Loss Distribution Approach in Practice*

Antoine Frachot, Olivier Moudoulaud, Thierry Roncalli
Groupe de Recherche Opérationnelle, Crédit Lyonnais, France†

This version: May 7, 2003

1 Introduction

An intense stream of research has been conducted over the past few years to address issues raised by the practical implementation of Basel II Advanced Measurement Approaches (AMA) and in particular the Loss Distribution Approach (LDA). Indeed, we believe that most of these issues are now sufficiently clarified to allow for a survey on operational risk quantitative techniques. This is the aim of this chapter.

The roots of quantitative LDA come from actuarial techniques as they have been developped by the insurance industry for years. It is of course the most natural idea apart from the fact that actuarial techniques could not be imported directly without any care because of the specificities of operational risks, most notably the reporting bias and the paucity of data. All quants who have looked closely to empirical data will agree on the idea that these two features of OpRisk data have a dramatic impact on capital charge and thus can definitely not be neglected even though it imposes to deal with more sophisticated computations than we may have expected initially.

This chapter aims at describing step by step how a full Loss Distribution Approach can be implemented in practice and how both quantitative and qualitative points of view can be reconciled. Our rule of conduct is to be as pragmatic as possible and not more sophisticated than necessary. In particular, we explicitly drop some maybe interesting issues when they should require too much effort in return for too few benefits in terms of capital charge accuracy. In this respect, we benefit from our experience at Credit Lyonnais and from all other related experiences and discussions we have been involved in over the last couple of years. In some sense, we mimick the process which gave birth to the so-called Internal Ratings Based (IRB) formulas proposed by the Basel Committee for credit risk: quants first started from a highly sophisticated credit risk model and downgraded it until it turned out to be an acceptable, implementable and pragmatic proxy of the "correct" capital charge. As an example it is worth noting that, at some time in the downgrading process, it appeared that simplicity demands assuming that credit risk is driven by only one source of risk which is furthermore assumed to be normally-distributed. All credit risk specialists will agree on the point that both one-factor and normal distribution assumptions are very simplistic and unrealistic assumptions but that the resulting capital charge is accurate enough to pretend representing an acceptable measure of risk. This is the spirit we try to adopt here.

The second contribution of this chapter is to give some numerical calculations of the accuracy of capital charge estimates. As by definition the severity and frequency estimations are processed with few available data, it is crucial to have a clear view of the inaccuracy of capital charge estimates. Furthermore, an estimate

*We warmly thank Maxime Pennequin, global head of OpRisk management at Crédit Agricole / Crédit Lyonnais. We would like to give a special thank to John S. Walter (Bank of America) for his valuable comments and many enlightning discussions. We are also grateful to Patrick de Fontnouvelle (Federal Reserve of Boston) and all ITWG members, in particular Andrea Colombo (Banca Intesa), Riccardo Cateni (Banca Intesa), Yimin Shih (Citigroup), Joseph Sabatini (JP Morgan), Georges Graziani (Bank of Montreal) and Tony Peccia (Bank of Montreal). Finally, we would like to thank David Kurtz (GRO, Crédit Lyonnais) for discussions on durations in Poisson processes.

†*address:* Crédit Lyonnais – Groupe de Recherche Opérationnelle, Le Centorial, 8, rue du Quatre Septembre – 75002 Paris – France; *E-mail:* antoine.frachot@creditlyonnais.fr or thierry.roncalli@creditlyonnais.fr

of the inaccuracy is the basic tool for addressing the issue of the number of losses (both external and internal) which are necessary for a reliable estimate of the capital charge.

This chapter follows the different steps necessary for implementing a LDA in practice:

- Step 1: Severity Estimation
- Step 2: Frequency Estimation
- Step 3: Capital Charge Computations
- Step 4: Confidence Interval
- Step 5: Self Assesment and Scenario Analysis

For each of these steps, we try to give illustrative examples and we gather all demanding mathematics into subsections named Technical Appendix. We hope it will allow for a more reader-friendly chapter.

2 Severity Estimation

This is probably the most difficult task as text-book techniques can not be used directly because available loss data are plagued by various sources of bias. This is the unfortunate case where our requirements – simplicity and accuracy – contradict each other: treating our data as if they were text-book, unbiased data, for the sake of simplicity is unacceptable as it may lead to highly inaccurate and entirely flawed capital charge (see Baud, Frachot and Roncalli, 2003 and Fontnouvelle et al., 2003). Therefore we really have to accept some complexity. It is however possible to make reasonable simplying assumptions which deteriorate the accuracy of capital charge to an acceptable extent.

2.1 Scaling Issues and Reporting Biases

Let us consider that the severity distribution has to be estimated from (say) m sets of loss data coming from m different "providers" (with "provider" meaning either a business unit within a bank or an external entity/consortium). Severity calibration can not be undertaken before deciding in which of the two following cases we fall:

- **Case 1: The m sources of loss data are assumed to be drawn from the same primary probability distribution but loss data are reported according to some (possible) different thresholds.** In other words, when we pool the m sources of data together, we are not mixing data which would be different in nature. We are just mixing similar data but these data is "packaged" differently.
- **Case 2: The m sources of loss data come from different primary probability distributions and thus have to be re-scaled. In addition, they may also be reported according to some different thresholds.** In this case, it really means that we try to mix data which are fundamentally different by nature.

We argue that Case 2 is obviously the most general case and certainly the most realistic case. Who would contradict the fact that (say) "external fraud" losses are structurally different from one country to another, from one large bank compared to a small one etc ? **However, at the state of knowledge of our OpRisk community, Case 2 is much more too complex to be addressed properly and the benefits we would (hardly) secure would probably not be worth the effort.** There are some reasons why we advocate this position:

- first, even though some tentative works have been done to find out a way to rescale severity distributions (see Shih, Samad-Khan and Medapa, 2000), such a task requires large sets of data and sets of data coming from different sources (i.e. external and internal) and it is unreasonable to take for granted that this is always feasible for all risk types. No doubts that the OpRisk community will investigate this point in the coming years when datasets will get larger and larger. As of today, it is unrealistic to consider that a reliable scaling function can be estimated.

- secondly, scaling formulas must be derived for each bank and nothing ensures that scaling formula can be imported from one bank to another bank. The scaling formula is a mechanical transformation which says how the "internal" severity distribution is to be compared with the "external" one. As a result, if we accept the logic of Case 2, we should acknowledge that each bank may have to estimate its own scaling formula.

- Third, there may be more differences between two business units within a bank (for example if they operate in rather different countries) than between one internal business line and an external one. Then, again if we accept the logic of Case 2, we should also think of deriving a scaling formula for data coming from two different business lines within the bank. Ths is certainly hard work which stays beyond the state of ressources dedicated to OpRisk in our institutions.

- Last but not least, some remarkable work done by Fontnouvelle et al., 2003 shows that, empirically speaking, scaling issues are not so dramatic as the authors evidence great similarities between different sources of loss data (OpVantage, OpRisk Analytics) **once reporting biases have been properly corrected.**

We then suggest to abandon Case 2 and to consider that reporting bias is the most important issue, in the same spirit as in Fontnouvelle et al., 2003. Therefore we leave Case 2 and scaling issue for a Basel III round. As a result, the task of calibrating severity distributions is made simpler as only reporting biases have to be considered.

