Operational VaR: a Closed-Form Approximation

Klaus Böcker Claudia Klüppelberg

Abstract

Klaus Böcker and Claudia Klüppelberg investigate a simple loss distribution model for operational risk. They show that, when loss data are heavy-tailed (which in practice they are), a simple closed-form approximation for the OpVaR can be obtained. They apply this approximation in particular to the Pareto severity model, for which they obtain also a simple time scaling rule for the operational VaR.

1 Introduction

According to the final proposals of the Basel Committee [1], in addition to market and credit risk, operational risk will also be a determinant of the new capital requirements as from 2007. Then every bank has to calculate explicit capital charges to cover operational risk by means of one of three approaches: the basic indicator approach, the standardised approach, and the advanced measurement approach (AMA). This "continuum of approaches" reflects different levels of sophistication and risk sensitivity. AMA as the most flexible approach for operational risk quantification allows the bank to build its own internal operational risk model and measurement system, comparable to market risk standards. Instead of prescribing a particular type of VaR model, the committee requires a set of quantitative and qualitative standards to be fulfilled. The following two are most relevant for the issues discussed in this paper:

(1) The operational-risk measure is a VaR at confidence level 99.9 % with a holding period of one year (cf. § 667)

(2) The measurement approach must capture potentially severe tail loss events (cf. § 667).

The most popular method in the industry to satisfy the AMA standards is the loss distribution approach (LDA), which is based on modelling the probability distribution of operational losses using bank-internal and external data.

Despite the current vivid debate, operational risk is not a new phenomenon. We all recall the operational risk event that happened on February 26, 1995, when the prestigious British merchant bank Barings had to declare bankruptcy. The reason was an accumulated loss of £ 625 million in its Singapore division caused by a trader, who was hiding loss-making positions in financial derivatives.

Obviously, the main objective of any financial institution is to manage and minimize operational risk. The only feasible way to manage risk is by identifying and minimize it. This can only be done successfully by the development of adequate quantification techniques. It seems to be worthwhile to consider the actuaries' approach to similar problems. Dealing with randomly occurring insurance claims, they have been at the very core of operational-risk-like issues for more than a hundred years. Hence, it is not surprising that LDA models have their roots in *insurance risk theory*, which goes back to the early work by Filip Lundberg in 1903. In this respect, modelling and quantifying operational risk can be viewed as a 100 year old science!

In this paper we suggest and investigate a model that indeed originates in insurance. We first introduce a Standard LDA, which contains the compound Poisson and the negative binomial model as special cases. Exploiting the common wisdom that severity distributions for operational risk are typically very heavy-tailed, we derive a closed-form approximation for the operational VaR (OpVaR). Extending the model from a static model to a dynamic one, we further show that such models have an in-built α-root-of-time rule for some $\alpha > 0$, which usually differs from the well-known square-root law. Finally, we introduce a new simple OpVaR estimate, which can be applied for scenario generation or for expert-based risk assessment.

2 The Loss Distribution Approach

In the context of LDA models, a widely accepted procedure consists of splitting up the total loss amount over a certain period into a frequency component, i.e. the number of losses, and a severity component, i.e. the individual loss amounts. The total loss is then obtained by compounding the frequency and the severity information. A prototypical model of this kind, which is currently best practise and implemented in various commercial software packages, is the following.

Definition 2.1. *(Standard LDA)*

(1) The severity process:
The severities $(X_k)_{k\in\mathbb{N}}$ *are positive independent and identically distributed (iid) random variables describing the magnitude of each loss event.*

(2) The frequency process:
The number $N(t)$ *of loss events in the time interval* $[0,t]$ *for* $t \geq 0$ *is random. The resulting* counting process $(N(t))_{t\geq 0}$, *is generated by a sequence of points* $(T_n)_{n\geq 1}$

of non-negative random variables satisfying

$$0 \le T_1 \le T_2 \le \cdots \quad a.s.$$

and

$$N(t) = \sup\{n \ge 1 : T_n \le t\}\,, \quad t \ge 0\,.$$

(3) *The severity process and the frequency process are assumed to be independent.*

(4) The aggregate loss process:
The aggregate loss $S(t)$ up to time t constitutes a process

$$S(t) = \sum_{i=1}^{N(t)} X_i\,, \quad t \ge 0\,.$$

Note that we do not require X_k to have finite mean and/or variance. This is in accordance with empirical research: Moscadelli [4] showed very convincingly that typical severity distributions for operational risk are very heavy-tailed such that even moments of low order may not exist; see also Section 2.2.

