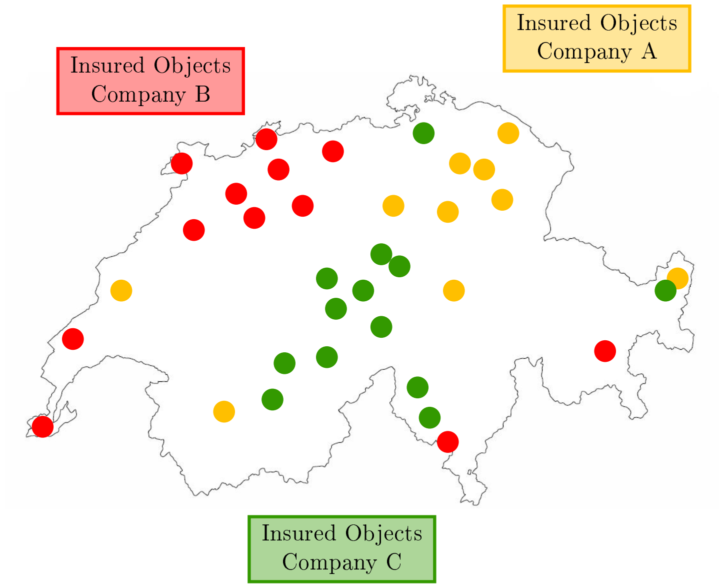
On the Transformation of Actuarial Loss Models into Synthetic NatCat Loss Tables

Simon Müller



Master's Thesis
Written at the Department of Mathematics, ETH Zürich
and SCOR Switzerland AG
14.08.2021

Supervisors: Prof. Dr. Patrick Cheridito
Dr. Philipp Arbenz

Contents

Chapter 1

Introduction

Let us consider the insurance companies A, B and C which all offer financial protection to property owners against flood damage in Switzerland (see Figure 1.1). All three companies wish to model their losses in order to charge adequate premiums. As floods belong to the category of *Natural Catastrophes*, insurers typically work with so called *NatCat models*, which need very precise exposure data such as building location and characteristics as input in order to model the losses. Modern NatCat models have in common that they produce so called *Year Loss Tables (YLT's)* as outputs, which can be used to compute *Occurrence Exceedance Probability Curves (OEP's)* and *Aggregate Exceedance Probability Curves (AEP's)*.

To be on the safe side, companies A, B and C contact the reinsurer Master Re in order to protect their portfolio against devastating floods (such as the ones in 2005 or 2021). In a classical case, Master Re can simply take the exposure data of all three companies and use it as input for a NatCat model in order to model the total loss. However, this procedure is only applicable if insurers A, B and C can provide sufficiently granular exposure data. Let us assume that companies B and C possess such information, while company A only knows the total insured value of its portfolio, but missed to collect precise exposure data. Thus, it is not possible to run a NatCat model for insurer A.

Nevertheless, Master Re wants to consider the NatCat risks of company A in a YLT setting, since it would be very convenient to use the well-established infrastructure and systems for NatCat models as well for company A. As an example, the computation of all-peril-region statistics (such as an all-peril OEP) is only possible if all NatCat risks of the reinsured companies are modelled with YLT's. Moreover, solvency considerations and portfolio management require a consistent modelling environment, see Figure 1.2.

All these arguments indicate that Master Re should integrate company A in its NatCat environment, i.e. Master Re needs to build an *alternative modelling approach* which generates a YLT out of past loss data of insurer A.

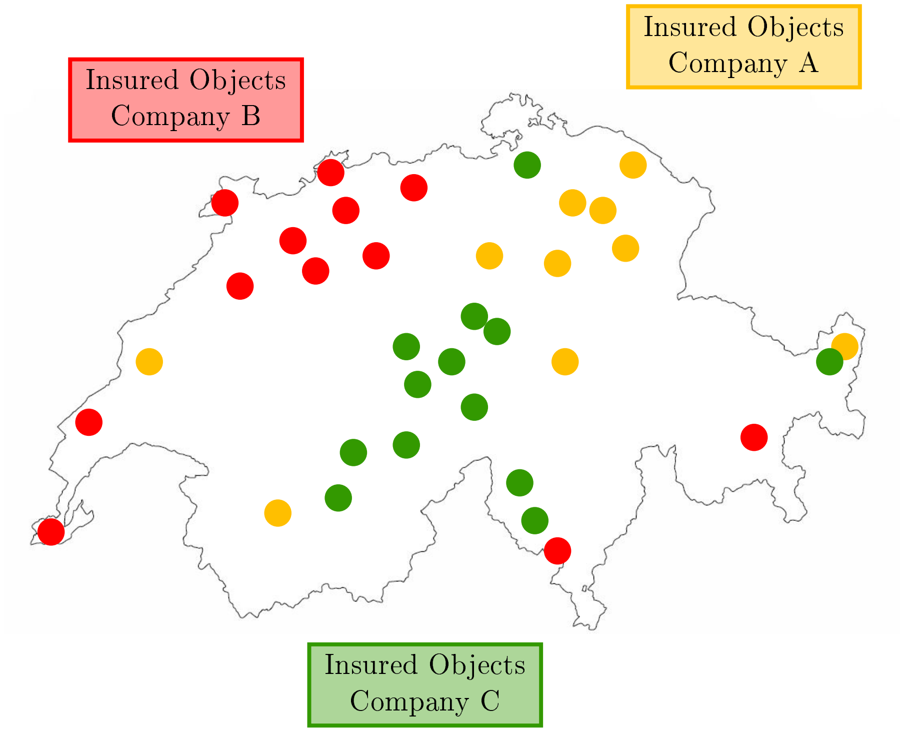


Figure 1.1: Portfolios of Insurers A, B and C

This procedure is only applied if there is no or not precise enough exposure data, since relying on past loss data is generally not a good idea in NatCat modelling (as we will explain later). Master Re decides to apply the following four transformations steps in order to model a *synthetic* YLT for insurer A (we use the term *synthetic* to distinguish the YLT's from the classical and alternative modelling approach):

1. Fit a frequency-severity model to past loss data of company A.
2. Transform the frequency-severity model into a frequency-severity model with Poisson(2) frequency.
3. Model losses for $n(= 100'000)$ years (i.e. generate a YLT).
4. Inject the YLT into a pre-generated *Year Event Rank Table (YERT)* to obtain a synthetic YLT.

Looking at this four steps critically, there might arise three questions: Why does Master Re not calibrate directly a Poisson(2) frequency-severity model to loss data and why do we impose this specific frequency choice at all? Moreover, why do we inject the YLT obtained in step 3 in a YERT and not use it directly?

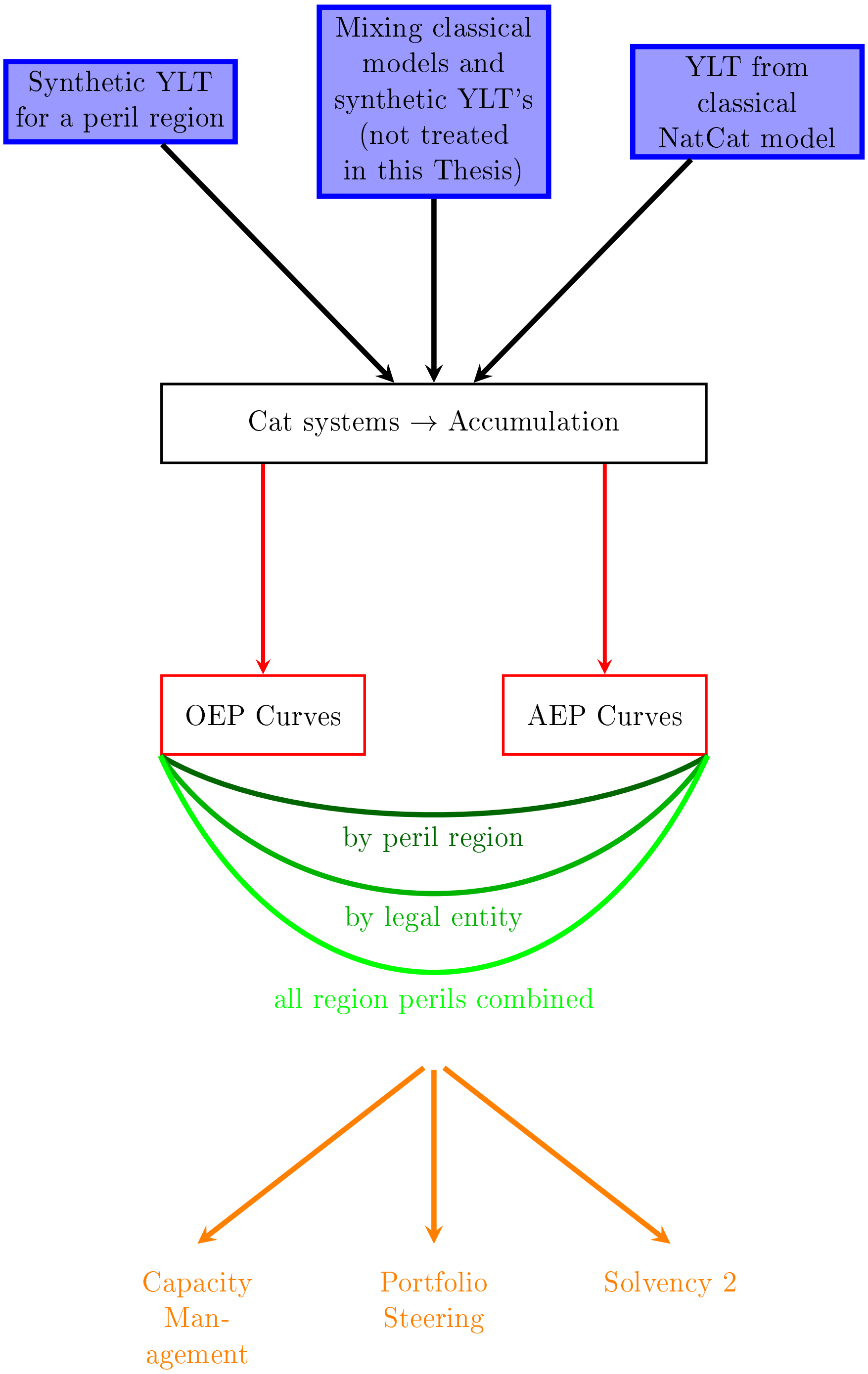


Figure 1.2: Importance of a coherent NatCat modelling environment

First of all, the reinsurer wants to include the alternative modelling approach seamlessly into its modelling framework. Allowing for general frequency-severity models, the respective actuary enjoys full modelling freedom in the first step, i.e. standard model calibration tools can be applied and it is not necessery to treat the calibration process in the first step separately.

Addressing the second question, one should point out that most NatCat vendor models work with Poisson frequencies. In order to obtain a better comparability, it is advisable to use the same distribution type as well in the alternative approach. Moreover, most natural catastrophes occur (locally) with a frequency of two or smaller, so a Poisson(2) frequency is able to "capture all losses". In summary, step 2 ensures that we obtain a coherent modelling framework which is in line with the process in vendor models.

For a detailed answer of the third question, we refer to Chapter 5. The injection into a YERT is basically performed in order to realize a certain level of compatibility.

The alternative modelling approach can not be executed without any transformation error. There are essentially three error types which arise in this context. First of all, we transform the calibrated frequency into a Poisson(2) frequency, which results in a *Frequency Transformation Error*. Moreover, we will also be forced (in general) to transform the severity in step 2 in order keep the expected loss unchanged, i.e. we obtain a *Severity Transformation Error*. Finally, some losses are always discarded in step 4 with positive probability, which gives rise to the so called *Loss Cutting Error*.

In this thesis, we analyze the alternative modelling approach for company A mathematically. In particular, we describe and investigate the three error types and try to limit them if possible by suitable tools such as reinsurance layers. We start with a brief description of classical NatCat models in Chapter 2. Afterwards, we discuss in Chapter 3 frequency-severity models and their most relevant properties. Chapter 4 investigates how the frequency-severity model that was calibrated to loss data of insurer A can be approximated with a frequency-severity model with Poisson(2) frequency. Finally, Chapter 5 discusses the transformation of frequency-severity models into synthetic NatCat loss tables via a pre-generated YERT.

Figure 1.3 visualizes the classical modelling process (green: exposure data available) as well as the alternative approach (orange: no exposure data available).