2.2 How to Adjust for Reporting Bias?

In theory, the data collection threshold affects severity estimation in the sense that the sample severity distribution (i.e. the severity distribution of reported losses) is different from the "true" one (i.e. the severity distribution one would obtain if all losses were reported). Unfortunately, the true distribution is the most relevant for calculating capital charge and also for being able to pool different sources of data in a proper way. As a consequence, linking the sample distribution to the true one is a necessary task.

It can be done by using conditional distributions and by solving Maximum Likelihood (ML) program accordingly. This point is now widely acknowledged (Baud, Frachot and Roncalli, 2002 and Fontnouvelle et al., 2003) even though solving such a conditional ML program is not so easy. Furthermore Baud, Frachot and Roncalli, 2003 have proven that neglecting reporting bias implies very poor estimates of the severity distribution. In particular, in absence of an appropriate adjustment, pooling data coming from different sources (with different data collection thresholds) results in strongly biased severity distribution which eventually appeared to be much riskier than it actually is.

For practical calculations, it is probably not necessary to go as far as Baud, Frachot and Roncalli, 2002 have suggested. In their paper, the authors consider that, for external databases, there are as many thresholds as contributors and therefore they propose to treat thresholds as stochastic. This is certainly true and realistic but is complicated to tackle as the corresponding likelihood is quite sophisticated. We then suggest to consider that each dataset has one threshold and that the stated threshold is the actual one unless there exist strong counter-arguments. We thus refer to Baud, Frachot and Roncalli, 2002 for a general treatment of threshold corrections.

Mathematically, calibration is done by maximizing the loglikelihood function. Suppose for sake of simplicity that the severity distribution is lognormal with parameters μ and σ then one has to solve:

$$\max_{(\mu,\sigma)} \ell_n(\mu,\sigma) = \sum_{i=1}^{n} \ell(\zeta_i, \mu, \sigma \mid H_i) \tag{1}$$

where n is the number of losses, $\ell(\zeta_i, \mu, \sigma \mid H_i)$ is the loglikelihood of the i^{th} loss (reported subject to the threshold H_i).

Technical Appendix 1 *The expression of the loglikelihood function is now common-knowledge. Let us consider a dataset whose threshold is H, the sample distribution is equal to:*

$$f_{\text{sample}}(x; \mu, \sigma \mid H) := \mathbf{1}\{x \geq H\} \cdot \frac{f(x; \mu, \sigma)}{\int_H^{+\infty} f(y; \mu, \sigma)\, \mathrm{d}y} = \mathbf{1}\{x \geq H\} \cdot \frac{f(x; \mu, \sigma)}{1 - \mathbf{F}(H; \mu, \sigma)}$$

where $f(x; \mu, \sigma)$ is the true probability density function (which is assumed to be lognormal distribution $LN(\mu, \sigma)$) and F is the corresponding cumulative distribution function. As a result, the loglikelihood function is:

$$\ell_n(\mu, \sigma) = \sum_{i=1}^{n} \ln f(\zeta_i; \mu, \sigma) - n \times \ln(1 - \mathbf{F}(H; \mu, \sigma))$$

where ζ_i is the i^{th} loss and n is the number of losses. This is the second term which corrects for reporting bias. We see in particular that it vanishes away when the threshold is equal to zero. As a consequence, if thresholds are low then this last term is negligible but, on the contrary has a huge impact when they are significant or if they differ significantly from one dataset to another. This bias correction is essential for ensuring that different sources of data are pooled together properly.

2.3 How to Treat Aggregate Loss Data?

In some instances, loss data are not reported on a single-loss basis but are instead aggregated together. In this case, only the aggregate value is reported into the internal database. Assuming that we also know the underlying number of events corresponding to this aggregate loss, we may wonder whether this aggregate loss carries valuable information for the purpose of severity estimation. Indeed it does but how to extract this information is not straightforward (see for example Frachot, Georges and Roncalli, 2001 for a discussion of the Generalized Method of Moment (GMM)). GMM is not too complicated to implement though it is more difficult to tackle both single-losses and aggregate losses simultaneously as Maximum Likelihood and GMM techniques do not marry easily.

This issue is also related to the data collection threshold: data losses are assumed to be reported on a single basis provided they are higher than the threshold. As a consequence, as soon as a precise threshold will be set, aggregate losses will become less and less a cause of concern. Therefore, we do not address this question here.

3 Frequency Estimation

If we still try to stick to simplicity, it is a good idea to assume that the frequency distribution is a Poisson distribution. This distribution has many appealing features: first it is widely used in the insurance industry for modeling problems similar to operational risks; secondly it needs only one parameter (called λ) to be entirely described and, third, the ML value of this parameter is simply the empirical average number of events per year. However some care is necessary when reporting biases operate. For obvious reasons, if one bank's reporting cut-off is set at a high level, then the average number of (reported) events will be low. It does not imply, in any sense, that the bank is allowed to put a lower amount of capital than another bank which uses a lower threshold and is otherwise identical to the first one. It simply means that the average number of events must be corrected for reporting bias as well.

It appears that the calibration of the frequency distribution comes as a second step (after having calibrated the severity distribution) because the forementionned correction needs an estimate of μ and σ for its calculation. This is rather straightforward: the difference (more precisely the ratio) between the number of reported events and the "true" number of events (which would be obtained if all losses were reported, i.e. with a zero-threshold) corresponds exactly to the probability of one loss being higher than the threshold. This probability is a direct by-product of the severity distribution. The following technical appendix gives the mathematical expression of this probability, which in turn provides a straightforward way to make the appropriate correction.

Technical Appendix 2 *The expression of the "true" frequency parameter is given by:*

$$\lambda = \frac{\lambda_{\text{sample}}}{\Pr\{\text{loss} > H\}}$$

which is mathematically equal to:

$$\lambda = \frac{\lambda_{\text{sample}}}{1 - \mathbf{F}(H; \mu, \sigma)}$$

In practical terms, one has to compute the average number of reported events by year (which is an estimate of λ_{sample}) and to use the previous estimates of μ and σ to uncover the true frequency distribution.

What about external data and scaling issues? Exactly as in the previous section, we may have information on frequency of events experienced by competitors or by the whole banking industry. The extent by which they can be seen as a valuable information remains unclear as long as the scaling function is unknown. Here the scaling function gives the link between the number of events experienced by one bank and its business size (or any variable which may be considered as relevant for evaluating the expected number of events). For example, past litterature has suggested that a square-root pattern may be appropriate for modelling the scaling function (i.e. the number of events of one bank is linked to the square-root of its business size). Other works have proposed to use credibility theory as a way to adjust internally-estimated frequencies (Frachot and Roncalli, 2002). However we should honestly acknowledge that these methods are hardly implementable because, again, few risk managers have enough data to test and calibrate such functionnal links. Before a Basel III round takes place, it may be preferable to take internal frequencies of events for granted, provided that they have been validated by bank's experts (and otherwise corrected by some expert-based adjustments when necessary). Quantitative adjustments by use of external frequencies require data which are out of reach as of today and therefore would be more confusing than relevant for our purpose.

4 Capital Charge Computation

Once the frequency and severity distributions have been calibrated, the computation of capital charge is quite simple, provided we agree on its precise definition. Capital charge processing is done thanks to Monte Carlo simulations which are standard skills among quants. We shall not spend much time here to detail how to set up a Monte Carlo scheme since many papers have done this before. We prefer to give our understanding of some widely-discussed issues related to capital charge calculations.