Typical examples for LDA models are obtained by specifying the frequency process in the following way:

Example 2.2. (a) *The* Poisson-LDA *is a Standard LDA, where $(N(t))_{t\ge0}$ is a homogenous Poisson process with intensity $\lambda > 0$, in particular,*

$$P(N(t) = n) = p_t(n) = e^{-\lambda t}\,\frac{(\lambda t)^n}{n!}, \quad n \in \mathbb{N}_0\,.$$

(b) *The* negative-binomial-LDA *is a Standard LDA, where $(N(t))_{t\ge0}$ is given by a negative binomial process satisfying for $\beta, \gamma > 0$*

$$P(N(t) = n) = p_t(n) = \binom{\gamma + n - 1}{n}\left(\frac{\beta}{\beta + t}\right)^{\gamma}\left(\frac{t}{\beta + t}\right)^{n}, \quad n \in \mathbb{N}_0\,.$$

The negative binomial distribution is a gamma mixture of a Poisson distribution, i.e. it can be viewed as a Poisson distribution whose parameter λ is a gamma distributed random variable. This allows for modelling *over-dispersion*, which means that for all $t > 0$ the variance of $N(t)$ is greater than its mean, whereas for the Poisson-LDA $\mathrm{var}(N(t)) = EN(t)$ holds. However, we will see later that as far as the OpVaR approximation is concerned, over-dispersion is of minor importance.

2.1 Subexponential Severity Distributions

Concerning the severity, we have to take—in accordance with Basel II—the heavy-tail property of operational losses into account. Some popular distributions are given in Table 1. All of them are heavy-tailed, more precisely, they belong to the class

Name	Distribution function	Parameters
Lognormal	$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$	$\mu \in \mathbb{R},\ \sigma > 0$
Weibull	$F(x) = 1 - e^{-(x/\theta)^\tau}$	$\theta > 0,\ 0 < \tau < 1$
Pareto	$F(x) = 1 - \left(1 + \frac{x}{\theta}\right)^{-\alpha}$	$\alpha, \theta > 0$

Table 1: Popular severity distributions with support $(0, \infty)$. (Φ is the standard normal distribution function).

of so-called *subexponential distributions*, meaning that their tails decay slower than any exponential tail.

The defining property of subexponential distributions is that the tail of the sum of n subexponential random variables has the same order of magnitude as the tail of the maximum variable among them, more precisely,

$$\lim_{x\to\infty} \frac{P(X_1 + \cdots + X_n > x)}{P(\max(X_1, \ldots, X_n) > x)} = 1 \quad \text{for some (all) } n \geq 2. \tag{1}$$

This means that the sum of n iid severities is most likely to be large because of one of the terms being large, or, with emphasis on operational risk, severe overall losses are mainly due to a single big loss rather than the consequence of accumulated small independent losses. Of course, this insight should have implications for operational risk management.

The goal of every LDA model is to determine the *aggregate loss distribution*, which for the Standard LDA can be written as

$$\begin{aligned} G_t(x) &= P(S(t) \leq x) \\ &= \sum_{n=0}^{\infty} p_t(n)\, P(S(t) \leq x | N(t) = n) \\ &= \sum_{n=0}^{\infty} p_t(n)\, F^{n*}(x), \quad x \geq 0, \quad t \geq 0\,, \end{aligned}$$

where $F(\cdot) = P(X_k \leq \cdot)$ is the distribution function of X_k, and $F^{n*}(\cdot) = P(\sum_{i=1}^n X_i \leq \cdot)$ is the n-fold convolution of F with $F^{1*} = F$ and $F^{0*} = I_{[0,\infty)}$.

For most choices of severity and frequency distributions, G_t cannot be calculated analytically. Approximation methods to overcome this problem include Panjer recursion, Monte Carlo simulation, and FFT (fast Fourier transform) methods, see Klugman, Panjer and Willmot [3] for an overview. The drawback of these methods is, however, that their results remain a "black box", and the interaction between different model parameters and their impact on the final result, i.e. the OpVaR is only interpretable through extensive sensitivity analyses.