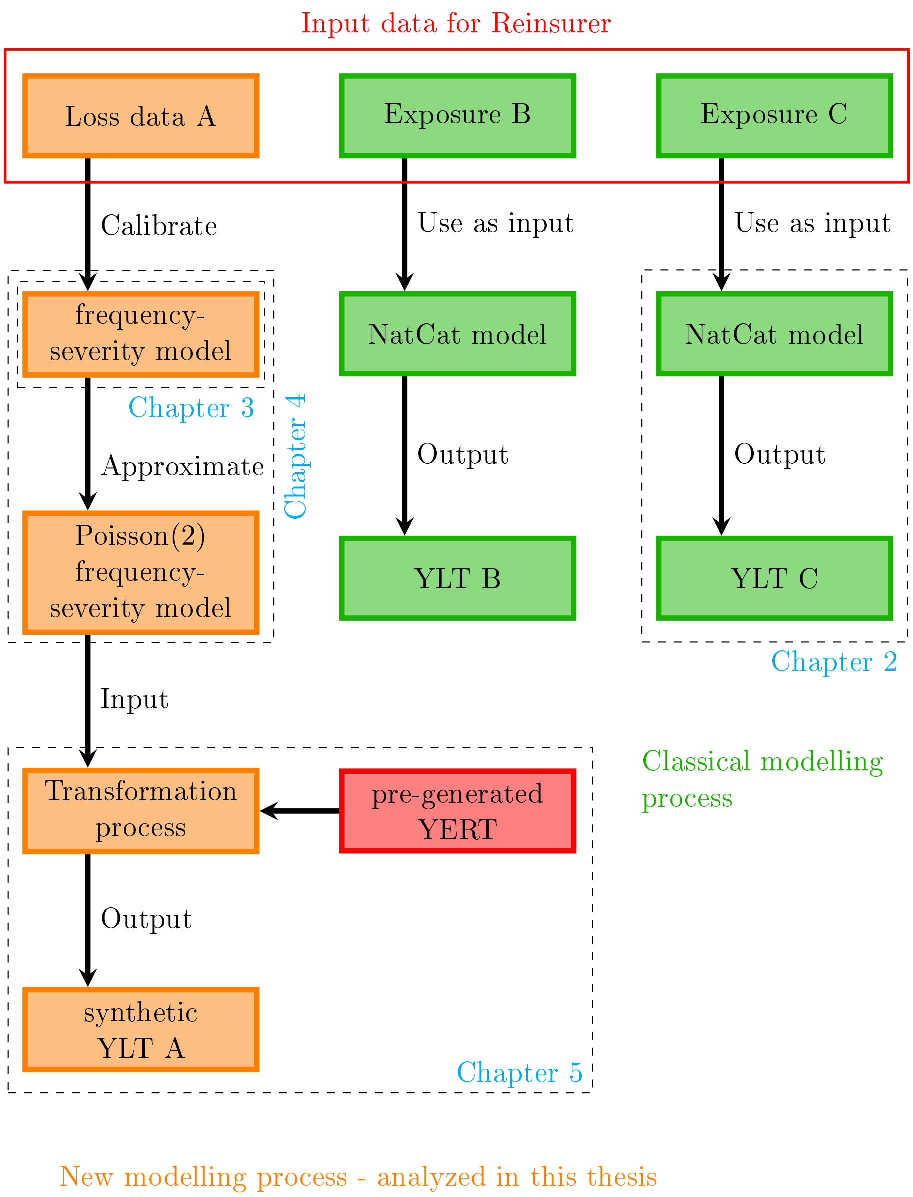


Figure 1.3: NatCat modelling process

Chapter 2

Classical Modelling of Natural Catastrophes

Let us begin this Thesis with a brief description of classical NatCat models. In this Chapter, we follow Arbenz [6], pages 83-109.

2.1 Structure of NatCat models

One might be tempted to calibrate frequency-severity models to loss data for the modelling of natural catastrophes, as it is usually done in a Non-Life Insurance context. Unfortunately, this does not turn out to be a good idea; Since the occurrence probabilities of events, the number of insured objects, the damage incurred to insured objects and the policy conditions are *not* constant over time, the calibration of a frequency-severity model to loss data will not lead to satisfying results. Moreover, certain catastrophes such as earthquakes occur (locally) with a very low frequency, such that there are not enough data points to calibrate a frequency-severity model, or the time difference between two data points is so big that the losses of the events can not be compared. Therefore, NatCat models generally follow a different approch; they are built out of 4 modules:

- The **Hazard module** models the local intensity and frequency of the physical effects of the underlying peril (e.g. earthquake: peak acceleration and velocity by location).
- In the **Exposure module**, exposure data such as (insured) value, building characteristics, location and policy conditions of the insured objects is collected.
- The **Vulnerability module** combines the results from the first two modules and models the groundup loss to the insured objects caused by the simulated hazards. This is done via the so called *vulnerability*

curves, which map the intensity of an event (e.g. earthquake magnitude) to the mean damage ratio of the object. The choice of the curve is based on the exposure of the respective object.

- In the **Financial module**, policy conditions and preceeding (re-) insurance contracts (such as Quota shares and Excess of Loss contracts (XL's), but *no* Catastrophe Excess of Loss contracts (Cat XL's)) are applied to the groundup loss to compute the *Net Pre-Cat loss*. Application of Cat XL's then gives the reinsured Cat loss, which is the quantity of interest in this context.

Figure 2.1 shows some vulnerability curves for different buildings for a wind storm. Depending on the building characteristics, the mean damage ratio (= average loss percentage of the sum insured) can vary significantly, i.e. recording precise information about the insured objects plays a crucial role.

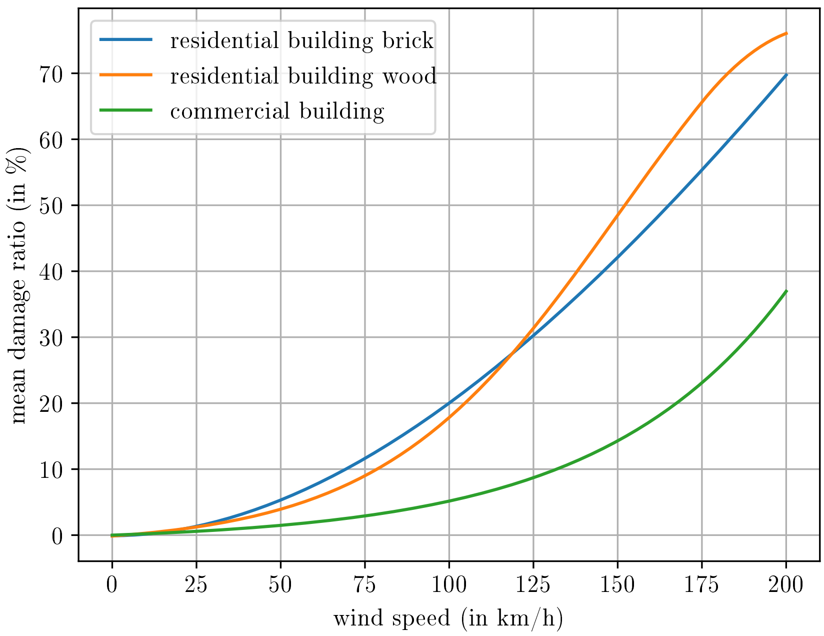


Figure 2.1: Vulnerability curves for different building types and wind speeds

In summary, NatCat models aim to model the reinsured Cat loss using physical models of the reinsured perils and are generally not based on past loss experiece.

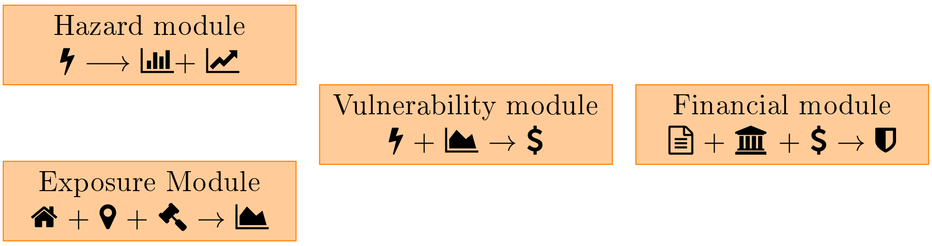


Figure 2.2: The four modules of a modern NatCat model

2.2 ELT's, YLT's and YERT's

2.2.1 Event Loss Tables (ELT's)

After having built a NatCat model (e.g. for Japanese Earthquake), the next step is usually the construction of an *Event Loss Table (ELT)*. For this purpose, a list of *potential events* (in our example: earthquakes of different magnitude and location) is created. Each event is equipped with an event ID, the annual occurrence frequency or rate (typically a very small value), the expected loss per occurrence, the loss standard deviation per occurrence and the total exposure. These values are obtained from the NatCat model: The Hazard module provides information about frequency and local severity of each event in the ELT. For each insured object, the appropriate vulnerability curve is selected to compute the expected loss and the standard deviation. Together with the financial module, this enables an estimation of the total loss and standard deviation caused by each event in the ELT. The following Table 2.1 gives a concrete example of an ELT for Japanese Earthquake, where the numbers are just made up and do not stem from a Cat model (numbers are in millions).

Table 2.1: ELT example for Japanese Earthquake

Event ID	Description	Rate	Expected loss	Standard deviation	Exposure
1	Tōhoku EQ 9	0.0015	1'500	400	4'500
2	Osaka EQ 6	0.02	30	20	60
3	Hokkaidō EQ 7.7	0.0075	800	300	2'200
4	Kumamoto EQ 7	0.01	250	100	650
5	Kuril Island EQ 8.3	0.0065	20	15	80

2.2.2 Year Loss Tables (YLT's)

The events from the ELT can now be used to simulate losses of different years. To do so, one chooses the number n of simulated years (typically

$n = 100'000$) and for each year, events from the ELT are simulated. The number of occurrences of an event is drawn from a Poisson distribution with the rate in the ELT as parameter. The expected loss and the standard deviation are used to simulate the loss via a scaled beta distribution.

The output of such a simulation is a table with simulated years where we have for each year simulated events plus the losses. We call such a table a *Year Loss Table (YLT)*. Table 2.2 shows an example of a YLT constructed out of the ELT in Table 2.1. Note that since frequencies of events are typically very low, we expect to see only very few events in a YLT. For illustratory purposes, the following YLT contains many events.

Table 2.2: YLT example for Japanese Earthquake with 10 simulated years

Simulated Year	Event ID	Loss (in mio)
1	2	45
1	4	180
3	1	1'700
6	5	25
6	2	28
6	3	650
10	4	375

2.2.3 Year Event Rank Tables (YERT's)

In the next Chapters, we will also encounter *Year Event Rank Tables*, which are obtained as follows: We start again with an ELT and simulate events for each year, but this time, we do *not* assign losses to the events. Instead, each event obtains a randomly assigned *rank*. The ranks represent an ordering of the events, i.e. lower ranks indicate that the corresponding losses are higher. YERT's will be used to produce synthetic YLT's in the alternative modelling process for company A. For more details, we refer to Chapter 5. Table 2.3 shows a YERT which was obtained from Table 2.1.

2.3 OEP and AEP curves

Other outputs of classical NatCat models are *Exceedance Probability Curves*. These curves represent the probability that a (single or accumulated) loss exceeds a certain threshold. More precisely, we have

Definition 2.1. Let $X_1, X_2, \ldots X_N$ denote the NatCat loss amounts of one year. Then the *OEP (Occurrence Exceedance Probability)* represents the

Table 2.3: YERT example for Japanese Earthquake with 10 simulated years

Simulated Year	Number of Events	Ranks of Events
1	2	5
		4
3	1	1
6	3	7
		6
		2
10	1	3

probability of getting any single event within (typically) one year with a particular loss size or greater. Formally,

$$\mathrm{OEP}(y) = \mathbb{P}[\max(X_1, X_2, \ldots X_N) \geq y].$$

Analogously, the *AEP (Aggregate Exceedance Probability)* represents the probability of getting total annual NatCat losses of a particular level or greater. Formally,

$$\mathrm{AEP}(y) = \mathbb{P}\left[\sum_{i=1}^{N} X_i \geq y\right].$$

In practice, these curves are obtained out of Year Loss Tables. Suppose we have a YLT modelling n years and denote by X^i the biggest modelled loss of year i, for $1 \leq i \leq n$. Then the empirical distribution function of the maximal loss is

$$\widehat{F}_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X^i \leq y\}}.$$

Moreover, we can approximate

$$\mathrm{OEP}(y) \approx 1 - \widehat{F}_n(y)$$

and an analogous procedure approximates the AEP. If we possess YLT's for all peril regions and reinsured companies, an all-peril OEP can be computed in a similar way. In that case, we just take for each year the maximal loss of all involved YLT's.

Chapter 3

Frequency-Severity Models

We want to continue with an introduction to frequency-severity models. These models including their calibration to loss data can be discussed very detailedly. As we consider them here only as a tool in a bigger modelling process, we content ourselves with a brief introduction and a discussion of the most important properties, where we omit some of the proofs. The calibration to loss data will not be covered in this Chapter, since we want to focus on the theoretical aspects. We follow Wüthrich [5], pages $25 - 75$ and Arbenz [6], pages $38 - 39$, which also provide more details for the interested reader.

A frequency-severity model writes the *total claim amount* S of an insurance portfolio as

$$S = \sum_{i=1}^{N} X_i,$$

where N is a random variable modelling the *number of claims occurred*, i.e. the *frequency* and the random variables X_i model the *sizes* of the losses, i.e. the *severities*. First of all, we discuss some often used frequency and severity distributions and then introduce the frequency-severity model formally.

3.1 Important probability distributions in (re-) insurance

3.1.1 Frequency distributions

To model the frequency of (re-)insurance losses, one usually uses one of the following three discrete probability distributions:

Definition 3.1. A random variable $N \in \mathbb{N}_0$ has a *Binomial distribution* with *volume* $v \in \mathbb{N}$ and *success parameter* $p \in (0, 1)$, if

$$\mathbb{P}[N = k] = \binom{v}{k} p^k (1-p)^{v-k} \qquad \text{for all } k \in \{0, 1, 2, \ldots, v\}.$$

The binomial model can be useful when we have v independent policies and each policy can either have one claim with probability p or zero claims (with probability $(1-p)$), which is a reasonable model for a death cover for v individuals of the same age. Much more important for our purposes is the following distribution

Definition 3.2. A random variable $N \in \mathbb{N}_0$ has a *Poisson distribution* with parameter $c > 0$, if

$$\mathbb{P}[N = k] = e^{-c} \cdot \frac{c^k}{k!} \qquad \text{for all } k \in \mathbb{N}_0.$$

Proposition 3.3. *Let X and Y be independent Poisson random variables with parameters λ_x and λ_y respectively. Then $X+Y$ is a Poisson random variable with parameter $\lambda_x + \lambda_y$.*

Proof. See Föllmer et al [2], page 31. □

Remark 3.4. This property of the Poisson distribution induces the Aggregation Property (Theorem 3.22) of the Compound Poisson model presented below.