4.1 What is the Definition of the Capital Charge?

There remain ambiguities around the definition of the regulatory capital charge. We are aware of at least three distinct definitions:

- **Definition 1** (OpVaR): The capital charge is the 99.9% percentile of the total loss distribution.
- **Definition 2** (OpVaR_Unexpected_Loss_Only): This is the previous OpVaR from which expected losses are substracted. The current Basel proposal seems to accept this definition as long as the bank can demonstrate that it has adequately provided for expected losses through pricing, reserves, and/or expensing practices.

- **Definition 3** (OpVaR_Above_Threshold): The capital charge is the 99.9% percentile of the total loss distribution where only above-the-threshold losses are considered.

The three definitions can be implemented through Monte Carlo simulations with roughly the same level of complexity but they obviously give different figures.

Technical Appendix 3 *If N is the (random) number of events, then the aggregate loss is $L = \sum_{i=0}^{N} \zeta_i$. The three definitions can then be expressed in mathematical terms as:*

- *Definition 1*
$$\Pr\{L > \text{OpVaR}\} = 0.1\%$$
- *Definition 2*
$$\Pr\{L > \text{OpVaR} + \text{EL}\} = 0.1\%$$
where EL *is the expected total loss* $\mathbb{E}\left[\sum_{i=0}^{N} \zeta_i\right]$.
- *Definition 3*
$$\Pr\left\{\sum_{i=0}^{N} \zeta_i \times \mathbf{1}\{\zeta_i \geq H\} > \text{OpVaR}\right\} = 0.1\%$$
where $1\{\zeta_i \geq H\}$ *equals 1 if the loss exceeds the threshold H and 0 otherwise.*

4.2 How does Reporting Bias Affect Capital Charge Estimate?

The way capital charge is influenced by the level of the data collection threshold depends on the definition.

- Definitions 1 and 2: Since these definitions never mention any reference to the data collection process, the two capital charges are independent of any threshold which may be used in practice. In both cases, capital charge intends to represent the amount of capital to put aside operational risk within a bank. As such, it has no reasons to depend on the way loss data are reported. In particular, the fact that bank's risk management policy has made up its mind for some specific reporting threshold says that only loss data higher than the threshold will be captured in risk management system but does not say anything on intrinsic riskiness of the bank. As a consequence, setting a reporting threshold should not affect the capital charge. If previous steps have been carefully followed, that is if appropriate correction has been taken to neutralize reporting bias, then the *numerically-calculated* capital charge is also invariant with respect to the threshold.
- Definition 3: As the threshold value enters explicitly in the definition, then the capital charge does depend on the threshold. Since in Definition 3 all losses below the threshold are excluded, the total aggregate loss is thus below the one which enters Definition 1. As a result, the capital charge in Definition 3 is getting lower as the threshold is set at a higher level.

More subtle is the fact that the threshold may affect the accuracy of the capital charge since frequency and severity parameters' accuracy do depend on the threshold. Intuitively accuracy is likely deteriorated when the threshold is set at a high level because the calibration of the severity distribution relies on too few data. In other words, the extrapolated part of the severity distribution becomes too important. Therefore, as far as Definitions 1 and 2 are concerned, the trade-off which results in an optimal threshold has more to do with the balance between collecting costs and the accuracy of the capital charge than with the level of the capital charge itself.

4.3 How to Aggregate Capital Charges for Different Loss Types and Business Lines?

This issue remains to be addressed. From a theoretical point of view, one can admit that aggregate losses by risk-type are not perfectly correlated and thus summing up all capital charges together is highly conservative. The first point to be discussed is to clarify which correlation we are talking about. As the capital charge results from two sources of randomness – frequency and severity – there are also two possible sources of correlation.

As an example, we may find that aggregate losses for (say) external fraud and internal fraud are correlated because either frequency of events or severity of events are correlated. In the former case, we should observe that, historically, the number of external fraud events is high (respectively low) when the number of internal fraud events is also high (respectively low). This is a sensible way to consider correlation between aggregate losses of two different event types. On the contrary, we feel much less confortable with the other way, i.e. severity correlation. In effect, a basic feature of actuarial models requires to assume that individual losses are jointly independent within one specific risk type. **Therefore it is conceptually difficult to assume simultaneously severity-independence within each class of risk and severity-correlation between two classes.**

Subsequently, **we would rather assume that correlation between aggregate losses by event-type is fundamentally conveyed by the underlying correlation between frequencies**. By analogy with credit risk models, we expect that, even with strong frequency-correlation, aggregate losses may show low level of correlation. Furthermore one may also guess that it is particularly true for high severity events since severity-independence likely dominates frequency-correlation. This point is confirmed by our calculations in the following technical appendix: **even if two risk types occur with highly-correlated frequencies, aggregate losses show low levels of correlations.**

Since strong frequency-correlation might not translate in strong correlation of aggregate losses, we conclude that diversification effects could be worth being taken into account as they may significantly reduce the total capital charge (compared with the full-correlation feature which is assumed when adding together capital charges of all event-types and/or business units).

Technical Appendix 4 *Let us consider two aggregate losses* $L_1 = \sum_{i=0}^{N_1} \zeta_i^1$ *and* $L_2 = \sum_{i=0}^{N_2} \zeta_i^2$*. In order to obtain tractable formula, we assume that the two frequency distributions have the same parameters —* $\lambda_1 = \lambda_2 = \lambda$*. If* N_1 *and* N_2 *are perfectly correlated, it comes that* $N_1 = N_2 = N$*. We have*

$$\begin{aligned}
\operatorname{cov}(L_1, L_2) &= \mathbb{E}[L_1 L_2] - \mathbb{E}[L_1]\,\mathbb{E}[L_2] \\
&= \mathbb{E}\left[\sum_{i=0}^{N} \zeta_i^1 \sum_{i=0}^{N} \zeta_i^2\right] - \mathbb{E}[L_1]\,\mathbb{E}[L_2] \\
&= \mathbb{E}\left[N^2 \mathbb{E}\left[\zeta^1\right]\mathbb{E}\left[\zeta^2\right]\right] - \lambda^2 \mathbb{E}\left[\zeta^1\right]\mathbb{E}\left[\zeta^2\right] \\
&= \left(\operatorname{var}(N) + \mathbb{E}^2[N] - \lambda^2\right)\mathbb{E}\left[\zeta^1\right]\mathbb{E}\left[\zeta^2\right] \\
&= \lambda \times \mathbb{E}\left[\zeta^1\right] \times \mathbb{E}\left[\zeta^2\right]
\end{aligned}$$

We deduce that an upper bound of the correlation between the two aggregate losses is

$$\operatorname{cor}^+(L_1, L_2) = \frac{\mathbb{E}\left[\zeta^1\right] \times \mathbb{E}\left[\zeta^2\right]}{\sqrt{\left(\operatorname{var}(\zeta^1) + \mathbb{E}^2[\zeta^1]\right) \times \left(\operatorname{var}(\zeta^2) + \mathbb{E}^2[\zeta^2]\right)}}$$

We remark that the correlation between the two aggregate losses does not depend on the Poisson parameter λ*. When the severity distributions are lognormal, we obtain after some computations the following results:*

$$\operatorname{cor}^+(L_1, L_2) = \exp\left(-\frac{1}{2}\sigma_1^2 - \frac{1}{2}\sigma_2^2\right)$$