As both regulatory capital and economic capital are based on a very high quantile of the aggregate loss distribution G_t, a natural estimation method for OpVaR is via asymptotic tail and quantile estimation. Instead of considering the entire distribution, it is sufficient to concentrate on the right tail $P(S(t) > x)$ for very large x. Now, in actuarial science, the tail behavior of G_t has been extensively studied both for small claims and large claims models. For the latter, a key result states that for a Standard LDA with subexponential severities one has under weak regularity conditions (see Theorem A.1, equation (A.1)) for every fixed $t > 0$,

$$\overline{G}_t(x) \sim EN(t)\overline{F}(x)\,, \quad x \to \infty\,, \tag{2}$$

where $EN(t)$ is the expected frequency and $\overline{F}(\cdot) = 1 - F(\cdot)$ and $\overline{G}_t(\cdot) = 1 - G_t(\cdot)$ are the tail distributions of severity and aggregrate loss, respectively. The symbol $\sim$ means that the quotient of right hand and left hand side tends to 1; i.e. $\lim_{x\to\infty} \overline{G}_t(x)/\overline{F}(x) = EN(t)$ for every fixed $t > 0$.

It has been shown in Examples 1.3.10 and 1.3.11 of Embrechts, Klüppelberg and Mikosch [2] that the tail estimate (2) holds for the Poisson-LDA and the negative-binomial-LDA.

2.2 A Closed-Form Approximation for OpVar

Given relation (2) it is straightforward to obtain an expression for the OpVaR, valid at very high confidence levels. Recall that VaR is just a quantile of a distribution function.

Definition 2.3 (Value at Risk). *Suppose G_t is the aggregate loss distribution. Then the* Value-at-Risk (VaR) *up to time t at confidence level κ is defined as the κ-quantile of the loss distribution:*

$$VaR_t(\kappa) = G_t^{\leftarrow}(\kappa)\,, \quad \kappa \in (0,1)\,,$$

where $G_t^{\leftarrow}(\kappa) = \inf\{x \in \mathbb{R} : \; G_t(x) \geq \kappa\}$, $0 < \kappa < 1$, is the generalized inverse of G_t. If G_t is strictly increasing and continuous, we may write $VaR_t(\kappa) = G_t^{-1}(\kappa)$.

Using (2) we obtain an asymptotic formula for the OpVaR:

Theorem 2.4 (Analytical OpVaR). *Consider the Standard LDA model for fixed $t > 0$ and a subexponential severity with distribution function F. Assume, moreover, that the tail estimate (2) holds. Then, the $VaR_t(\kappa)$ satisfies the approximation*

$$VaR_t(\kappa) = F^{\leftarrow}\left(1 - \frac{1-\kappa}{EN(t)}(1+o(1))\right), \quad \kappa \to 1\,. \tag{3}$$

Proof. Note first that $\kappa \to 1$ is equivalent to $x \to \infty$. Then recall that $o(1)$ always stands for a function, which tends to 0, if its argument tends to a boundary, in our case if $\kappa \to 1$ or $x \to \infty$. With this notation relation (2) can be rewritten as

$$G_t(x) = 1 - EN(t)\,\overline{F}(x)(1 + o(1))\,, \quad x \to \infty\,.$$

Setting the right hand side equal to κ gives an asymptotic solution

$$F(x) = 1 - \frac{1-\kappa}{EN(t)}(1 + o(1))\,, \quad x \to \infty\,,$$

and, finally,

$$x = G_t^{\leftarrow}(\kappa) = F^{\leftarrow}\left(1 - \frac{1-\kappa}{EN(t)}(1 + o(1))\right), \quad \kappa \to 1\,. \qquad \square$$

This result, which holds for a quite general class of LDA models, is remarkable for two reasons. First, it says that the OpVaR at high confidence levels only depends on the tail and not on the body of the severity distribution. Therefore, if one is only interested in VaR calculations, modelling the whole distribution function F is superfluous. Second, because the frequency enters in expression (3) only with its expectation $EN(t)$, it is also not necessary to calibrate a specific counting process; estimating the sample mean of the frequency suffices. As a consequence thereof, over-dispersion as modelled by the negative binomial distribution, has asymptotically no impact on the OpVaR.

In order to obtain a first order approximation for the OpVaR for a specific LDA model, it suffices to combine (3) with the tail of the (subexponential) severity distribution F. Furthermore, even closed-form solutions for the (asymptotic) OpVaR are available (see Table 2).