Proposition 3.5 (Expectation and Variance of Poisson random variables)**.** *Let $X \sim$ Poisson(λ). Then*

$$\mathbb{E}[X] = Var(X) = \lambda.$$

Proof. We have

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!} \\ &= e^{-\lambda} \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 &= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} - \lambda^2 \\ &= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} - \lambda^2 = e^{-\lambda} \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!} - \lambda^2 \\ &= e^{-\lambda} \lambda \left(\underbrace{\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!}}_{\lambda e^{\lambda}} + \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{e^{\lambda}} \right) - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

□

Remark 3.6. Sometimes, one writes $c = \lambda \cdot v$, with $\lambda > 0$, and $v > 0$, where v represents the *volume* (e.g. the number of policies) and λ measures the *frequency*.

Remark 3.7. Note that if N has a Binomial distribution, then $\mathbb{E}[N] = vp$ and $\operatorname{Var}(N) = vp(1-p)$, i.e. $\operatorname{Var}(N) < \mathbb{E}[N]$. On the other hand, if N has a Poisson distribution with parameter λ, then $\operatorname{Var}(N) = \mathbb{E}[N] = \lambda$.

Definition 3.8. A random variable $N \in \mathbb{N}_0$ has a *Negative Binomial distribution* with parameters $r > 0$, and $\beta > 0$, if

$$\mathbb{P}[N = k] = \binom{k+r-1}{k} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^k \qquad \text{for all } k \in \mathbb{N}_0.$$

Proposition 3.9. *Let X and Y be independent random variables with $X \sim$ NegBin(r_1, β) and $Y \sim$ NegBin(r_2, β). Then $X + Y \sim$ NegBin$(r_1 + r_2, \beta)$.*

Proof. The moment generating function of a $\operatorname{NegBin}(r, \beta)$ random variable can be computed as

$$M(t) = \left(\frac{1}{1+\beta(1-e^t)}\right)^r,$$

see Wüthrich [5], Proposition 2.21, where the different parametrizations are linked via

$$\gamma = r, \quad p = \frac{\beta}{1+\beta}.$$

Therefore (due to independence of X and Y),

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = M(t) = \left(\frac{1}{1+\beta(1-e^t)}\right)^{r_1+r_2},$$

i.e. $X + Y \sim \operatorname{NegBin}(r_1 + r_2, \beta)$. □

Proposition 3.10. *Suppose $X \sim$NegBin(r, β). Then $\mathbb{E}[X] = r\beta$ and Var$(X) = r\beta(1+\beta)$.*

Proof. By equation (1.3) of Wüthrich [5], page 17, we can compute

$$\frac{dM}{dt}(t) = \frac{dM}{dt}(1+\beta(1-e^t))^{-r} = r\beta(1+\beta(1-e^t))^{-r-1}e^t,$$

Hence,

$$\mathbb{E}[X] = \left.\frac{dM}{dt}\right|_{t=0} = r\beta.$$

An analogous computation can be performed for the second moment to establish the result for the variance. □

At a first glance, there is no connection between the Poisson and the Negative Binomial distribution. However, the Negative Binomial distribution can be seen as a generalization of the Poissen distribution, which is a very useful property from an actuarial point of view.

Definition 3.11. Suppose $\Lambda \sim H$, s.t. $H(0) = 0$ (i.e. $\Lambda > 0$ $\mathbb{P}$- a.s.), with $\mathbb{E}[\Lambda] = \lambda$ and $\mathrm{Var}(\Lambda) > 0$ (i.e. Λ is non- deterministic). Assume moreover that conditionally, given Λ,

$$N \sim \mathrm{Poisson}(\Lambda v)$$

for some fixed volume $v > 0$. Then N has a *Mixed Poisson distribution*.

Mixed Poisson models are very suitable if we want to model frequencies in different accounting years when there is an underlying *risk driver* that changes over time: We can model the number of claims in every year with a Poisson distribution, where the parameter changes from year to year, which reflects the changes in the risk profile. Because of this property, such distributions are for example used when the number of Hurricanes is modelled. Since Mixed Poisson models introduce uncertainity in the parameter, one might expect that they generate higher variances. Indeed:

Proposition 3.12. *Suppose N has a Mixed Poisson distribution, then*

$$Var(N) > \mathbb{E}[N].$$

Proof. By the tower property (Wüthrich [5], page 20), we get

$$\mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N|\Lambda]] = \mathbb{E}[\Lambda v] = \lambda v.$$

Moreover, we get for the variance (again by the tower property)

$$\mathrm{Var}(N) = \mathbb{E}[\mathrm{Var}(N|\Lambda)] + \mathrm{Var}(\mathbb{E}[N|\Lambda]) = \underbrace{v\mathbb{E}[\Lambda]}_{=\lambda v} + \underbrace{v^2\mathrm{Var}(\Lambda)}_{>0} > \lambda v.$$

□

Remark 3.13. Together with Remark 3.7, we find for all three cases $\mathrm{Var}(N) < \mathbb{E}[N]$, $\mathrm{Var}(N) = \mathbb{E}[N]$ and $\mathrm{Var}(N) > \mathbb{E}[N]$ a corresponding frequency distribution. In practice, one can for instance compare the first two empirical moments to choose one of the three presented frequencies. A question that arises in this context is how Λ should be chosen. We answer this question together with the above mentioned connection between Poisson and Negative Binomial models.

Definition 3.14. A positive, absolutely continuous (with respect to the Lebesgue measure) random variable X is said to have a *Gamma distribution* with *shape parameter* $\gamma > 0$ and *scale parameter* $c > 0$, if it has a density of the form

$$f_X(x) = \frac{c^\gamma}{\Gamma(\gamma)} x^{\gamma-1} e^{-cx}, \qquad \text{for all } x > 0.$$

Theorem 3.15. *Let $\Theta \sim Gamma(\gamma, \gamma)$ for $\gamma > 0$. Assume moreover that conditionally given Θ,*

$$N\big|_{\Theta} \sim Poisson(\Theta\lambda v)$$

for a frequency λ and a volume v. Then N has a Negative Binomial distribution with parameters

$$r = \gamma, \qquad \beta = \frac{\lambda v}{\gamma}.$$

Proof. By the tower property, we get

$$\begin{aligned}
\mathbb{P}[N = k] &= \mathbb{E}[\mathbb{P}[N = k|\Theta]] = \mathbb{E}\left[e^{-\Theta\lambda v}\frac{(\Theta\lambda v)^k}{k!}\right] \\
&= \int_0^\infty e^{-x\lambda v}\frac{(x\lambda v)^k}{k!}\frac{\gamma^\gamma}{\Gamma(\gamma)}x^{\gamma-1}e^{-\gamma x}\,\mathrm{d}x \\
&= \frac{(\lambda v)^k\gamma^\gamma}{\Gamma(\gamma)k!}\frac{\Gamma(\gamma+k)}{(\gamma+\lambda v)^{\gamma+k}}\underbrace{\int_0^\infty \frac{(\gamma+\lambda v)^{\gamma+k}}{\Gamma(\lambda+k)}x^{\gamma+k-1}e^{-(\gamma+\lambda v)x}\,\mathrm{d}x}_{=1} \\
&= \frac{\Gamma(\gamma+k)}{\Gamma(\gamma)k!}\left(\frac{\gamma}{\gamma+\lambda v}\right)^\gamma\left(\frac{\lambda v}{\gamma+\lambda v}\right)^k \\
&= \binom{k+r-1}{k}\left(\frac{1}{1+\beta}\right)^r\left(\frac{\beta}{1+\beta}\right)^k.
\end{aligned}$$

□

3.1.2 Severity distributions

Beside the number of claims, the size or severity of a specific claim is also of interest. There are various possible choices for severity distributions. However, we will focus our attention on the *Pareto distribution*, since this is the most relevant model for severities in a reinsurance context.

Definition 3.16. A random variable X is said to have a *Pareto distribution* with *threshold* $x_0 > 0$ and *tail parameter* $\alpha > 0$, if X has a density of the form

$$f_X(x) = \frac{\alpha}{x}\left(\frac{x_0}{x}\right)^\alpha \quad \text{for all } x \geq x_0.$$

Remark 3.17. The $k-$th moment of the Pareto distribution exists if and only if $k < \alpha$. Since the Pareto distribution has a relatively slowly decaying tail, it is called *heavy-tailed*.

3.2 The frequency-severity model

After we have introduced frequencies and severities of (re-)insurance losses, we can combine them to model the *total claim amount* S of a (re-)insurance company. We focus our attention on the following model, which imposes additional independence conditions. In practice, these conditions are often (approximately) satisfied and they equip our model with good properties in terms of computability.

Definition 3.18. Let N be a random variable taking values in $\mathbb{N}_0$ and let $Y_1, Y_2, \ldots \overset{\text{iid}}{\sim} G$ with $G(0) = 0$ (i.e. $Y_i > 0$ $\mathbb{P}$- a.s. $\forall i$). Assume moreover that N and $(Y_1, Y_2, Y_3, \ldots)$ are independent. Then we call

$$S = \sum_{i=1}^{N} Y_i \tag{3.1}$$

a *frequency-severity model* for the *total claim amount* S.

Remark 3.19. If we consider for instance a frequency-severity model S with $N \sim \text{Poisson}(\lambda)$ and $Y_i \overset{\text{iid}}{\sim} \text{Pareto}(x_0, \alpha)$, then we call S a *Poisson-Pareto model.* Alternatively, one calls the distribution of S coming from a frequency-severity model a *Compound distribution* and we write in this example $S \sim \text{CompPoi}(\lambda, G)$, where G is the cdf of Y_i.

Remark 3.20. The assumption $G(0) = 0$ in Definition 3.18 is crucial since it guarantees that all claims are positive with probability 1.

Let us have a look at Figure 3.1 which visualizes cumulative distribution functions of Poisson(λ)-Pareto(α, x_0) models with different parameters λ and α, where we set $x_0 = 1$. In the plot on the left-hand side, we see for integer values on the x-axis a bulge in the graph. This is due to the relatively high tail parameter of the Pareto distribution, which means that most of the severities are close to 1 and the probability for much higher claims decays substantially. A total claim amount of e.g. 1.9 can only be achieved if there is exactly one claim (as $x_0 = 1$) and since this claim is likely to be close to one, the cdf is flat for values which are slightly below 2. A total claim amount of 2 or larger can be achieved by one or two claims, which results in a big incline of the cdf for values slightly bigger than 2. If we compare the two graphs, we see that the tail parameter α has a bigger effect on the cdf than the Poisson parameter λ (in terms of increase rate) and the bulges become less prominent for smaller tail parameters, since in this case the severities are more likely to attain large values. In total, we see that the shape of compound distributions heavily depends on the choice of the parameters.

The next result investigates expectation and variance of frequency-severity models.

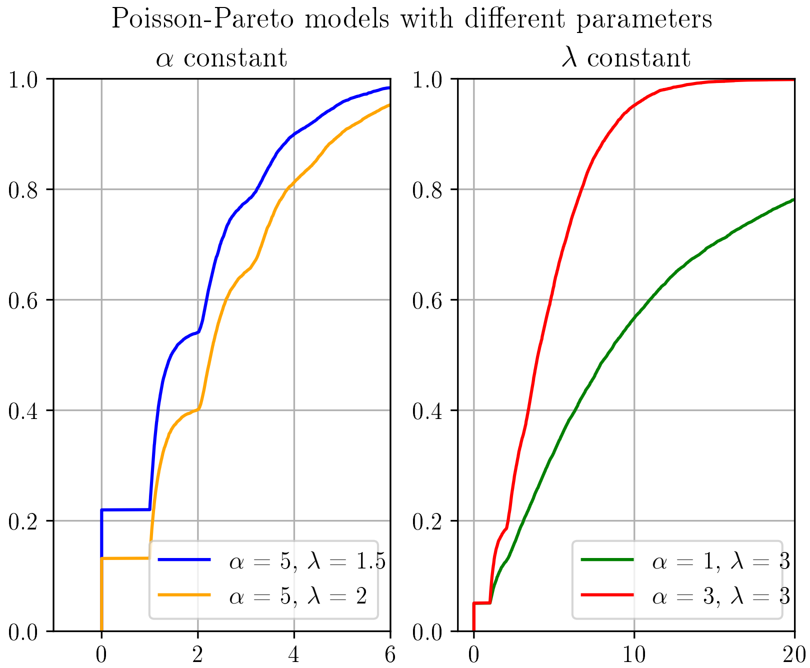


Figure 3.1: Poisson-Pareto models with different parameters

Theorem 3.21 (Wald)**.** *Let S be a frequency-severity model such that $Var(N) < \infty$ and $Var(Y_1) < \infty$. Then we have*

1. $\mathbb{E}[S] = \mathbb{E}[N] \cdot \mathbb{E}[Y_1]$.

2. $Var(S) = Var(Y_1) \cdot \mathbb{E}[N] + Var(N) \cdot \mathbb{E}[Y_1]^2$.