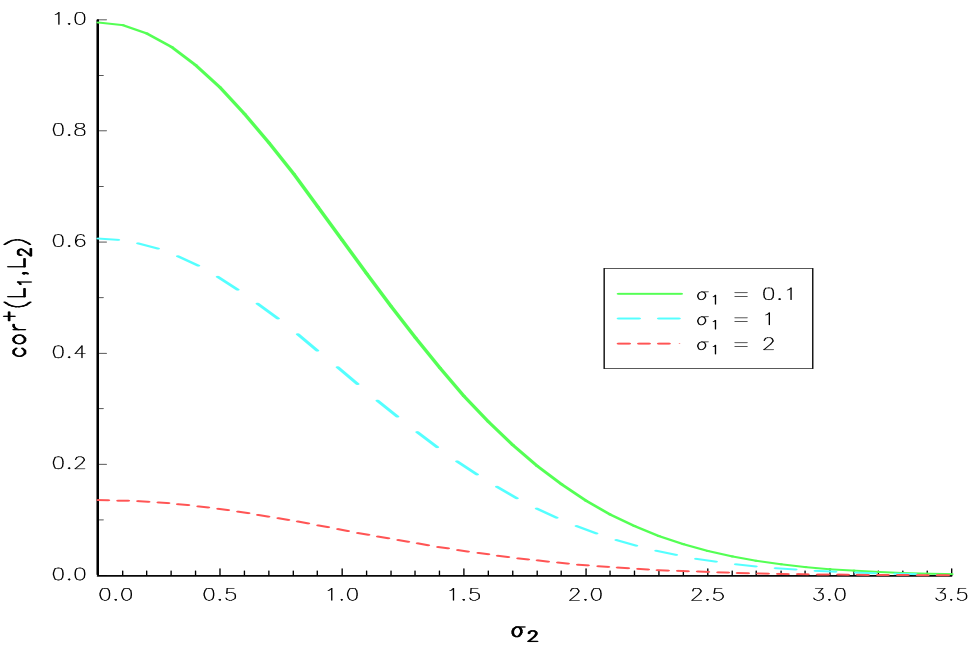

Figure 1: Upper bound of the correlation between two aggregate losses

The correlation is then a very simple formula which depends only on the values σ_1 and σ_2. Moreover, we remark that the function is decreasing with respect to these parameters. We report in Figure 1 the relationship between σ_1, σ_2 and $\mathrm{cor}^+(L_1, L_2)$. For high severity loss types, $\mathrm{cor}^+(L_1, L_2)$ is very small. For example, when $\sigma_1 = \sigma_2 = 2$ and $\sigma_1 = \sigma_2 = 2.2$, $\mathrm{cor}^+(L_1, L_2)$ takes respectively the values 1.8% and 0.8%.

Remark 1 *We point out that even for low severity loss types, the correlation between the aggregate losses is very small: it can not be bigger than 10%, which is certainly very conservative — it corresponds to the case where $\sigma_1 = \sigma_2 = 1.50$.*

As a conclusion, according to the basic principle of actuarial / LDA models, correlations between aggregate losses are necessarily low. Our numerical computations suggest to set the correlation well below 10%. Finally, aggregation of capital charges can be performed by using the Normal approximation presented in Frachot, Georges and Roncalli, 2001.

5 Confidence Interval

The previous procedure provides an estimation of the capital charge which is uncertain by nature. One can guess that our regulators will expect from us to demonstrate that the estimate is not too far from its fair value. This is a crucial point in operational risk modelling because of the paucity of data which normally should translate into poor accuracy. As it will become clear in the sequel, confidence interval is the basic tool for justifying the computed capital charge as well as for addressing numerous issues.

The inaccuracy of the capital charge is directly linked to the inaccuracy of the estimators of the three underlying parameters λ, μ and σ. Therefore building a confidence interval of the capital charge can proceed as follows:

- first, derive the (in some cases, approximate) distribution of the underlying estimators;
- draw from these distributions a sufficiently large number of simulations;

- finally, for each path, compute the capital charge and then obtain its empirical distribution.

5.1 Parameters' Accuracy

The most critical point is of course the methodology for building an approximate distribution of the estimators of the underlying parameters. For the frequency parameter λ, this task is straightforward because we know the exact distribution of the estimator which is also Poisson-distributed. For the two remaining parameters μ and σ of the severity distribution, we can follow two different methodologies:

- Bootstrap methods
- Gaussian Approximation (for example because ML theory applies)

We are in favour of the second method because most of people (in particular regulators) are much more familiar with Gaussian distribution than with Bootstrap methods. Furthermore, commercial package can implement this methodology at almost no cost while Bootstrap methods may require some further developments. The last question which remains to be answered is to give the precise Gaussian distribution satisfied by the estimators of the severity parameters. This is done in the next Technical Appendix. It is worth saying that it can be derived as a by-product of step 1 (severity calibration) and thus does not require significantly more efforts.

Finally, we see that the accuracy of the frequency estimator improves when the number of recorded years T grows, and with an order of magnitude of $\sqrt{T}$. As an example, using a 5-year historical length instead of a 3-year length improves accuracy (as measured by the standard deviation) to an extent of 30% (i.e. $\sqrt{5/3} = 1.29$). Similarly, the accuracy of the severity estimators follows the same pattern and behaves as $\sqrt{n}$ where n is the number of (both internal and external) losses.

Technical Appendix 5 *If the number of events is asumed to be Poisson, then the ML estimator of the annual number of events $\hat{\lambda}_T$ is the average number of events per year for the last T years (if T is the number of recorded years). The estimator is Poisson-distributed in the following sense*[1]

$$T \times \hat{\lambda}_T \sim \mathcal{P}(T \times \lambda)$$

From ML theory, it comes that the estimators $\hat{\mu}$ and $\hat{\sigma}$ are approximately Gaussian with:

$$\begin{pmatrix} \hat{\mu}_n \\ \hat{\sigma}_n \end{pmatrix} \approx \mathcal{N}\left(\begin{pmatrix} \mu \\ \sigma \end{pmatrix}, \Omega_n\right)$$

where Ω_n is the inverse of the so-called Fisher information matrix:

$$\Omega_n = -\begin{pmatrix} \partial^2 \ell_n/\partial\mu^2 & \partial^2 \ell_n/\partial\mu\partial\sigma \\ \partial^2 \ell_n/\partial\mu\partial\sigma & \partial^2 \ell_n/\partial\sigma^2 \end{pmatrix}^{-1}$$

ℓ_n is the loglikelihood function introduced in step 1. Therefore we just have to compute its second derivative with respect to the parameters. Since step 1 requires to maximize the loglikelihood, this second derivative is also computed in the course of the optimization (if a standard Newton-Raphson algorithm is used). In this sense, it appears to be a by-product of step 1.