Finally, we want to emphasize that the problem of finding a severity distribution that accurately describes empirical loss data is a non-trivial task, and that the parameterization of appropriate severity and frequency distributions is an integral part of every AMA model. A textbook treatment concerning such statistical issues as data analysis, parameter estimation, and hypothesis testing in the context of general loss models can be found in Klugman, Panjer and Willmot [3].

3 The Pareto Severity Model

Operational loss data are usually very heavy-tailed. Moscadelli [4] investigated empirical loss data collected by the Basel Committee during the financial year 2001. Motivated by extreme value methods, for a generalized Pareto distribution model

Name	$\mathrm{VaR}_t(\kappa)$
Lognormal	$\exp\left[\mu - \sigma\,\Phi^{-1}\left(\frac{1-\kappa}{EN(t)}\right)\right]$
Weibull	$\theta\left[\ln\left(\frac{EN(t)}{1-\kappa}\right)\right]^{\frac{1}{\tau}}$
Pareto	$\theta\left[\left(\frac{EN(t)}{1-\kappa}\right)^{1/\alpha} - 1\right]$

Table 2: First order approximations of the $\mathrm{VaR}_t(\kappa)$ as $\kappa \to 1$ for the aggregate loss distribution for popular severity distributions. Set $EN(t) = \lambda\, t$ for a Poisson distributed and $EN(t) = \gamma\, t/\beta$ for a negative binomially distributed severity.

(GPD model), he estimated $1/\alpha$ in a range between approximately 0.6 and 1.5, corresponding to α roughly between 0.7 and 1.7. For all such α the severity distribution has infinite variance and for $\alpha \leq 1$ even the mean value does not exist.

Recall that the GPD comprises distributions with compact support, exponential distributions and Pareto distributions, corresponding to $1/\alpha$ being negative, 0, and positive, respectively. For such small positive values of α as observed by Moscadelli above, it is quite clear that it suffices to consider the GPD family for positive finite values of α, corresponding to a Pareto distribution.

The Pareto distribution has further properties, which we will exploit in this section.

Example 3.1 (Pareto-LDA). *The Pareto-LDA is a Standard LDA as given in Definition 2.1, where the loss severities $(X_k)_{k\in\mathbb{N}}$ are Pareto distributed; i.e. for parameters $\alpha, \theta > 0$*

$$\overline{F}(x) = \left(1 + \frac{x}{\theta}\right)^{-\alpha}, \quad x > 0\,.$$

For simplicity we denote $EN(t) = \lambda t$ with the obvious understanding that for a negative binomial process λ has to be replaced by γ/β. As a result of Theorem 2.4, we obtain for the OpVaR

$$VaR_t(\kappa) \sim \theta\left(\frac{\lambda\, t}{1-\kappa}\right)^{1/\alpha}, \quad \kappa \to 1\,. \tag{4}$$

(Actually, any severity distribution satisfying $\overline{F}(x) \sim (x/\theta)^{-\alpha}$ as $x \to \infty$ yields approximation (4)).

Figure 1 compares the analytical VaR estimate (4) with the results of a Monte Carlo simulation for the Pareto-LDA with different shape parameters α and $\theta = 1$. We see that the best approximation is obtained for extremely heavy-tailed data, i.e.

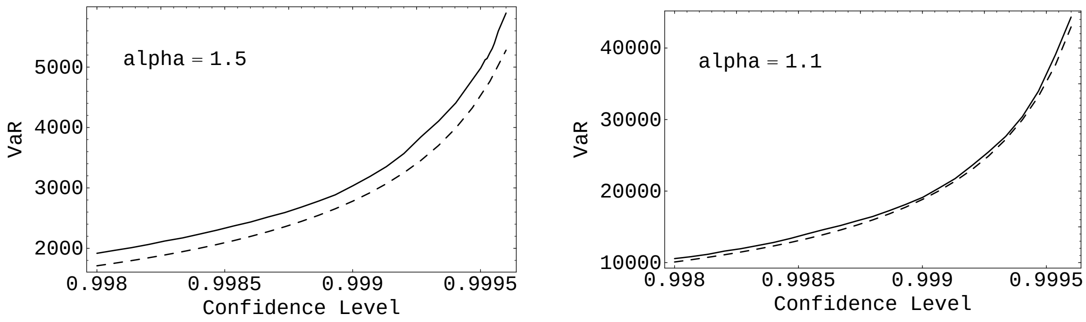

Figure 1: Comparison of the approximted VaR given by (4) (dashed line) and the simulated VaR (solid line) for the Pareto-Poisson LDA with $\theta = 1$.

for small values of α. Consequently, for operational loss data, our approximation should be very good.