Proof. For the first part, note that

$$\begin{aligned}\mathbb{E}[S] &= \mathbb{E}\left[\sum_{i=1}^{N} Y_i\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N} Y_i \middle| N\right]\right] = \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{E}[Y_i|N]\right] \\ &= \mathbb{E}[N \cdot \mathbb{E}[Y_1]] = \mathbb{E}[N] \cdot \mathbb{E}[Y_1].\end{aligned}$$

In order to prove the second statement, we apply the tower property for variances:

$$\begin{aligned}\operatorname{Var}(S) &= \operatorname{Var}\left(\sum_{i=1}^{N} Y_i\right) = \operatorname{Var}\left(\mathbb{E}\left[\sum_{i=1}^{N} Y_i \middle| N\right]\right) + \mathbb{E}\left[\operatorname{Var}\left(\sum_{i=1}^{N} Y_i \middle| N\right)\right] \\ &= \operatorname{Var}\left(\sum_{i=1}^{N} \mathbb{E}[Y_i|N]\right) + \mathbb{E}\left[\sum_{i=1}^{N} \operatorname{Var}(Y_i|N)\right] \\ &= \operatorname{Var}\left(\sum_{i=1}^{N} \mathbb{E}[Y_i]\right) + \mathbb{E}\left[\sum_{i=1}^{N} \operatorname{Var}(Y_i)\right] \\ &= \operatorname{Var}(N) \cdot \mathbb{E}[Y_1]^2 + \mathbb{E}[N] \cdot \operatorname{Var}(Y_1).\end{aligned}$$

□

As we have already seen in Proposition 3.3, the sum of Poisson random variables is again a Poisson random variable. In the next step we show that this useful property translates to Compound Poisson distributions, which is essentially the reason why Compound Poisson models play an important role in loss modelling and in this thesis.

Theorem 3.22 (Aggregation property of CompPoi distributions)**.** *Assume $S_1, \dots, S_n$ are independent with $S_j \sim CompPoi(\lambda_j, G_j)$ for all $1 \le j \le n$. The aggregated claim has a Compound Poisson distribution*

$$S = \sum_{j=1}^{n} S_j \sim CompPoi(\lambda, G),$$

where

$$\lambda = \sum_{j=1}^{n} \lambda_j, \qquad G(x) = \sum_{j=1}^{n} \frac{\lambda_j}{\lambda} G_j(x).$$

Proof. For a detailed proof using moment generating functions, see Wüthrich [5], Theorem 2.12. □

Corollary 3.23. *Let S_1 and S_2 be two independent Compound Poisson models as in Theorem 3.22. Let*

$$U \sim Bernoulli\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$$

be independent of X_i and Y_j for all i, j. Then

$$S \overset{(d)}{=} \sum_{i=1}^{N} X_i^s,$$

with

$$N \sim Poisson(\lambda_1 + \lambda_2), \qquad X_i^s \overset{iid}{\sim} U \cdot X_1 + (1 - U) \cdot Y_1.$$

Proof. It suffices to show that X_i^s has cdf G (as in Theorem 3.22). Indeed, using independence of U, X_i and Y_j yields

$$\begin{aligned}\mathbb{P}[X_1^s \leq x] &= \mathbb{P}[U \cdot X_1 + (1-U) \cdot Y_1 \leq x] \\ &= \mathbb{P}[U=1] \cdot \mathbb{P}[X_1^s \leq x | U=1] + \mathbb{P}[U=0] \cdot \mathbb{P}[X_1^s \leq x | U=0] \\ &= \frac{\lambda_1}{\lambda_1+\lambda_2} \cdot G_X(x) + \frac{\lambda_2}{\lambda_1+\lambda_2} \cdot G_Y(x) \\ &= G(x),\end{aligned}$$

where G_X and G_Y denote the cdf's of X and Y respectively. □

This last result is very useful when we want to use the aggregation property for simulations, as in the next example:

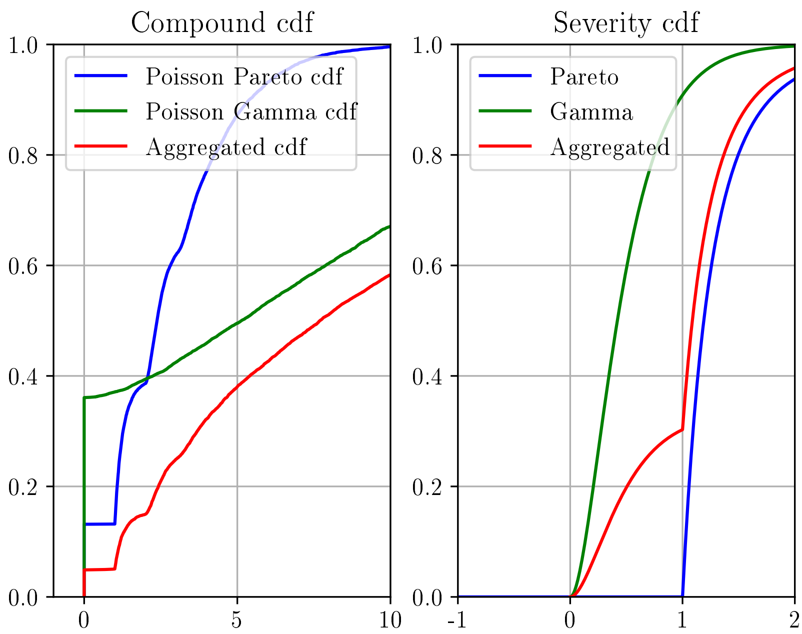


Figure 3.2: Visualization of the Aggregation property

In Figure 3.2, the graph on the left-hand side shows the cdf of a Poisson($\lambda = 2$)-Pareto($\alpha = 4, x_0 = 1$) model S_1 and the cdf of a Poisson($\lambda = 1$)-Gamma($\gamma = 2, c = 1/4$) model S_2. The red graph shows the cdf of the aggregated model S according to Theorem 3.22. Note that the cdf of the aggregated model is always smaller than the cdf's of the individual models. It is also interesting to have a look at the intercept of the y-axis. Note that $S = 0$ if and only if $S_1 = 0$ and $S_2 = 0$. Therefore, $\mathbb{P}[S = 0] = \mathbb{P}[S_1 = 0] \cdot \mathbb{P}[S_2 = 0]$ due to independence. This equality can be seen in the graph: The y-axis intercept of the red graph is the product of the y-axis intercepts of the green and blue graphs.

The graph on the right-hand side shows the severity cdf's of the three models. Note that according to Theorem 3.22, the severity cdf of the aggregated model is a convex combination of the individual severity cdf's.

3.3 Reinsurance layers and their properties

In a reinsurance context, frequency-severity models are usually not used as presented in the previous section; It is a common feature of contracts that losses are only absorbed above a *deductible* D up to a *cover* C. These two parameters must be taken into account when the total loss amount S of a reinsurance company is modelled. A very useful tool to implement D and C are the so called *Layer functions*.

Definition 3.24. A *Layer function* is a map

$$L_{D,C}\colon \mathbb{R} \to \mathbb{R}, \quad t \mapsto \min(C, \max(0, t - D)),$$

with *deductible* $D \geq 0$ and *cover* $C > 0$. In a *Per Risk Excess-of-Loss (Per Risk XL)* reinsurance contract, the layer function is applied to each and every risk, i.e. the total claim amount S is given by

$$S = \sum_{i=1}^{N} L_{D,C}(X_i).$$

Remark 3.25. If the X_i's model losses of *events* and not the losses of single risks, then a contract where a layer is applied to each of the X_i's is called a *Per Event Excess-of-Loss* reinsurance contract.

Let us investigate the behaviour of such Per Risk XL models. First of all, we discuss a useful result about the computation of moments:

Theorem 3.26 (Darth Vader Rule)**.** *Let X be a real valued random variable which is non-negative almost surely. Assume that the n-th moment of X exists, then*

$$\mathbb{E}[X^n] = n \int_0^\infty x^{n-1}(1 - F_X(x))\,\mathrm{d}x,$$

where F_X denotes the cumulative distribution function of X.

Proof. A proof using Lebesgue integration theory can be found in Muldowney et al [8]. □

Lemma 3.27. *Let X be a non-negative real valued random variable and let $D \geq 0$. Then*

$$\mathbb{E}[\max(0, X - D)] = \mathbb{E}[(X - D)^+] = \mathbb{E}[X \cdot \mathbb{1}_{\{X>D\}}] - D \cdot \mathbb{P}[X > D].$$

Proof. We have by the Darth Vader rule

$$\mathbb{E}[(X-D)^+] = \int_D^\infty (x-D)\,\mathrm{d}F_X(x) = \mathbb{E}[X \cdot \mathbb{1}_{\{X>D\}}] - D \cdot \mathbb{P}[X > D],$$

as desired. □

Proposition 3.28. *Let $L_{D,C}$ be a layer function and let X be a non-negative real valued random variable such that the n-th moment of $L_{D,C}(X)$ exists. Then we have*

$$\mathbb{E}[L_{D,C}(X)^n] = n\int_D^{D+C} (x-D)^{n-1}(1-F_X(x))\,\mathrm{d}x.$$

Proof. The integral in the statement reminds us of the Darth Vader rule, which we want to apply here. In order to do so, we first of all find the cdf of $L_{D,C}(X)$. We have

- $F_{L_{D,C}(X)}(t) = 0$ for all $t < 0$,
- $F_{L_{D,C}(X)}(0) = F_X(D)$,
- $F_{L_{D,C}(X)}(t) = F_X(t+D)$, for $0 < t < C$,
- $F_{L_{D,C}(X)}(C) = 1$.

Now, the Darth Vader rule gives us

$$\begin{aligned}\mathbb{E}[L_{D,C}(X)^n] &= n\int_0^\infty x^{n-1}(1-F_{L_{D,C}(X)})\,\mathrm{d}x \\ &= n\int_0^C x^{n-1}(1-F_X(x+D))\,\mathrm{d}x \\ &= n\int_D^{D+C} (x-D)^{n-1}(1-F_X(x))\,\mathrm{d}x.\end{aligned}$$

□

Example 3.29. Figure 3.3 visualizes the impacts of a Layer function on a cdf. The upper graph shows the cdf of a Pareto($\alpha = 1.5, x_0 = 1$) random variable X (green graph). Moreover, the deductible $D = 1.25$ and the cover $C = 1.75$ are also marked. When we apply the layer $L_{D,C}$ to X, the cdf is first of all shifted to the left by D, since the layer reduces each loss by D. Moreover, all losses above $D + C$ are trimmed, since loss amounts above the cover are not paid by the reinsurer. To emphasize the relation between the two cdf's further, we have a look at the red dotted lines in both graphs which are a segment of the respective cdf's. They are essentially the same curve, but the segment in the lower graph is shifted to the left by D.

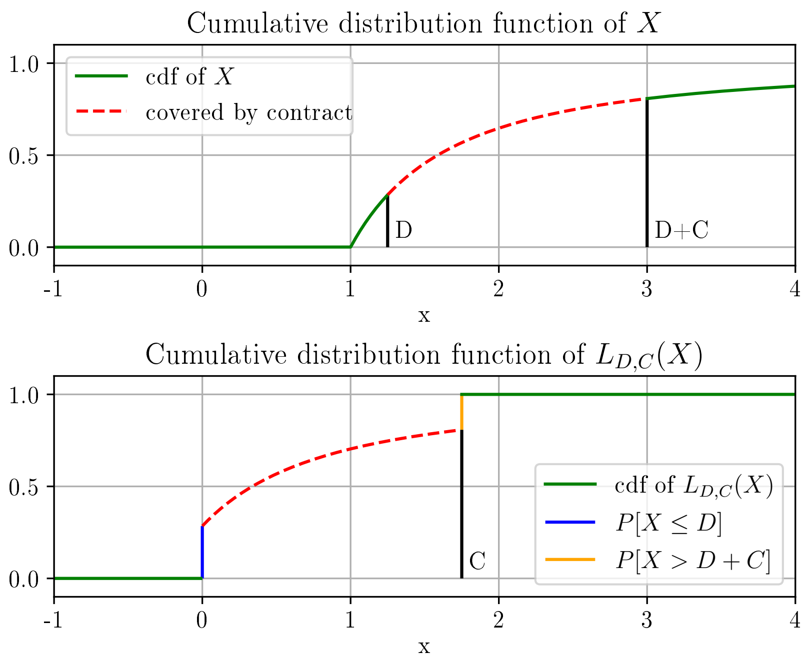


Figure 3.3: Visualization of the effects of a Layer function

The length of the blue segment in the lower graph is equal to the probability that a loss falls below the deductible. Similarly, the length of the orange segment is equal to the probability that the loss exceeds $D + C$, which means that the claim is only partially covered by the reinsurer, so the red segment in the lower graph visualizes the range which is covered by the reinsurer.

We finally want to visualize the impact of a Per Risk Excess of Loss layer on a frequency severity model.

Example 3.30. Figure 3.4 visualizes the effect of a Per Risk XL layer applied to a Poisson(2)-Pareto($\alpha = 1.5, x_0 = 2$) model. The different graphs show the cdf's of the total claim amount where we apply a layer $L_{D,C}$ to each and every loss for different values D and C. Let us discuss a few extreme cases: A layer with $D = 0$ and $C = \infty$ (red graph) has no impact on the losses and shows the total claim amount as if no layer is applied. In the case $D = 0$, $C = 1$, we have $D + C = 1 < x_0 = 2$, i.e. every loss is cut and therefore, the resulting cdf is a step function. If we finally look at the case $D = 5$, $C = 7$, we immediately notice the big y-axis intercept. This comes from the relatively high deductible D, i.e. many losses fall below it and are therefore counted as zero losses in the layer model.