Remark 2 *We assume that the estimator $\hat{\lambda}$ is independent with respect to the estimators of the parameters of the severity distribution. This assumption seems natural. However, we can not assume that $\hat{\mu}_n$ and $\hat{\sigma}_n$ are independent. From ML theory, $\hat{\mu}_n$ and $\hat{\sigma}_n$ are asymptotically independent only when the threshold H is zero (see Figure 2).*

[1] We verify that

$$\operatorname{plim} \frac{\mathcal{P}(T \times \lambda)}{T} = \lambda$$

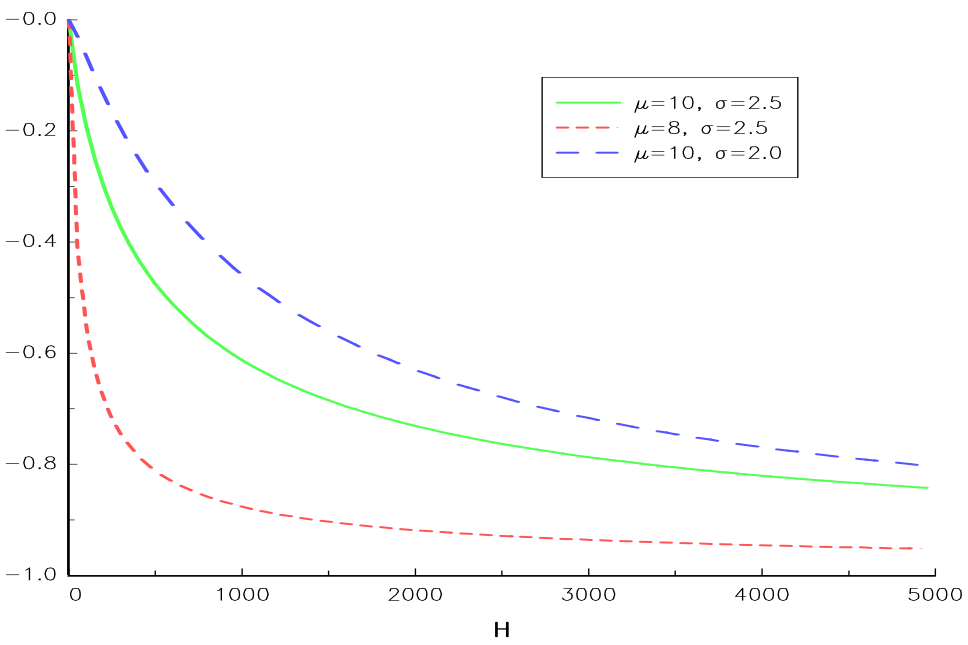

Figure 2: Asymptotic correlation between $\hat{\mu}_n$ and $\hat{\sigma}_n$

5.2 Capital Charge Accuracy

As capital charge is directly linked to the frequency and severity parameters, its accuracy is easily derived from above results. Generally speaking, the probability distribution of any function of the underlying parameters can be computed provided we know the distribution of the underlying parameter estimators and we are able to calculate its first derivatives with respect to the parameters.

In practice, we simulate paths of frequency and severity parameters according to the approximate distribution derived in the previous subsection, then, for each path, we compute the capital charge and we eventually obtain its probability distribution. Following is a numerical example.

Suppose that $\lambda = 100$, $\mu = 9$ and $\sigma = 2$. The number of recorded years is $T = 5$ years and the number of losses is $n = 1000$. Since the average number of events per year is $\lambda = 100$ and the number of recorded years is 5, it is rather unlikely that all the 1000 losses may come from internal databases. As a result, this example assumes that internal databases are supplemented with external losses. Finally, we also assume that the threshold is equal to 5000 euros. In Figure 3, we report the following ratios:

$$\begin{aligned} R_{\text{Freq}} &= \frac{\text{OpVaR}\left(\hat{\lambda}_T, \mu, \sigma\right)}{\text{OpVaR}\left(\lambda, \mu, \sigma\right)} \\ R_{\text{Sev}} &= \frac{\text{OpVaR}\left(\lambda, \hat{\mu}_n, \hat{\sigma}_n\right)}{\text{OpVaR}\left(\lambda, \mu, \sigma\right)} \\ R_{\text{Freq+Sev}} &= \frac{\text{OpVaR}\left(\hat{\lambda}_T, \hat{\mu}_n, \hat{\sigma}_n\right)}{\text{OpVaR}\left(\lambda, \mu, \sigma\right)} \end{aligned}$$

R is then the ratio between the capital charge estimator (for various cases) and the true capital charge. R_{Freq} (and respectively R_{Sev}) corresponds to the case where the frequency parameter (resp. the severity parameters) is the only parameter assumed to be random. It permits to assess which part of the total

inaccuracy is attributable to each parameter. The real-life case where both the frequency and severity parameters are random is captured by $R_{\text{Freq+Sev}}$. In order to give an idea of the accuracy of the capital

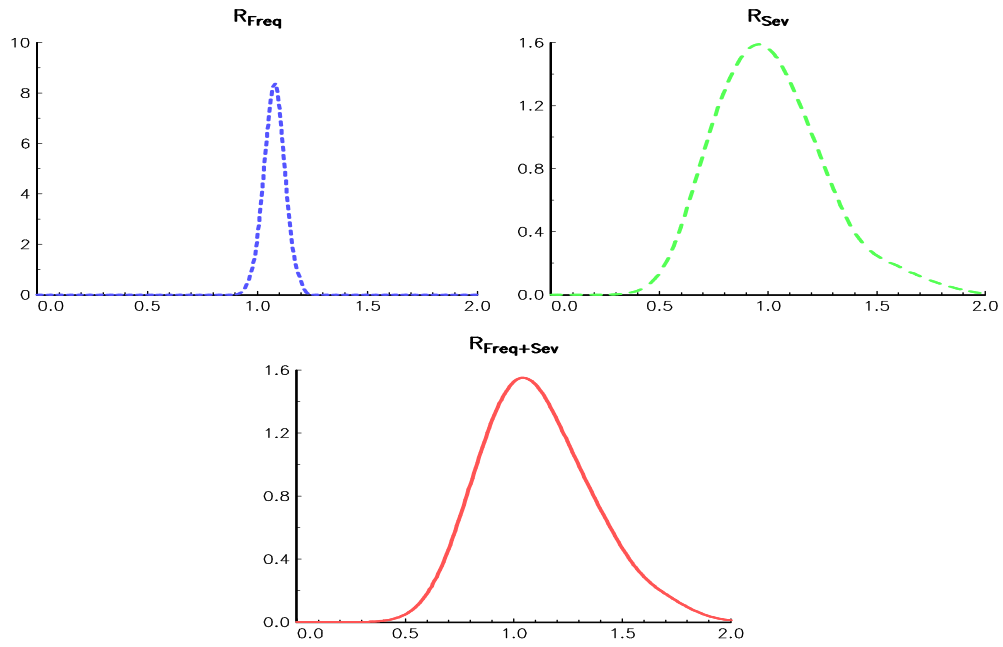

Figure 3: Probability density function of $\widehat{\text{OpVaR}}$

charge estimate, we may estimate the value c defined as follows

$$\Pr\left\{\widehat{\text{OpVaR}} \geq (1-c) \times \text{OpVaR}\right\} = \alpha$$

where $\widehat{\text{OpVaR}}$ is the capital charge estimate and OpVaR its true value[2]. **This criterion is well-suited for regulatory purposes since regulators will probably focus on the risk of under-estimating the capital charge.**

Using the previous numerical values for λ, μ and σ, and defining $\widehat{\text{OpVaR}}$ as $\text{OpVaR}\left(\lambda, \hat{\mu}_n, \hat{\sigma}_n\right)$ – we treat here the frequency parameter as given and not random – we obtain the following results for c:

	α		
n	75%	90%	95%
100	42%	60%	68%
1000	16%	26%	31%
10000	5%	9%	12%

As an example, for $n = 10000$ losses, the capital charge may be underestimated by less than 15% (if we consider a 95% level of confidence). We remark that the error increases with the confidence level α and decreases with the number of losses of the database. Now, if we consider that both the frequency parameter and the parameters of the severity loss distribution are random – $\widehat{\text{OpVaR}}$ is now defined as

[2]Here OpVaR is defined according to definition 1 of the previous section.