3.1 Time Scaling in the Pareto Severity Model

A well-known formula in risk management is the square-root-of-time rule for deriving multi-period VaR values from 1-period values. This scaling law is based on the well-known property of the normal distribution, which says that the sum of n iid centered normal random variables, when scaled by $\sqrt{n}$ is again normally distributed. As a generalisation, the central limit theorem guarantees that the sum of n iid random variables with finite variance (with arbitrary distribution and centered by its mean) converges for $n \to \infty$, when scaled by $\sqrt{n}$, to a standard normal distribution. It can be shown that the central limit theorem holds also for Pareto-LDA models, when proper adjustments have been made for the random number $N(t)$ of summands; see Embrechts, Klüppelberg and Mikosch [2], Theorems 2.5.7 and 2.5.9. Note that for $\alpha < 2$ neither is scaling by $\sqrt{n}$ correct nor does the normal distribution appear as a limit for $n \to \infty$. Instead scaling has to follow a $1/\alpha$-root and the limit is a so-called stable distribution, which is much heavier-tailed than the normal law.

We are, however, not aiming at a limit law for $n \to \infty$, respectively $N(t) \to \infty$ (which means $t \to \infty$), but for a simple multi-period VaR based on 1-period values. Moreover, we consider approximations in the very far out tail of a heavy-tailed distribution. Consequently, a central limit argument may be misleading, and scaling with the square-root factor is even for a finite variance model not justified.

We may, however, infer from (4) that for all fixed $t > 0$,

$$\mathrm{VaR}_t(\kappa) \sim t^{1/\alpha}\,\mathrm{VaR}_1(\kappa)\,, \quad \kappa \to 1\,. \tag{5}$$

Consequently, in the case of a Pareto-LDA model, we have an α-root-of-time rule for the OpVaR. Inserting typical values for α, (5) implies that the threat of losses due to operational risk increases rapidly (and much faster than the outcome of the square-root-rule) when considering future time horizons. To put it simply, operational risk can be a long-term killer!

3.2 Maxima of Operational Losses

Consider a VaR at confidence level κ and time horizon $t = 1$ year, i.e. the potential 1-year loss that is exceeded only with small probability $1 - \kappa$. From the law of large numbers we know that for large N an event with probability p occurs on average Np times in a series of N observations. Therefore, in case of yearly data, for $\kappa = 0.1\%$, VaR can be heuristically interpreted as the once-in-a-thousand-year event. There is, however, a different interpretation of VaR that is closely related to the *sample maxima* among a sequence of $N(t)$ iid loss variables X_i within a given time period $[0, t]$,

$$M(t) = \max(X_1, \ldots, X_{N(t)}), \qquad t \geq 0.$$

For the Standard LDA from Definition 2.1, setting $P(N(t) = n) = p_t(n)$ and defining $M_n = \max(X_1, \ldots, X_n)$ for $n \in \mathbb{N}$, we can immediately calculate the distribution function G_M of $M(t)$ for any fixed $t > 0$.

$$G_M(x) = P(M(t) \leq x) = \sum_{n=0}^{\infty} p_t(n)\, P(M_n \leq x) = \sum_{n=0}^{\infty} p_t(n)\, F^n(x)\,, \quad x \geq 0\,.$$

Example 3.2. *(Poisson-Pareto-LDA)*
If the frequency follows a Poisson process with intensity $\lambda > 0$, we obtain

$$G_M(x) = \sum_{n=0}^{\infty} e^{-\lambda t}\, \frac{(\lambda t)^n}{n!}\, F^n(x) = e^{-\lambda\, t\, \overline{F}(x)}\,, \quad x \geq 0\,. \tag{6}$$

We now ask for the most probable value x_{mp} of the maximum, the mode *of G_M. If F has a differentiable density f with derivative f', then also G_M has a differentiable density g_M with derivative g_M'. In this case, the mode of G_M is determined as the solution x_{mp} to*

$$g_M'(x) = e^{-\lambda t\, \overline{F}(x)}\, \lambda\, t\, \left[\lambda\, t\, f^2(x) + f'(x)\right] = 0$$

and, thus, x_{mp} is the solution to

$$\lambda\, t\, f^2(x) + f'(x) = 0\,.$$

For most realistic severity distributions x_{mp} will be unique. In the important example of a Pareto distribution we have

$$x_{mp} = \theta \left[\left(\frac{\alpha\, \lambda\, t}{1 + \alpha} \right)^{1/\alpha} - 1 \right] \approx \theta \left(\frac{\alpha\, \lambda\, t}{1 + \alpha} \right)^{1/\alpha}\,. \tag{7}$$