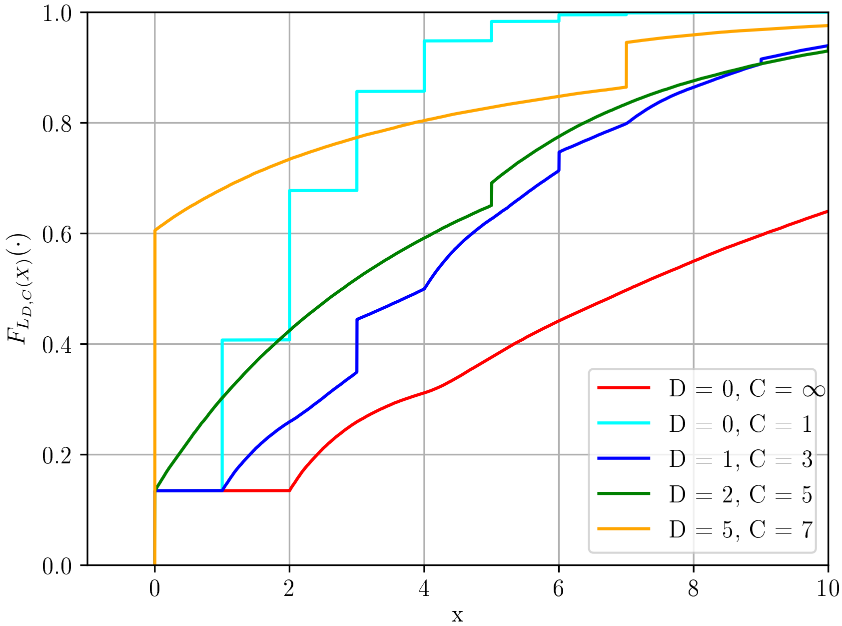


Figure 3.4: Different Layers applied to a frequency-severity model.

Chapter 4

Approximation of frequency-severity models with Poisson(2)-models

4.1 Outline

Let us have a quick look back at the alternative modelling process described in the introduction and the topics discussed so far. We have seen how classical NatCat models work and we also introduced the main concepts of frequency-severity models. The next step in the alternative modelling process as described in the introduction is the approximation of frequency-severity models by those with a Poisson(2) frequency. Let us recall why we do this step in the alternative modelling process:

As we have seen in Chapter 2, classical NatCat models produce Year Loss Tables where the number of events is drawn from a Poisson distribution. This is a justification why we also rely on this frequency in the alternative modelling approach. Moreover, Natural Catastrophes usually occur with a rather small frequency, i.e. if we choose a Poisson parameter of 2 (or lower), we can ensure that the frequency is able to model all losses (i.e. we do not "lose" events). It is very useful if we always work with the same parameter, as it makes the results more comparable and provides a general modelling framework which is applicable for different perils. In case there is a peril that occurs with a higher rate, it is also possible to select higher Poisson parameters, but for most relevant cases, a choice of $\lambda = 2$ is suitable.

In this Chapter, we focues on the *Frequency Transformation Error* and the *Severity Transformation Error*. However, we will not analyze them separately, but measure the quality of our approximations with expectation and variance.

The goal of this Chapter is to establish and justify the following

Transformation 4.1. Let $S = \sum_{i=1}^{N} X_i$ be a frequency-severity model. Then we approximate S with the frequency-severity model $T = \sum_{j=1}^{M} Y_j$, with

$$M \sim \text{Poisson}(2), \quad Y_j \overset{\text{iid}}{\sim} H_X,$$

for

$$H_X : \mathbb{R} \to [0,1], \quad t \mapsto \max\left(0, \frac{2-\mathbb{E}[N]}{2} \cdot \mathbb{1}_{\{t\geq 0\}} + \frac{\mathbb{E}[N]}{2} \cdot F_X(t)\right). \quad (4.1)$$

We start with restrictive assumptions on N and find a first transformation. Afterwards, we will gradually generalize our model until we can finally justify the choice of H_X as cdf for our transformed severity.

4.2 The case $\mathbb{E}[N] = 2$

Let us begin our discussion with the simplest case where $\mathbb{E}[N] = 2$, i.e. our model S has already the desired expected frequency. If we choose $Y_j := X_i$, $M \sim \text{Poisson}(2)$ and $T := \sum_{j=1}^{M} Y_j$, then we get by Theorem 3.21

$$\mathbb{E}[S] = \underbrace{\mathbb{E}[N]}_{=2} \cdot \mathbb{E}[X_1] = \underbrace{\mathbb{E}[M]}_{=2} \cdot \underbrace{\mathbb{E}[Y_1]}_{=\mathbb{E}[X_1]} = \mathbb{E}[T]. \quad (4.2)$$

Since S and T have the same expectation, we can use this as motivation to approximate S with T:

Transformation 4.2. Let $S = \sum_{i=1}^{N} X_i$ be a frequency-severity model such that $\mathbb{E}[N] = 2$. Then we approximate S with the frequency-severity model $T = \sum_{j=1}^{M} Y_j$, where

$$M \sim \text{Poisson}(2), \quad F_Y = F_X.$$

Remark 4.3. We will see later (Remark 4.15) that we can justify Transformation 4.2 more rigorously. As a summary, note that we just exchanged the frequency distribution and left the severities untouched. Moreover, the cdf of Y_j is equal to H_X for this special case.

Proposition 4.4. *Let S and T be as in Transformation 4.2. Then*

1. $\mathbb{E}[S] = \mathbb{E}[T]$,
2. $Var(T) = Var(S) + \mathbb{E}[X_1]^2 \cdot (\mathbb{E}[N] - Var(N))$.

Proof. The first part was already established in equation (4.2). For the second part, we get with Theorem 3.21

$$\begin{aligned} \mathrm{Var}(T) &= \mathrm{Var}(Y_1) \cdot \underbrace{\mathbb{E}[M]}_{=2=\mathbb{E}[N]} + \underbrace{\mathrm{Var}(M)}_{=2=\mathbb{E}[N]} \cdot \mathbb{E}[Y_1]^2 \\ &\overset{X_1 \overset{(d)}{=} Y_1}{=} \mathrm{Var}(X_1) \cdot \mathbb{E}[N] + \mathbb{E}[N] \cdot \mathbb{E}[X_1]^2 \end{aligned} \tag{4.3}$$

and

$$\mathrm{Var}(S) = \mathrm{Var}(X_1) \cdot \mathbb{E}[N] + \mathrm{Var}(N) \cdot \mathbb{E}[X_1]^2. \tag{4.4}$$

Putting equations (4.3) and (4.4) together, we receive

$$\mathrm{Var}(T) = \mathrm{Var}(S) + \mathbb{E}[X_1]^2 \cdot (\mathbb{E}[N] - \mathrm{Var}(N)),$$

as desired. □

Corollary 4.5. *Let S and T be as in Transformation 4.2. Then:*

- *If N has a Binomial distribution, then Var$(S) <$ Var(T).*
- *If N has a Poisson distribution, then Var$(S) =$ Var(T).*
- *If N has a Negative Binomial distribution, then Var$(S) >$ Var(T).*

Proof. The result follows immediately from Proposition 4.4 and Remark 3.13 □

Remark 4.6. In case $N \sim \mathrm{Poisson}(2)$, S and T have the same expectation and variance. In fact, they both have the *same* distribution as Transformation 4.2 does not change S in that case.

Example 4.7 (Transformation of Negative Binomial models)**.** In order to illustrate the above transformation, let

$$N_\beta \sim \mathrm{NegBin}(r, \beta), \quad X_i \overset{\text{iid}}{\sim} \mathrm{Pareto}(\alpha = 3, x_0 = 2), \quad S_\beta = \sum_{i=1}^{N_\beta} X_i$$

for $r > 0$, $\beta > 0$ such that $\mathbb{E}[N_\beta] = r\beta = 2$ (i.e. we choose $\beta > 0$ freely and set $r = 2/\beta$). By Propositon 3.10, we get

$$\mathbb{E}[N] = r\beta = 2, \qquad \mathrm{Var}(N) = r\beta(1+\beta) = 2(1+\beta) = \mathbb{E}[N] \cdot (1+\beta)$$

and therefore

$$\mathrm{Var}(N) - \mathbb{E}[N] = 2\beta. \tag{4.5}$$

Moreover, Proposition 4.4 tells us

$$\begin{aligned}\mathrm{Var}(S) - \mathrm{Var}(T) &= \mathbb{E}[X_1]^2 \cdot (\mathrm{Var}(N) - \mathbb{E}[N]) \\ &= 12 \cdot (r\beta(1+\beta) - r\beta) = 12 \cdot r\beta^2 \\ &= 24 \cdot \beta.\end{aligned}$$

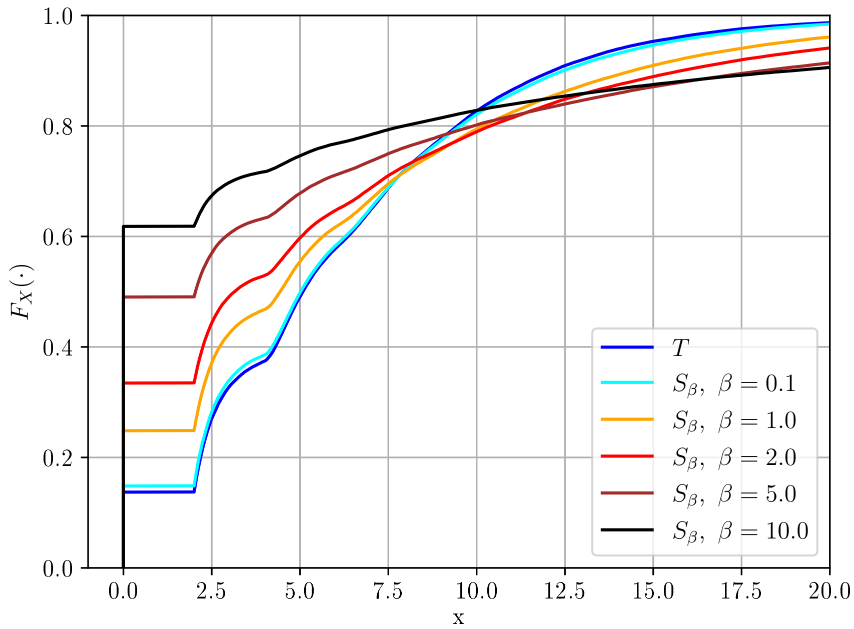


Figure 4.1: Cumulative distribution functions of the models S_β for some selected parameters β and their approximation T.

Since we kept the expectation constant for all our models N_β, the variance of N_β and therefore also the quality of our approximation (in terms of the variance) basically depends on the choice of β, as we have seen in the last computation. If β increases, then $\mathrm{Var}(S) - \mathrm{Var}(T)$ and the difference of the first two moments of N_β increase as well (see equation (4.5)). Since the first two moments of a Poisson random variable agree, one can argue (informally) that an increasing β brings N_β further away from a Poisson distribution and therefore we expect our approximation to be worse, which is reflected in the second part of Proposition 4.4.

Figure 4.1 visualizes the cdfs of S_β for some selected parameters as well as their transformation T. Note that all S_β are transformed to the same Poisson model, since they all have the same expected frequency and the same severity distribution. This is one of the crucial weaknesses of Transformation 4.2: Only the expected frequency of N and the severity distribution are taken into account in T. Unfortunately, we will not be able to correct this mistake in the remainder of this Chapter, since we explicitly want T to have a Poisson frequency. This can be regarded as the price for enforcing our approximation to have a Poisson(2) frequency .The figure suggests that bigger β lead to a worse approximation, which is in line with the result

$$\mathrm{Var}(S) - \mathrm{Var}(T) = 24 \cdot \beta.$$

4.3 The case $\mathbb{E}[N] \leq 2$ for Poisson models

After we have discussed a first approach to approximate compound models with expected frequency equal to 2, we will generalize Transformation 4.2 to the case $\mathbb{E}[N] \leq 2$. However, an extension of our approach is not straight forward. We first of all construct a transformation for Poisson models, since there is a natural way to extend our thoughts in that case. Afterwards, we will study the properties of the new approximation for arbitrary frequency distributions.

Let N have a Poisson distribution with $\mathbb{E}[N] \leq 2$. Note that the model S does not produce as many events (in expectation) as a Compound model with expected frequency 2. We can compensate this mismatch if we simply add to S some claims of size zero. Together with Theorem 3.22 we show that S can be written as a Compound model with Poisson(2) frequency:

Theorem 4.8. *Let $S = \sum_{i=1}^{N} X_i$ be a frequency-severity model with $N \sim$ Poisson such that $\mathbb{E}[N] \leq 2$. Then*

$$S \overset{(d)}{=} T,$$

where T is a frequency-severity model with Poisson(2) frequency and severity cdf

$$G_X : \mathbb{R} \to [0,1], \quad t \mapsto \frac{2 - \mathbb{E}[N]}{2} \cdot \mathbb{1}_{\{t \geq 0\}} + \frac{\mathbb{E}[N]}{2} \cdot F_X(t), \tag{4.6}$$

where F_X denotes the cdf of X_i.

Proof. If $\mathbb{E}[N] = 2$, then N and M have the same distribution. Moreover $G_X(t) = F_X(t)$ for all $t \in \mathbb{R}$. Altogether we get $S \overset{(\mathrm{d})}{=} T$, as desired. Otherwise, let

$$\widetilde{S} = \sum_{i=1}^{\widetilde{N}} \widetilde{X}_i,$$

with $\widetilde{N} \sim \text{Poisson}(2 - \mathbb{E}[N])$ and $\widetilde{X}_i = 0$ for all i such that $\widetilde{N}$ and all $\widetilde{X_i}$ are independent. By construction, $\widetilde{S} = 0$ $\mathbb{P}$-a.s. and therefore $S \overset{(\mathrm{d})}{=} S + \widetilde{S}$. Next, we apply Theorem 3.22 to see that $T := S + \widetilde{S}$ has a Compound

Poisson distribution with frequency parameter $\mathbb{E}[N] + (2 - \mathbb{E}[N]) = 2$ and severity cdf

$$G_X(t) = \frac{2 - \mathbb{E}[N]}{2} \cdot \mathbb{1}_{\{t \geq 0\}} + \frac{\mathbb{E}[N]}{2} \cdot F_X(t),$$

where F_X denotes the cdf of X_i. Altogether, choosing $M \sim \text{Poisson}(2)$ and $Y_i \overset{\text{iid}}{\sim} G_X$ yields $S \overset{\text{(d)}}{=} S + \widetilde{S} = T$, i.e. we can represent S exactly as a Compound Poisson model with frequency 2. ☐

Corollary 4.9. *Let S be as in Theorem 4.8. Let*

$$U \sim \textit{Bernoulli}\left(\frac{\mathbb{E}[N]}{2}\right)$$

be independent of N and X_i for all i. Then

$$S \overset{(d)}{=} \sum_{j=1}^{M} Y_j,$$

for $M \sim \textit{Poisson}(2)$ and $Y_j \overset{iid}{\sim} U \cdot X_1$ for all j.