$\text{OpVaR}\left(\hat{\lambda}_T, \hat{\mu}_n, \hat{\sigma}_n\right)$ – the results become:

	α		
n	75%	90%	95%
100	26%	50%	58%
1000	5%	18%	24%
10000	-1%	5%	7%

Finally, if we suppose the severity as given and non random ($\mu = 9$ and $\sigma = 2$), $\widehat{\text{OpVaR}} = \text{OpVaR}\left(\hat{\lambda}_T, \mu, \sigma\right)$ is a function of the number of years T, and we may verify that the error c decreases with T.

As a conclusion, we must stress the fact that capital accuracy depends on the value of the frequency and severity parameters. In particular, capital accuracy is probably different whether it is computed for low severity/high frequency or high severity/low frequency events. In this sense, there is no one-size-fits-all rule for deriving capital charge accuracy. Therefore we suggest to compute systematically capital accuracy. **Finally it is worth saying that our results confirm that external data may be necessary in some cases.**

6 Other Issues

Previous sections have shown how to build a sound and pragmatic LDA which addresses most important issues. We now turn to some remaining issues/questions.

6.1 Goodness of Fit Tests

It is probably interesting to search for the distribution which fits best loss severity or to wonder if the Poisson distribution is well-suited to modelling frequency. Again we back the idea that these questions should be left for future research as they imply endless discussions. If the entire OpRisk community commits itself to use one definite set of distributions (say Poisson × Lognormal), it would greatly simplify comparisons between banks' capital charges, and these benefits would largely encompass the (probably small) loss of accuracy due to the use of one-size-fits-all distributions.

6.2 Data Sufficiency

This question can now receive a rigorous answer thanks to confidence intervals. Since capital charge accuracy depends directly on the number of observed losses, we just have to check whether the capital charge is calculated with an acceptable accuracy where "acceptable" means that the confidence interval is not too wide.

Let us consider the problem of the previous section. We have computed the value c satisfying $\Pr\left\{\widehat{\text{OpVaR}} \geq (1-c) \times \text{OpVaR}\right\} = \alpha$ for a given value of α – n and T are fixed. Now, c is set to an "acceptable" level and we want to find $n^\star$ such that $\Pr\left\{\widehat{\text{OpVaR}} \geq (1-c) \times \text{OpVaR}\right\} \geq \alpha$ if $n \geq n^\star$. It then answers the question of the number of external losses necessary to achieve an acceptable accuracy.

Example 1 *Suppose that $\lambda = 100$, $\mu = 9$ and $\sigma = 2$. We assume that T is 5 years. Moreover, we assume that the threshold is equal to 5000 euros. In Figure 4, we report the relationship between the probability $\Pr\left\{\widehat{\text{OpVaR}} \geq (1-c) \times \text{OpVaR}\right\}$ and the number of losses n. It is now easy to find the minimum observation $n^\star$ for a given value of α. For example, if $c = 20\%$ and $\alpha = 80\%$, $n^\star$ is approximatively equal to 580 if $\mu = 9$ and $\sigma = 2$ and 940 if $\mu = 7$ and $\sigma = 2.5$. If $c = 25\%$ and $\alpha = 80\%$, $n^\star$ takes respectively the values 240 and 490, etc.*

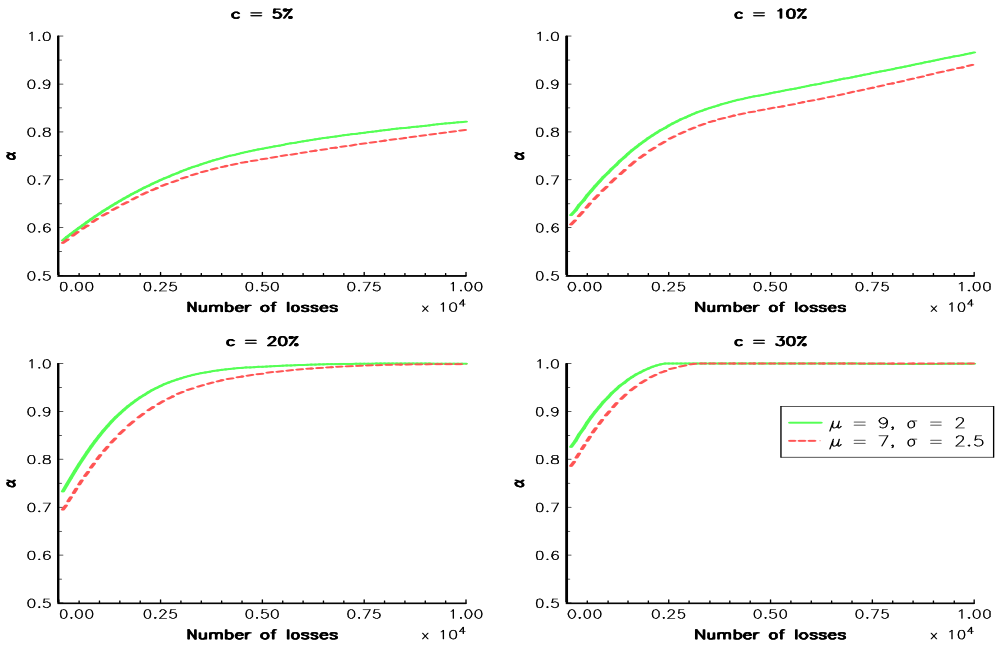

Figure 4: Relationship between the probability $\Pr\left\{\widehat{\text{OpVaR}} \geq (1-c) \times \text{OpVaR}\right\}$ and the number of losses n

Technical Appendix 6 *If we suppose the frequency as given and non random, we may find an analytical expression of* $n^\star$. *From ML theory, we recall that the estimators* $\hat{\mu}$ *and* $\hat{\sigma}$ *are approximately Gaussian with*

$$\sqrt{n}\left(\begin{pmatrix}\hat{\mu}_n\\ \hat{\sigma}_n\end{pmatrix}-\begin{pmatrix}\mu\\ \sigma\end{pmatrix}\right)\longrightarrow\mathcal{N}\left(\mathbf{0},\mathcal{J}^{-1}\right)$$

where $\mathcal{J}$ *is the Fisher information matrix. Because we define* $\widehat{\text{OpVaR}}$ *as* $\text{OpVaR}\left(\lambda,\hat{\mu}_n,\hat{\sigma}_n\right)$, *we have*

$$\sqrt{n}\left(\text{OpVaR}\left(\lambda,\hat{\mu}_n,\hat{\sigma}_n\right)-\text{OpVaR}\left(\lambda,\mu,\sigma\right)\right)\longrightarrow\mathcal{N}\left(0,\mathcal{J}_h^{-1}\right)$$

with

$$\mathcal{J}_h^{-1}=\begin{bmatrix}\partial_\mu\text{OpVaR}\left(\lambda,\hat{\mu}_n,\hat{\sigma}_n\right) & \partial_\sigma\text{OpVaR}\left(\lambda,\hat{\mu}_n,\hat{\sigma}_n\right)\end{bmatrix}\mathcal{J}^{-1}\begin{bmatrix}\partial_\mu\text{OpVaR}\left(\lambda,\hat{\mu}_n,\hat{\sigma}_n\right)\\ \partial_\sigma\text{OpVaR}\left(\lambda,\hat{\mu}_n,\hat{\sigma}_n\right)\end{bmatrix}$$

