash. *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation.* 1990.

[20] Directive 2006/48/EC of the European Parliament and of the Council. Official journal of the european union, 2006.

[21] Serious Fraud Office. http://www.sfo.gov.uk/news/prout/pr255.asp?id=255.

[22] Basel Committee on Banking Supervision. http://www.bis.org/bcbs/index.htm.

[23] Basel Committee on Banking Supervision. Consultative document - operational risk, January 2001.

[24] Basel Committee on Banking Supervision. Basel ii: The new basel capital accord - third consultative paper, April 2003.

[25] Basel Committee on Banking Supervision. Basel ii: International convergence of capital measurement and capital standards: A revised framework - comprehensive version, June 2006.

[26] Martin Ottmann and Andreas Seibt. Ama insurance - the necessity for something new.

[27] Fabio Piacenza, Daniele Ruspantini, and Aldo Soprano. Operational risk class homogeneity, 2006.

[28] RMG - Risk Management Group. *Scenario - based AMA*, New York, May 2003. www.ny.frb.org/newsevents/events/banking/2003/con0529d.pdf.

[29] Markus Rossmiller. Operational risk in banks, 2003.

[30] Joseph A. Sabatini. Implementing an ama for operational risk, 2005.

[31] Joachim Hartung und Bärbel Elpet und Karl-Heinz Klösener. *Statistik. Lehr- und Handbuch der angewandten Statistik.* Oldenbourg, 2002.