Proof. The result follows from the construction in the proof of Theorem 4.8 and Corollary 3.23. ☐

Motivated by Theorem 4.8, we formulate the following

Transformation 4.10. Let $S = \sum_{i=1}^{N} X_i$ be a frequency-severity model where N has a Poisson distribution such that $\mathbb{E}[N] \leq 2$. Then we transform S to $T = \sum_{j=1}^{M} Y_j$, with

$$M \sim \text{Poisson}(2), \quad Y_j \overset{\text{iid}}{\sim} G_X,$$

for G_X is as in equation (4.6).

Clearly, $S \overset{\text{(d)}}{=} T$ by 4.8, so the word "transformation" is perhaps not optimal. The situation becomes however more interesting in the next section where we extend Transformation 4.10 to arbitrary frequency distributions, and the above transformation will then lead to an approximation rather than an exact representation.

In Figure 4.2, we see the visualization of $G_X(\cdot)$ for different parameter values $\lambda = \mathbb{E}[N]$ and X having a Pareto$(\alpha = 1.5, x_0 = 2)$ distribution. For $\lambda = 2$ we get the original cdf of X. Note that the y-axis intercept increases if λ decreases. A lower λ implies that we add more zero claims in the transformation process and therefore, the severity cdf of T must as well allow for more zero claims.

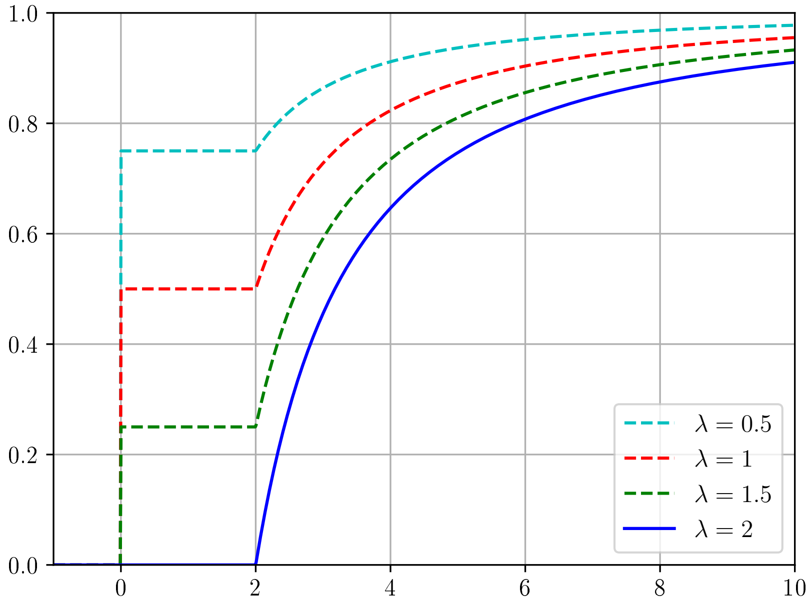


Figure 4.2: Visualization of $G_X(\cdot)$ for different parameters

Remark 4.11. As we have already pointed out, S and T are both frequency-severity models with $S \overset{(d)}{=} T$, i.e. $\mathbb{E}[S] = \mathbb{E}[T]$. Therefore

$$\mathbb{E}[N] \cdot \mathbb{E}[X_1] = \underbrace{\mathbb{E}[M]}_{=2} \cdot \mathbb{E}[Y_1].$$

Hence, if $Y \sim G_X$, then

$$\mathbb{E}[Y] = \frac{\mathbb{E}[N]}{2} \cdot \mathbb{E}[X].$$

We can generalize this connection to arbitrary moments (if they exist):

Lemma 4.12. *Let $Y \sim G_X$ and assume that the n-th moment of Y exists, then*

$$\mathbb{E}[Y^n] = \frac{\mathbb{E}[N]}{2} \cdot \mathbb{E}[X^n].$$

Proof. We apply Theorem 3.26 twice:

$$\begin{aligned}\mathbb{E}[Y^n] &= n \cdot \int_0^\infty t^{n-1}(1 - G_X(t))\,\mathrm{d}t \\ &= n \cdot \int_0^\infty t^{n-1}\left(1 - \frac{2-\mathbb{E}[N]}{2} - \frac{\mathbb{E}[N]}{2}\cdot F_X(t)\right)\mathrm{d}t \\ &= n \cdot \int_0^\infty t^{n-1}\left(\frac{\mathbb{E}[N]}{2} - \frac{\mathbb{E}[N]}{2}\cdot F_X(t)\right)\mathrm{d}t \\ &= \frac{\mathbb{E}[N]}{2}\cdot n \cdot \int_0^\infty t^{n-1}(1 - F_X(t))\,\mathrm{d}t \\ &= \frac{\mathbb{E}[N]}{2}\cdot \mathbb{E}[X^n].\end{aligned}$$

□

4.4 The case $\mathbb{E}[N] \leq 2$ for arbitrary frequencies

We now want to find similar results for other frequency distributions. For this purpose we simply apply Transformation 4.10 defined in the previous section to arbitrary frequency distributions and investigate the behaviour.

Transformation 4.13. Let $S = \sum_{i=1}^N X_i$ be a frequency-severity model such that $\mathbb{E}[N] \leq 2$. Then we approximate S with $T = \sum_{j=1}^M Y_j$, for

$$M \sim \text{Poisson}(2), \quad Y_j \overset{\text{iid}}{\sim} G_X,$$

where G_X is as in equation (4.6).

Theorem 4.14. *Let S and T be as in Transformation 4.13. Then*

1. $\mathbb{E}[S] = \mathbb{E}[T]$.
2. $\mathit{Var}(T) = \mathit{Var}(S) + \mathbb{E}[X]^2 \cdot (\mathbb{E}[N] - \mathit{Var}(N))$.

Proof. By Lemma 4.12,

$$\mathbb{E}[T] = \underbrace{\mathbb{E}[M]}_{=2} \cdot \mathbb{E}[Y] = 2 \cdot \frac{\mathbb{E}[N]}{2} \cdot \mathbb{E}[X] = \mathbb{E}[N] \cdot \mathbb{E}[X] = \mathbb{E}[S],$$

which proves the first part. By the Wald identities (Theorem 3.21), we get

$$\begin{aligned}\mathrm{Var}(T) &= \underbrace{\mathrm{Var}(Y_1)}_{=\mathbb{E}[Y_1^2]-\mathbb{E}[Y_1]^2} \cdot \underbrace{\mathbb{E}[M]}_{=2} + \underbrace{\mathrm{Var}(M)}_{=2} \cdot \mathbb{E}[Y_1]^2 \\ &= 2 \cdot E[Y_1^2] = \mathbb{E}[N] \cdot \mathbb{E}[X_1^2], \qquad (4.7)\end{aligned}$$

where we used again Lemma 4.12 in the last step. Again by Wald, we compute for S

$$\begin{aligned}
\mathrm{Var}(S) &= \mathrm{Var}(X_1) \cdot \mathbb{E}[N] + \mathrm{Var}(N) \cdot \mathbb{E}[X_1]^2 \\
&= (\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2) \cdot \mathbb{E}[N] + \mathrm{Var}(N) \cdot \mathbb{E}[X_1]^2 \\
&= \underbrace{\mathbb{E}[N] \cdot \mathbb{E}[X_1^2]}_{\mathrm{Var}(T)} + \mathbb{E}[X_1]^2 \cdot (\mathrm{Var}(N) - \mathbb{E}[N]). \qquad (4.8)
\end{aligned}$$

All in all, equations (4.7) and (4.8) yield

$$\mathrm{Var}(T) = \mathrm{Var}(S) + \mathbb{E}[X_1]^2 \cdot (\mathbb{E}[N] - \mathrm{Var}(N)), \qquad (4.9)$$

as desired. □

Remark 4.15. First of all, note that Theorem 4.14 gives us *exactly* the same statements for $\mathbb{E}[T]$ and $\mathrm{Var}(T)$ as Propositon 4.4, which tells us that Transformation 4.13 is a sensible generalization of Transformation 4.2. This is due to the fact that 4.2 is a special case of 4.13, which gives us the mathematical justification of the approach in the first section of this Chapter. Moreover, note that for a Poisson random variable N, we have $\mathbb{E}[N] = \mathrm{Var}(N)$, so equation (4.9) reads $\mathrm{Var}(T) = \mathrm{Var}(S)$ in that case, which is in line with the previously derived result $S \overset{(d)}{=} T$ (in other words, 4.10 is a special case of 4.13).

There is also a beautiful formula for the probability of a total claim amount of zero in the transformed model:

Proposition 4.16. *Let T be as in Transformation 4.13. Then*

$$\mathbb{P}[T = 0] = e^{-\mathbb{E}[N]}.$$

Proof. We have

$$\begin{aligned}
\mathbb{P}[T = 0] &= \sum_{k=0}^{\infty} \mathbb{P}[M = k] \cdot G_X(0)^k = \sum_{k=0}^{\infty} \frac{e^{-2} 2^k}{k!} \cdot \left(\frac{2 - \mathbb{E}[N]}{2} \right)^k \\
&= e^{-2} \sum_{k=0}^{\infty} \frac{(2 - \mathbb{E}[N])^k}{k!} = e^{-2} \cdot e^{2 - \mathbb{E}[N]} \\
&= e^{-\mathbb{E}[N]},
\end{aligned}$$

where we used $F_X(0) = 0$. □

We see in particular that $\mathbb{P}[T = 0]$ is independent of the underlying severity.

The following table summarizes the impacts of Theorem 4.14 on different commonly used frequency dsitributions.

Table 4.1: Transformation of different frequencies

Frquency N	$\mathbb{E}[N]$	$\mathrm{Var}(N)$	$\mathrm{Var}(T)$
$\mathrm{Binomial}(n, p)$	np	$np(1-p)$	$> \mathrm{Var}(S)$
$\mathrm{Poisson}(\lambda)$	λ	λ	$= \mathrm{Var}(S)$
$\mathrm{NegBin}(r, \beta)$	$r\beta$	$r\beta(1+\beta)$	$< \mathrm{Var}(S)$

Intuitively speaking, Theorem 4.14 tells us that the approximation becomes worse (in terms of the variance) if the difference between the expectation and the variance of the frequency is big. Note that this observation is in line with Corollary 4.5. Let us illustrate Theorem 4.14 with a few examples.

Example 4.17. In this example, we want to illustrate the impact of the chosen frequency distribution in Theorem 4.14. Let

$$S_1 = \sum_{i=1}^{N_1} X_i, \; S_2 = \sum_{j=1}^{N_2} Y_j, \; S_3 = \sum_{k=1}^{N_3} Z_k,$$

with $X_i, Y_j, Z_k \overset{\text{iid}}{\sim} \mathrm{Pareto}(\alpha = 3, x_0 = 1)$ and

- $N_1 \sim \mathrm{NegBin}(r = 1, \beta = 1{,}5)$,
- $N_2 \sim \mathrm{Binomial}(n = 2, p = 0.75)$,
- $N_3 \sim \mathrm{Poisson}(\lambda = 1.5)$.

In particular, we have $\mathbb{E}[N_1] = \mathbb{E}[N_2] = \mathbb{E}[N_3] = 1.5$ and all three models share the same severity distribution. Therefore, if we apply Transformation 4.13, S_1, S_2 and S_3 are all transformed to S_3. This can also be seen by looking at equation (4.6): the transformed severiy depends on the expected frequency and the severity distribution, which both coincide for all models in our example.

Let us have a look at Figure 4.3. The blue, orange and red graphs are the cdf's of S_1, S_2, S_3 respectively. By Theorem 4.14, all models have the same expectation and by equation (4.9), the variance is only driven by the original frequencies, since all models share the same severity. We see that the cdf of the Compound Binomial model is the fastest growing curve. This is due to the variance of the Binomial distribution, wich is smaller than the variances of the Negative Binomial and Poisson model (at least when all three distributions share the same expectation, as in our case). Similarly, the cdf of the Negative Binomial model is the slowest increasing curve, since the variance of the underlying frequency is the biggest of all three. As equation (4.9) suggests, the mismatch of the approximated model compared to the original one increases if the mismatch of variance and expectation of the original frequency increase.