If we solve the equation

$$\Pr\left\{\widehat{\text{OpVaR}}\geq(1-c)\times\text{OpVaR}\right\}=\alpha$$

we find that

$$n^\star=\left(\frac{\Phi^{-1}\left(\alpha\right)}{c\times\text{OpVaR}\left(\lambda,\mu,\sigma\right)}\right)^2\mathcal{J}_h^{-1}$$

If we apply this formula to our previous example, we obtain the following results:

	α			
c	60%	70%	80%	90%
10%	240	1040	2700	6260
20%	60	260	670	1560
30%	20	110	300	690
40%		60	160	390
50%		40	100	250

$\lambda = 100,\ \mu = 9,\ \sigma = 2,\ H = 0$

	α			
c	60%	70%	80%	90%
10%	400	1740	4500	10430
20%	100	430	1120	2600
30%	40	190	500	1150
40%	20	100	280	650
50%	10	60	180	410

$\lambda = 100,\ \mu = 9,\ \sigma = 2,\ H = 5000$

	α			
c	*60%*	*70%*	*80%*	*90%*
10%	*410*	*1760*	*4540*	*10530*
20%	*100*	*440*	*1130*	*2630*
30%	*40*	*190*	*500*	*1170*
40%	*20*	*110*	*280*	*650*
50%	*10*	*70*	*180*	*420*

$\lambda = 100$, $\mu = 7$, $\sigma = 2.5$, $H = 0$

	α			
c	*60%*	*70%*	*80%*	*90%*
10%	*820*	*3530*	*9110*	*21120*
20%	*200*	*880*	*2270*	*5280*
30%	*90*	*390*	*1010*	*2340*
40%	*50*	*220*	*560*	*1320*
50%	*30*	*140*	*360*	*840*

$\lambda = 100$, $\mu = 7$, $\sigma = 2.5$, $H = 5000$

7 Self Assesment and Scenario Analysis

The concept of Scenario Analysis should deserve further clarification. Roughly speaking, when we refer to Scenario Analysis, we want to express the idea that banks' experts and experienced managers have some reliable intuitions on the riskiness of their business and that these intuitions are not entirely reflected in the bank's historical, internal data. As a first requirement, we expect that experts should have the opportunity to give their approval to capital charge results. In a second step, one can imagine that experts' intuitions are directly plugged into severity and frequency estimations.

Experts' intuition can be captured through scenario building. More precisely, a scenario is given by a potential loss amount and the corresponding probability of occurrence. As an example, an expert may assert that a loss of (say) one million euros or higher is expected to occur once every (say) 5 years. This is a valuable information in many cases, either when loss data are rare and do not allow for statistically sound results or when historical loss data are not sufficiently forward-looking. The issue which has to be addressed is how one can extract useful information from experts' scenarios and how it can be plugged into a conventional LDA framework.

It is quite easy if we notice that scenarios can be translated into restrictions on the parameters of frequency and severity distributions. Once these restrictions have been identified, a calibration strategy can be designed where parameters are calibrated by maximizing some standard criterion (such as Maximum Likelihood) subject to these restrictions being satisfied (at least approximately). As a result, parameter estimators can be seen as a mixture of the loss-data-based estimator and the scenario-based implied estimator. The following appendix details how it can be done provided one is able to weigh scenario-based information relatively to loss-data-based information.

Technical Appendix 7 *Let us consider a scenario defined as: "a loss of x or higher occurs once every d years". Let us also assume that the frequency distribution is a Poisson distribution (with parameter λ) and that the severity distribution is a lognormal distribution (with parameters μ and σ). With these notations, λ is the average number of losses per year, $\lambda \times (1 - \mathbf{F}(x; \mu, \sigma))$ is the average number of losses higher than x and finally $\frac{1}{\lambda \times (1 - \mathbf{F}(x;\mu,\sigma))}$ is the average duration[3] between two losses exceeding x. As a result, for a given scenario (x, d), parameters are restricted to satisfy:*

$$d = \frac{1}{\lambda \times (1 - \mathbf{F}(x; \mu, \sigma))}$$

It is obvious that three different scenarios suffice to calibrate the three parameters λ, μ, σ. Suppose that we face to different scenarios $\{(x_j, d_j), j = 1, \ldots, p\}$. We may estimate the implied parameters underlying the expert judgements using a quadratic creterion:

$$(\hat{\mu}, \hat{\sigma}) = \arg\min \sum_{j=1}^{p} w_j \left(d_j - \frac{1}{\lambda \times (1 - \mathbf{F}(x_j; \mu, \sigma))} \right)^2$$

[3] A rigourous proof of this result is given in the appendix.

where w_j is the weight associated to the j^{th} scenario. Our experience shows that it works best with standard optimal weights (ie. proportional to the inverse of the variance of d_j). Let us consider the following example:

x (in millions of euros)	*1*	*2.5*	*5*	*7.5*	*10*	*20*
d (in years)	$\frac{1}{4}$	*1*	*3*	*6*	*10*	*40*

Using the standard optimal weights, we obtain $\hat{\lambda} = 654$, $\hat{\mu} = 8.60$ and $\hat{\sigma} = 2.08$. We may compare directly these estimates to those calibrated using loss data. Moreover, if loss data are available, calibration can be achieved by maximizing some criterion[4] *obtained as a combination of maximum likelihood and the previous restrictions:*

$$\max_{(\mu,\sigma)} (1-\varpi)\,\ell_n(\mu,\sigma) - \varpi \sum_{j=1}^{p} w_j \left(d_j - \frac{1}{\lambda \times (1 - \mathbf{F}(x_j;\mu,\sigma))} \right)^2$$

where ϖ is a weight reflecting the confidence one places on expert's judgements.

8 Conclusion

This chapter was aimed at providing a comprehensive survey of all technical issues raised in the course of an LDA implementation. Though technical, these issues are nevertheless important as they have a major impact on capital charges if incorrectly tackled. Our experience has taught us that reporting bias is probably one of the most proeminent issues but fortunately can be overcome by appropriate maximum likelihood techniques. Secondly, we have shown that confidence intervals are very useful tools to address some issues like data sufficiency. Furthermore, we have derived an approximate though reliable way to compute this confidence interval. Finally, further research will have to focus on scaling issues and goodness of fit tests as databases become larger.

[4] It corresponds to the Penalized Maximum Likelihood method.

References

[1] **Basel Committee on Banking Supervision**, 2001, Operational Risk — Consultative Document, Supporting document to the New Basel Capital Accord, January.

[2] **Basel Committee on Banking Supervision**, 2001, Working Paper on the Regulatory Treatment of Operational Risk, September.

[3] **Basel Committee on Banking Supervision**, 2002, Quantitative Impact Study 3. Technical Guidance, October.