α	κ		
	99.0 %	99.9 %	99.95 %
1.2	77	524	934
1.0	200	2 000	4 000
0.8	871	15 496	36 857

Table 3: The factor $\left(\frac{1+1/\alpha}{1-\kappa}\right)^{1/\alpha}$ of equation (8) for α and κ in a realistic range.

Note the similarity between the VaR formula (4) and the right hand side of (7). We finally arrive at the following approximate relationship between the OpVaR at time horizon t and the most probable value of the maximum loss event during that time period for κ near 1,

$$VaR_t(\kappa) \approx \left(\frac{1+1/\alpha}{1-\kappa}\right)^{1/\alpha} x_{mp} \, . \tag{8}$$

It is worth mentioning that this result does not depend on the frequency process, but only on the shape parameter α and the confidence level κ. For any given x_{mp}, Table 3 clearly shows the sensitivity of the corresponding OpVar of the shape parameter and the confidence level.

The question arises, whether (8) can be used as an alternative approximation for OpVar. Unfortunately, estimating x_{mp} by a reliable empirical method would require a vast amount of loss data, which are currently not available. The underlying data should consist of annual maximal losses for the last years, which define a histogram, from which x_{mp} can be read off. Therefore, a large amount of annual maxima would have to be collected before x_{mp} could be estimated, where presumably the iid property would be violated simply by non-stationarity in a long time series.

However, the right hand side of (8) can, for instance, be estimated by scenario analyses and expert-based risk assessment. An experienced risk manager may guesstimate the maximum-one-year loss caused by a single event within the next year. Annual maximal losses of previous years may guide the way. Such estimates, interpreted as the most probable value x_{mp}, then yield an expert-approximation of the OpVaR as it is required by the Basel Committee.

References

[1] Basel Committee on Banking Supervision (2004) *International Convergence of Capital Measurement and Capital Standards.* Basel.

[2] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[3] Klugman, S., Panjer, H. and Willmot, G. (2004) *Loss Models - From Data to Decisions.* Wiley, Hoboken, New Jersey.

[4] Moscadelli, M. (2004) *The Modelling of Operational Risk: Experience with the Analysis of the Data Collected by the Basel Committee.*
Banca D'Italia, Termini di discussione No. 517.

About the Authors

Klaus Böcker works in the risk control department at HypoVereinsbank in Munich and develops quantitative models and methods for risk integration and economic capital calculations. Claudia Klüppelberg holds the Chair of Mathematical Statistics at the Center for Mathematical Sciences of the Munich University of Technology. Her main research interest lies in the area of risk management in finance and insurance.

Acknowledgement and Disclaimer

Klaus Böcker would like to thank Martin Ottmann for stimulating discussions about OpRisk issues. The opinions expressed in this article are those of the authors and do not reflect the views of HypoVereinsbank.

Appendix: Tail Behaviour of the Aggregate Loss Distribution

The following theorem covers the Standard LDA with the two frequency models of Example 2.2.

Theorem A.1 (Embrechts, Klüppelberg and Mikosch [2], Theorem 1.3.9). *Consider the standard LDA $S(t) = \sum_{n=0}^{N(t)} X_i$, $t \geq 0$, from Definition 2.1. Assume that the severities X_i are subexponential with distribution function F. Fix $t > 0$ and define the frequency distribution by $P(N(t) = n) = p_t(n)$ for $n \in \mathbb{N}_0$. Then, the aggregate loss distribution is given by*

$$G_t(x) = \sum_{n=0}^{\infty} p_t(n)\, F^{n*}(x)\,, \quad x \geq 0, \quad t \geq 0\,.$$

Assume that for some $\varepsilon > 0$,

$$\sum_{n=0}^{\infty} (1+\varepsilon)^n p_t(n) < \infty\,. \tag{A.1}$$

Then, G_t is subexponential with tail behaviour given by

$$\overline{G}_t(x) \sim EN(t)\overline{F}(x)\,, \qquad x \to \infty\,.$$