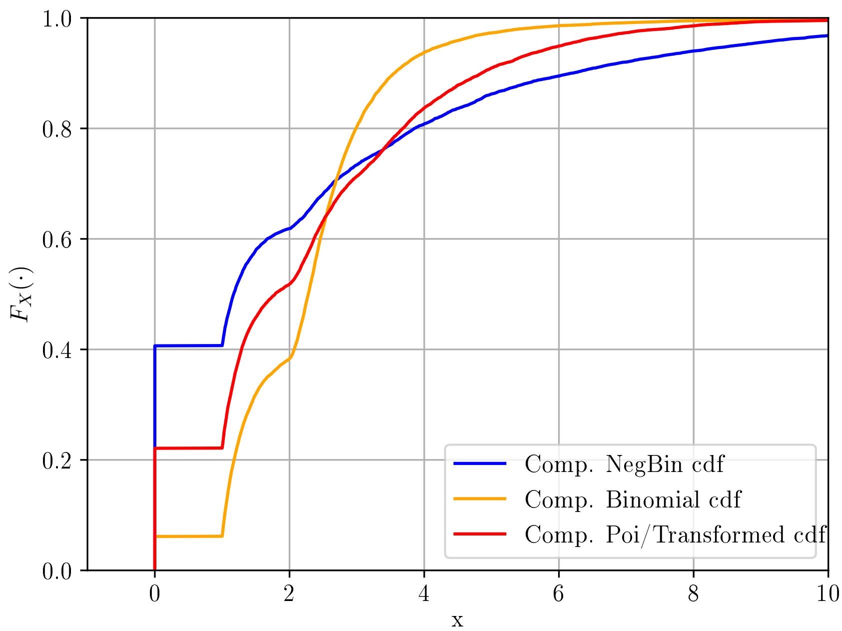


Figure 4.3: Transformation of compound cdf's with same severity and expected frequency

Example 4.18. We now focus on the behaviour of different severities under our transformation. Let

$$S_1 = \sum_{i=1}^{N_1} X_i, \quad S_2 = \sum_{j=1}^{N_2} Y_j,$$

where $N_1, N_2 \overset{\text{iid}}{\sim} \text{NegBin}(r = 1, \beta = 1.5)$ and

$$X_i \overset{\text{iid}}{\sim} \text{Pareto}(\alpha = 3, x_0 = 2), \qquad Y_j \overset{\text{iid}}{\sim} \text{Gamma}(\gamma = 3, c = 1).$$

Note that N_1 and N_2 have the same distribution. We made this choice to lay our focus solely on the behaviour of the severities under our transformation. Moreover, we have

$$\mathbb{E}[X_i] = \frac{\alpha x_0}{\alpha - 1} = 3, \quad \mathbb{E}[Y_j] = \frac{\gamma}{c} = 3$$

and thus

$$\mathbb{E}[S_1] = 4.5 = \mathbb{E}[S_2].$$

One can compute

$$\mathbb{E}[X_i^2] = \frac{x_0^2 \alpha}{\alpha - 1} = 6, \qquad \mathbb{E}[Y_j^2] = \frac{\gamma + \gamma^2}{c^2} = 12.$$

By Theorem 4.14, we have

$$\begin{aligned}\mathrm{Var}(S_1) - \mathrm{Var}(T_1) &= 6 \cdot 2.25 = 13.5,\\ \mathrm{Var}(S_2) - \mathrm{Var}(T_2) &= 12 \cdot 2.25 = 27.\end{aligned}$$

The last observation is very interesting. Even though the fundamental structure of the severity is preserved by Transformation 4.13 (up to scaling and translation), the shape of the original severity has a big impact on the quality of the approximation (in terms of the variance). Note moreover that S_1 and S_2 are not transformed into the same Poisson model (in contrast to the previous example). Figure 4.4 illustrates S_1 and S_2 as well as their transformations T_1 and T_2.

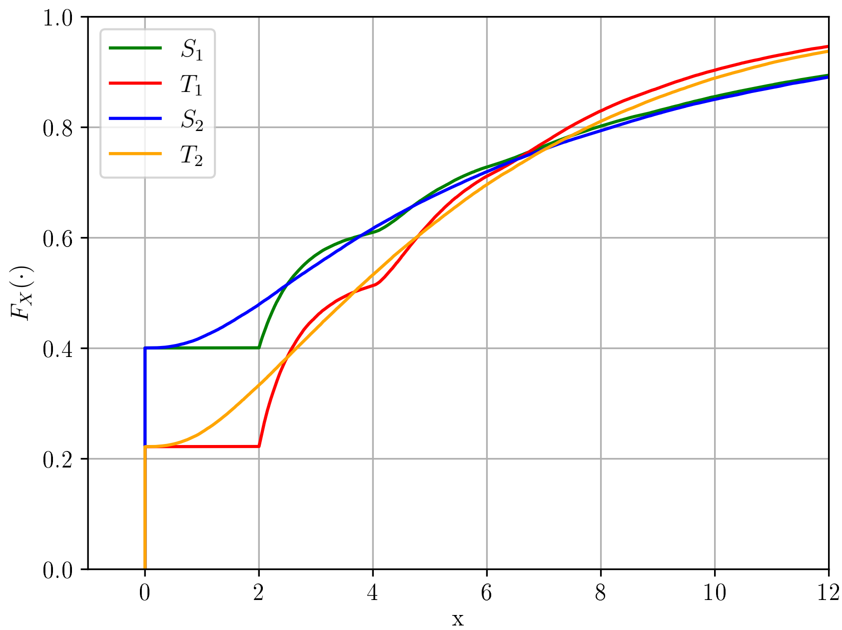


Figure 4.4: Transformation of compound cdf's with same frequency and different severities

We see that the cdf's related to the Pareto distribution have more bumps than the cdf's related to the Gamma distribution. This is due to the relatively high tail parameter α, which results in a concentration of the losses coming from the Pareto distribution around 2, see Figure 4.4. Note that

$$\mathbb{P}[S_1 = 0] = \mathbb{P}[S_2 = 0] = \mathbb{P}[N_1 = 0] = \left(\frac{1}{1+\beta}\right)^r \overset{\text{here}}{=} 0.4$$

and

$$\mathbb{P}[T_1 = 0] = \mathbb{P}[T_2 = 0] = e^{-r\beta} = e^{-1.5} \approx 0.223$$

by Proposition 4.16, so the cdf's of the original models S_1 and S_2 have a higher y-axis intercept, but then increase slower than T_1 and T_2 (since $\mathbb{E}[S_i] = \mathbb{E}[T_i]$). Let us elaborate our example a bit further and suppose that we wish to minimize the difference $\mathbb{P}[S_i = 0] - \mathbb{P}[T_i = 0]$: This value can be seen as a measure for the quality of our approximation. According to the above computations, we have

$$\mathbb{P}[S_i = 0] - \mathbb{P}[T_i = 0] = \left(\frac{1}{1+\beta}\right)^r - e^{-r\beta} =: f(\beta).$$

For $\beta \to 0$, this expression converges to zero (and also for $\beta \to \infty$, but this case is not relevant in practice). Figure 4.5 visualizes f (for $r = 1$).

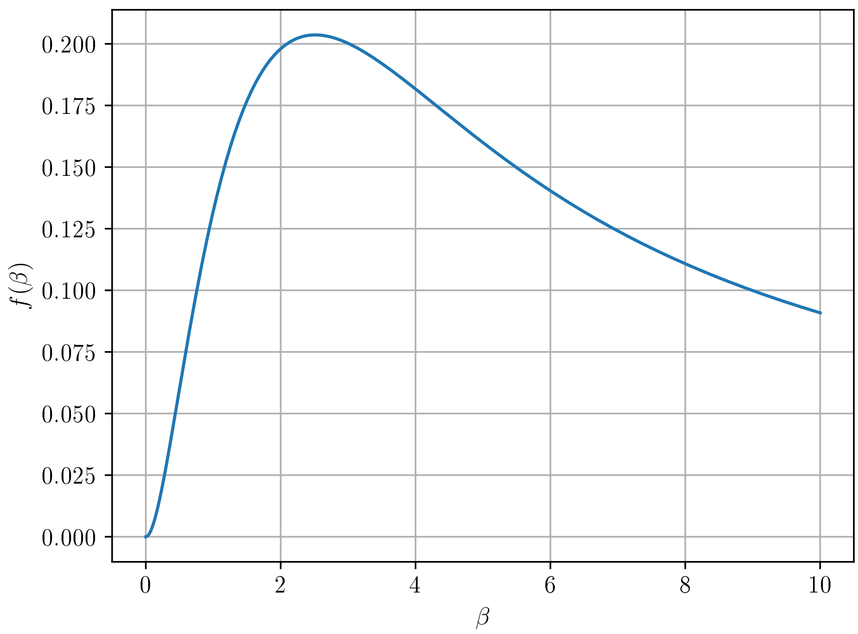


Figure 4.5: $f(\beta)$

However, let us recall that

$$\mathrm{Var}(N_i) = r\beta \cdot (1+\beta) = \mathbb{E}[N_i] \cdot (1+\beta),$$

and if we let β go to zero, then not only the y-axis intercepts of the cdf's of S_i and T_i converge (i.e. f goes to zero), but also

$$\mathrm{Var}(S_i) \to \mathrm{Var}(T_i), \quad \mathrm{Var}(N_i) \to \mathbb{E}[N_i].$$

In conclusion, we can say that minimizing the difference of the y-axis intercepts and minimizing the difference of the first two moments of N go hand in hand, so they describe (qualitatively) the same "goodness-of-fit" measure.

4.5 Generalization to arbitrary frequencies (including the case $\mathbb{E}[N] > 2$)

So far, we could always solve the problem that N underestimates (in expectation) the number of claims by adding some zero losses. However, this trick will no longer be applicable if we move on to the case where $\mathbb{E}[N] > 2$. We now face the problem that our initial frequency produces (in expectation) too many losses. Therefore, we must think about cutting certain events instead of adding zero claims. It will turn out that the problems which arise in this context can be solved smartly if we apply suitable Per Risk XL layers (the deductible will play the crucial role). As our models are mainly intended for a reinsurance context where layers are always applied, the problems which arise in this section can be circumvented in practice for most of the relevant cases if the deductible of the layer meets certain requirements.

4.5.1 Loss cutting

As we have already pointed out, we now must cut certain losses in order to attain an expected claims frequency of two. It lies at hand that the impact on the total loss amount S can be minimized if we cut the smallest losses. More concretely, if we have a frequency N such that $\mathbb{E}[N] > 2$, then we want to cut (in expectation) the $\mathbb{E}[N] - 2$ lowest losses, i.e. the lowest $\frac{\mathbb{E}[N]-2}{\mathbb{E}[N]}$ proportion of all losses modelled by N. Starting with the severity X of S, we wish to transform F_X such that it reflects this cutting. We suggest the following severity transformation:

Definition 4.19. For a severity cdf

$$F_X\colon \mathbb{R} \longrightarrow [0,1],$$

we define

$$\begin{aligned} H_X\colon \mathbb{R} &\longrightarrow [0,1] \\ t &\longmapsto \max\left(0, \frac{2-\mathbb{E}[N]}{2} \cdot \mathbb{1}_{\{t\geq 0\}} + \frac{\mathbb{E}[N]}{2} \cdot F_X(t)\right) \\ &= \max(0, G_X(t)), \end{aligned} \tag{4.10}$$

where N denotes the frequency of the underlying compound model.

Remark 4.20. Definition 4.19 can be seen as a generalization of G_X to any expected frequency, since $H_X = G_X$ if $\mathbb{E}[N] \leq 2$.

Before we discuss a new transformation for compound models, let us briefly check that H_X does indeed reflect the desired severity transformation. Note that taking the maximum in the definition of H_X is inevitable, since we now assume $\mathbb{E}[N] > 2$ and therefore $G_X(0) < 0$. It is straightforward that for any severity cdf F_X, the transformation H_X is normalized, increasing and continuous, i.e. H_X is a cdf.

We want that H_X reflects the cutting of the (in average) lowest $\mathbb{E}[N]-2$ losses, i.e. the lowest $\frac{\mathbb{E}[N]-2}{\mathbb{E}[N]}$ proportion of all losses should be cut. We can achieve that if we first look at F_X and its $\frac{\mathbb{E}[N]-2}{\mathbb{E}[N]}$ quantile t^*. If t^* is exactly the point where H_X stops to be zero and starts to increase, then H_X does indeed represent the desired cutting, since the lowest $\frac{\mathbb{E}[N]-2}{\mathbb{E}[N]}$ proportion of losses coming from F_X is no longer considered. More precisely, we have

Lemma 4.21. *Let S be a frequency-severity model with underlying continuous severity cdf F_X and expected claims frequency $\mathbb{E}[N] > 2$. Moreover, define*

$$t^* := q_F\left(\frac{\mathbb{E}[N]-2}{\mathbb{E}[N]}\right) \geq 0, \tag{4.11}$$

where q_F denotes the right-quantile of F_X. Then we have

$$H_X(t) = 0 \ \forall t \leq t^*, \qquad H_X(t) > 0 \ \forall t > t^*.$$

Proof.

$$\begin{aligned} G_X(t^*) &= \frac{2-\mathbb{E}[N]}{2} + \frac{\mathbb{E}[N]}{2} \cdot \underbrace{F_X\left(q_F\left(\frac{\mathbb{E}[N]-2}{\mathbb{E}[N]}\right)\right)}_{=\frac{\mathbb{E}[N]-2}{\mathbb{E}[N]}} \\ &= \frac{2-\mathbb{E}[N]}{2} + \frac{\mathbb{E}[N]-2}{2} = 0, \end{aligned}$$

so t^* is indeed the point where H_X starts to increase, since q_F denotes the right-quantile function. □

Corollary 4.22. *Let $t \geq t^*$, then $H_X(t) = G_X(t)$.*

Proof. By the Proof of Lemma 4.21, we have

$$G_X(t^*) = 0.$$

Since G_X is an increasing function, we get

$$H_X(t) = \max(0, G_X(t)) = G_X(t) \quad \text{for all } t \geq t^*.$$

□

Example 4.23. Lemma 4.21 is visualized in Figure 4.6, where we look at the cdf of a Pareto($\alpha = 1.5, x_0 = 2$) distribution and its transformation for different expected frequencies $\mathbb{E}[N]$. Note that depending on $\mathbb{E}[N]$, we get either a transform G_X or H_X. The dotted horizontal lines are on the levels

$$\frac{\mathbb{E}[N_3]-2}{\mathbb{E}[N_3]} = \frac{3-2}{3} = \frac{1}{3}, \quad \frac{\mathbb{E}[N_4]-2}{\mathbb{E}[N_4]} = \frac{4-2}{4} = \frac{1}{2}$$

respectively. We see that their F_X-quantiles agree with the points where the transformed severities start to increase, which justifies our approach graphically.