[4] **Baud, N., A. Frachot, and T. Roncalli,** 2002, "Internal data, external data and consortium data for operational risk measurement: How to pool data properly?", Working Paper, Crédit Lyonnais, Groupe de Recherche Opérationnelle.

[5] **Baud, N., A. Frachot, and T. Roncalli**, 2003, "How to avoid over-estimating capital charge for operational risk?", *OperationalRisk – Risk's Newsletter,* February.

[6] **Davidson, R. and J. MacKinnon**, 1993, *Estimation and Inference in Econometrics*, Oxford University Press.

[7] **De Fontnouvelle, P., V. DeJesus-Rueff, J. Jordan, and E. Rosengren**, 2003, "Using loss data to quantify operational risk", *Working Paper*, Federal Reserve Bank of Boston, Department of Supervision and Regulation.

[8] **Frachot, A., P. Georges and T. Roncalli**, 2001, "Loss Distribution Approach for operational risk", *Working Paper*, Crédit Lyonnais, Groupe de Recherche Opérationnelle.

[9] **Frachot, A. and T. Roncalli**, 2002, "Mixing internal and external data for managing operational risk", *Working Paper*, Crédit Lyonnais, Groupe de Recherche Opérationnelle.

[10] **Shih, J., A. Samad-Khan and P. Medapa**, 2000, "Is the size of operational loss related to firm size?", *Operational Risk*, January.

A The Method of Maximum Likelihood

The method of maximum likelihood is a very popular estimation technique. We recall here some results which can be found in every handbook on statistical estimation (see for example Davidson and MacKinnon, 1993).

Let θ be the vector of parameters to be estimated and Θ the parameter space. The likelihood for the i^{th} observation, that is the probability density of the observation i considered as a function of θ, is denoted $L_i(\theta)$. Let $\ell_i(\theta) \equiv \ln L_i(\theta)$ be the loglikelihood of $L_i(\theta)$. Given n independent observations, the loglikelihood function is

$$\ell_n(\theta) = \sum_{i=1}^{n} \ell_i(\theta)$$

$\hat{\theta}_n$ is the maximum likelihood estimator if

$$\ell_n\left(\hat{\theta}_n\right) \geq \ell_n(\theta) \qquad \forall\, \theta \in \Theta$$

The main properties of the ML estimator are *consistency*, *asymptotic normality* and *asymptotic efficiency*. In particular, we have

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \longrightarrow \mathcal{N}\left(\mathbf{0}, \mathcal{J}^{-1}(\theta_0)\right)$$

with $\mathcal{J}(\theta_0)$ the Fisher information matrix and θ_0 the 'true' value of the vector of parameters. We remind that

$$\mathcal{J}(\theta_0) = E_{\theta_0}\left[-\frac{\partial^2 \ell_i(\theta_0)}{\partial\theta\,\partial\theta^\top}\right]$$

Let $h(\theta)$ be a real function of the vector of parameters θ. Then, $h\left(\hat{\theta}_n\right)$ converges almost surely to $h(\theta_0)$ and we have

$$\sqrt{n}\left(h\left(\hat{\theta}_n\right) - h(\theta_0)\right) \longrightarrow \mathcal{N}\left(\mathbf{0}, \frac{\partial h(\theta_0)}{\partial\theta^\top}\mathcal{J}^{-1}(\theta_0)\frac{\partial h(\theta_0)}{\partial\theta}\right)$$

B The Distribution of the Duration between two losses exceeding a given value x

We assume that the number of losses is a Poisson process with intensity λ. We note T_i the time when the i^{th} loss occurs. It means that the durations $e_i = T_i - T_{i-1}$ between two consecutive losses are independent and exponential with parameter λ. We assume that the losses ζ_i are *i.i.d.* with distribution $\mathbf{F}$. We note now d_j the duration between two losses exceeding x. It is obvious that the durations are *i.i.d.* It suffice now to characterize d_1. We have[5]

$$\begin{aligned}
\Pr\{d_1 > t\} &= \sum_{i\geq 1} \Pr\{T_i > t; \zeta_1 < x, \ldots, \zeta_{i-1} < x; \zeta_i \geq x\} \\
&= \sum_{i\geq 1} \Pr\{T_i > t\}\, \mathbf{F}(x)^{i-1}(1 - \mathbf{F}(x)) \\
&= \sum_{i\geq 1} (1 - \mathbf{F}(x))\, \mathbf{F}(x)^{i-1} \sum_{k=0}^{i-1} e^{-\lambda t}\frac{(\lambda t)^k}{k!} \\
&= (1 - \mathbf{F}(x)) \sum_{k=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^k}{k!} \sum_{i=k}^{\infty} \mathbf{F}(x)^i \\
&= e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!}\mathbf{F}(x)^k \\
&= e^{-\lambda(1-\mathbf{F}(x))t}
\end{aligned}$$

[5] To establish this result, we use the fact that a finite sum of exponential times is an Erlang distribution.

It comes that d_1 follows an exponential distribution with parameter $\lambda(1-\mathbf{F}(x))$. The average duration between two losses exceeding x is also the mean of d_1 or $\frac{1}{\lambda(1-\mathbf{F}(x))}$.

In the next figure, we have simulated a Poisson process with a parameter λ equal to 5. The losses are lognormal distribution with $\mu = 9$ and $\sigma = 2$. The circles indicate the times when the loss exceeds 100 000 euros.

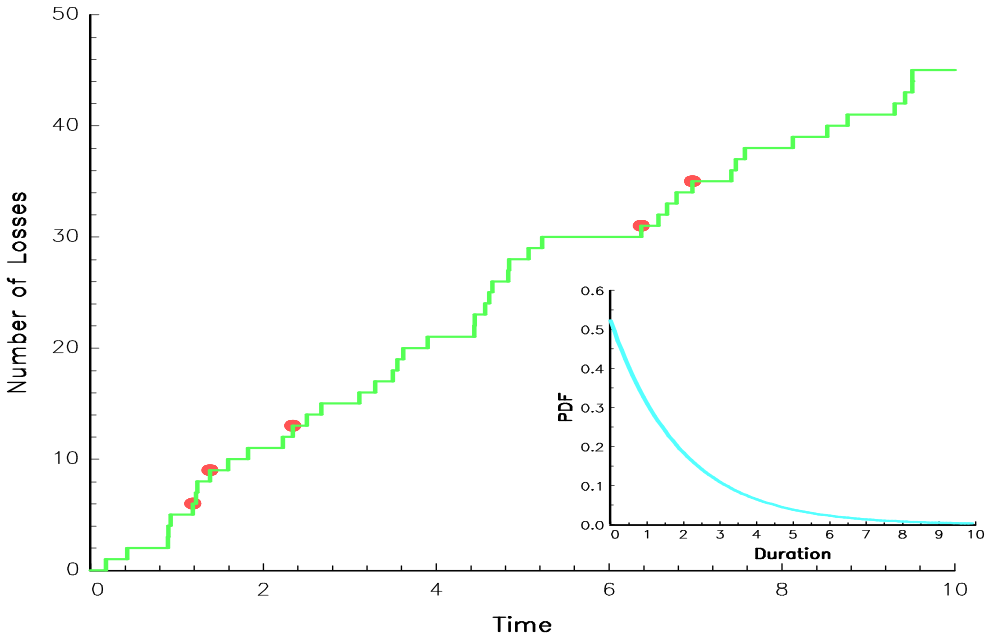

Figure 5: An example of Poisson process