The black solid line is drawn at the modelling threshold $x_0 = 2$ of the Pareto severity. In the case where $\mathbb{E}[N] \le 2$, F_X is scaled and translated and for $x \ge x_0 = 2$ losses have positive probability. This discussion leads us to the following

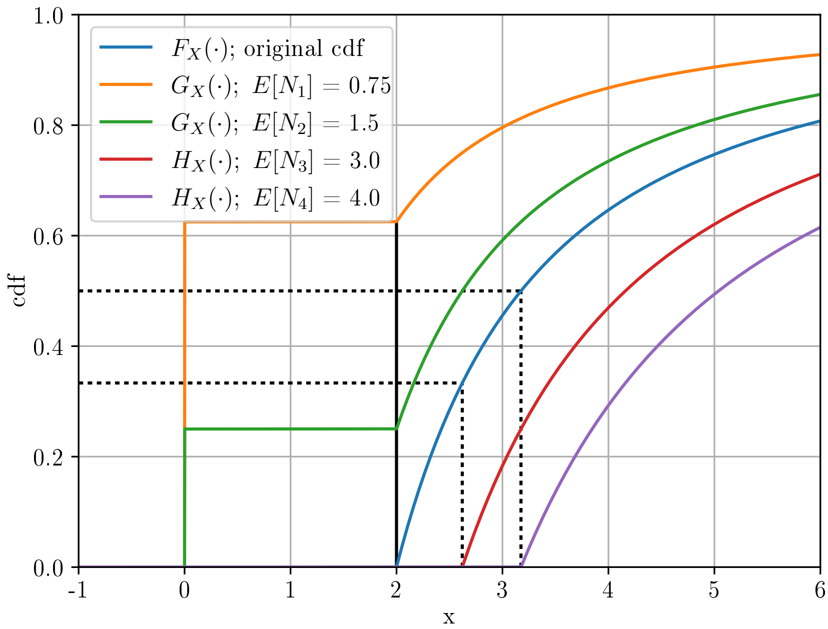


Figure 4.6: Transformation of a Pareto severity for different expected frequencies

Transformation 4.24. Let $S = \sum_{i=1}^{N} X_i$ be a frequency-severity model such that $\mathbb{E}[N] > 2$. Then we approximate S with $T = \sum_{j=1}^{M} Y_j$, where

$$M \sim \text{Poisson}(2), \quad Y_j \overset{\text{iid}}{\sim} H_X,$$

and H_X is as in equation (4.10).

As Figure 4.6 already suggests, the shape of H_X is quite different than the shape of G_X and thus it is not clear which properties we can expect from Transformation 4.24. This will be elaborated in the remainder of this section. Since T is a model where some losses have been cut, we expect that

T underestimates S, which can be shown mathematically in the next result:

Theorem 4.25. *Let S and T be as in Transformation 4.24. Then*

$$\mathbb{E}[T] \leq \mathbb{E}[S].$$

Proof. Let as before

$$t^* := q_F\left(\frac{\mathbb{E}[N]-2}{\mathbb{E}[N]}\right)$$

for the right quantile function q_F of F_X. Then we have by the Darth Vader rule

$$\begin{aligned}
\mathbb{E}[T] &= \mathbb{E}[M] \cdot \mathbb{E}[Y] = 2 \cdot \int_0^\infty (1 - H_X(t))\,\mathrm{d}t \\
&= 2 \cdot \left(\int_0^{t^*} 1\,\mathrm{d}t + \int_{t^*}^\infty (1 - G_X(t))\,\mathrm{d}t \right) \\
&= \mathbb{E}[N] \cdot \left(\int_0^{t^*} \frac{2}{\mathbb{E}[N]}\,\mathrm{d}t + \int_{t^*}^\infty (1 - F_X(t))\,\mathrm{d}t \right). \qquad (4.12)
\end{aligned}$$

Note that

$$1 - F_X(t^*) = \frac{\mathbb{E}[N]}{\mathbb{E}[N]} - \frac{\mathbb{E}[N]-2}{\mathbb{E}[N]} = \frac{2}{\mathbb{E}[N]}$$

and since F_X is increasing, $1 - F_X$ is decreasing. Therefore, we can estimate

$$\int_0^{t^*} \frac{2}{\mathbb{E}[N]}\,\mathrm{d}t \leq \int_0^{t^*} (1 - F_X(t))\,\mathrm{d}t \qquad (4.13)$$

and thus putting equations (4.12) and (4.13) together:

$$\mathbb{E}[T] \leq \mathbb{E}[N] \cdot \int_0^\infty (1 - F_X(t))\,\mathrm{d}t = \mathbb{E}[N] \cdot \mathbb{E}[X] = \mathbb{E}[S]$$

□

Remark 4.26. We get an equality in Theorem 4.25 if and only if F_X is constant in $[0, t^*]$. For the Pareto distribution, this can only happen when $t^* \leq x_0$, which implies that no losses are cut (or in other words, we would need $\frac{\mathbb{E}[N]-2}{\mathbb{E}[N]} = 0$, which contradicts our assumption in this section). Let us reformulate this finding in a

Corollary 4.27. *Let S and T be frequency-severity models as in Transformation 4.24, where we assume that S has a Pareto severity. Then*

$$\mathbb{E}[T] < \mathbb{E}[S].$$

Since Pareto distributions are very important in applications, we are not satisfied with this result; Our transformation is not even able to correctly estimate the expectation of the total claim amount S. This is the reason why we start to work with layers in the next section.

4.5.2 Application of Per Risk XL layers

Note that we usually construct frequency-severity models for the ground up losses X_i (or more precisely, we model e.g. in the Pareto case the losses above a modelling threshold $x_0 < D$), the layer is applied afterwards. If the deductible D is chosen such that in average all cut losses are below D, then the cutting process should have no impact on the transformation of the total loss amount in a Per Risk XL contract. Let us first of all show that Lemma 4.12 can be generalized for layers.

Lemma 4.28. *Let $Y \sim H_X$ and assume that the n-th moment of Y exists. Moreover, suppose that $D \geq t^*$, for t^* as in equation* (4.11). *Then*

$$\mathbb{E}[L_{D,C}(Y)^n] = \frac{\mathbb{E}[N]}{2} \cdot \mathbb{E}[L_{D,C}(X)^n].$$

Proof. By Proposition 3.28, we have

$$\mathbb{E}[L_{D,C}(X)^n] = n \int_D^{D+C} (x-D)^{n-1}(1-F_X(x))\,\mathrm{d}x$$

and

$$\begin{aligned}\mathbb{E}[L_{D,C}(Y)^n] &= n \int_D^{D+C} (x-D)^{n-1}(1-H_X(x))\,\mathrm{d}x \\ &\overset{D\geq t^*}{\equiv} n \int_D^{D+C} (x-D)^{n-1}(1-G_X(x))\,\mathrm{d}x \\ &= n \int_D^{D+C} \frac{\mathbb{E}[N]}{2} \cdot (x-D)^{n-1}(1-F_X(x))\,\mathrm{d}x \\ &= \frac{\mathbb{E}[N]}{2} \cdot \mathbb{E}[L_{D,C}(X)^n],\end{aligned}$$

where we used Corollary 4.22 in the second line. □

Interestingly, Theorem 4.14 does also have an equivalent result if we work with layers:

Theorem 4.29. *Let S and T be as in Transformation 4.24. Moreover, let $L_{D,C}$ be a layer function applied to each and every loss such that $D \geq t^*$, where t^* is as in* (4.11). *Then we have*

- $\mathbb{E}[T] = \mathbb{E}[S]$.
- $Var(T) = Var(S) + \mathbb{E}[L_{D,C}(X)]^2 \cdot (\mathbb{E}[N] - Var(N))$.

Proof. The proof works in the exact same way as the proof of Theorem 4.14, where we just use Lemma 4.28 instead of Lemma 4.12. □

Theorem 4.29 shows us, that our previously derived results can be generalized to the case where $\mathbb{E}[N] > 2$, under application of a suitable layer to each and every loss. In summary, we formulate

Transformation 4.30. Let $S = \sum_{i=1}^{N} L_{D,C}(X_i)$ be a frequency-severity model such that $\mathbb{E}[N] > 2$ and $D \geq t^*$, where t^* is defined as in equation (4.11). Then we approximate S with $T = \sum_{j=1}^{M} L_{D,C}(Y_j)$, where

$$M \sim \text{Poisson}(2), \quad Y_j \overset{\text{iid}}{\sim} H_X,$$

and H_X is as in equation (4.10).

and Theorem 4.29 ensures that $\mathbb{E}[S] = \mathbb{E}[T]$.

Example 4.31. Let us apply Transformation 4.30 to a concrete compound distribution. Suppose

$$S = \sum_{i=1}^{N} L_{D,C}(X_i),$$

where $X_i \overset{\text{iid}}{\sim} \text{LogNormal}(\mu = 0,\ \sigma = 3)$, $N \sim \text{NegBin}(r = 0.5,\ \beta = 7)$ and $D = 1,\ C = 3$. Then we have $\mathbb{E}[N] = 3.5 > 2$, and

$$t^* = q_F\left(\frac{\mathbb{E}[N] - 2}{\mathbb{E}[N]}\right) \approx 0.58 < D,$$

so all conditions of Transformation 4.30 are satisfied. Figure 4.7 shows on the left-hand side the frequency and severity of S, where the subpicture on the bottom left shows the cdf of X and $L_{D,C}(X)$ as well as t^* on the x-axis and $\frac{\mathbb{E}[N]-2}{\mathbb{E}[N]}$ on the y-axis.

The right-hand side shows the transformed frequency and severity. Finally, Figure 4.8 shows the cdf's of S and T with and without the layer applied to each and every loss. It stands out immediately that our models S and T have substantially different probabilities to produce a total claim amount of zero. The main reason for this phenomenon lies in the frequencies of the two models. As the upper subgraphs of Figure 4.7 suggest, the Negative Binomial frequency has a much higher probability to produce zero events than the Poisson frequency, which is in line with the bigger y-axis intercept of the cdf of S in Figure 4.8. Additionally, a look at Figure 4.7 clarifies that X lies below D with probability ≈ 0.5, while the layer has a smaller impact on the transformed severity. This is also a reason why S and T have different probabilities to produce a total loss amount of zero. Moreover, Figure 4.8 hints as well that the probability to produce a total claim amount of zero is bigger when we apply the layer. This is very reasonable as the total claim amount can be zero if there are eiter no losses, or if all losses are below the deductible in the model with the layer.

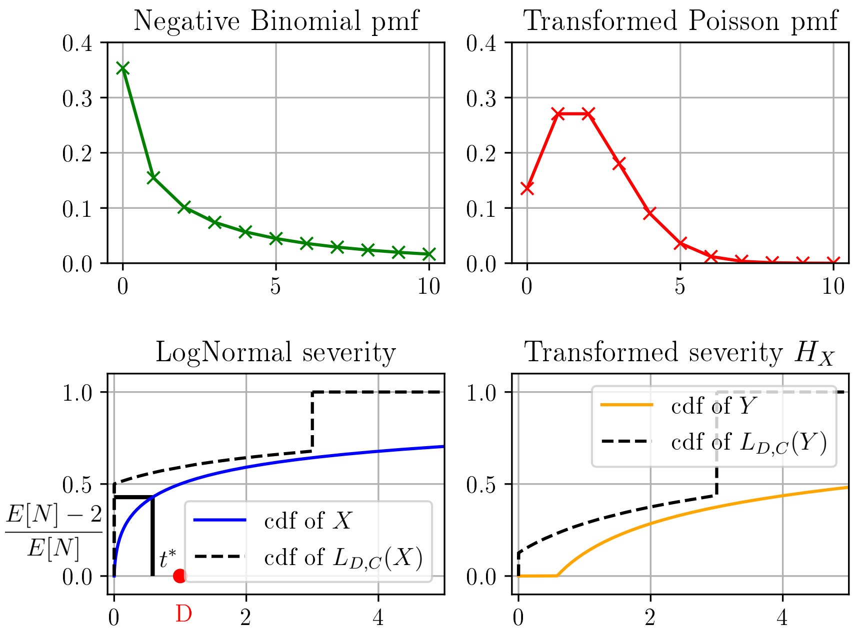


Figure 4.7: Behaviour of a Negative Binomial frequency and a LogNormal severity under Transformation 4.30

Let us check that our example verifies the statements of Theorem 4.25 and 4.29. In a simulation, empirical distribution functions of S and T had been generated based on $1'000'000$ samples with and without the application of layers. When no layers had been applied, the simulation yielded

$$\widehat{\mathbb{E}[S]} \approx 313.45, \quad \widehat{\mathbb{E}[T]} \approx 309.95,$$

so there is still a relative gap of $\approx 1.1\%$ and the expectation of T is smaller than the expectation of S, which reproduces Theorem 4.25. On the other hand, we get the following simulated values when we apply layers:

$$\widehat{\mathbb{E}[S]} \approx 4.092, \quad \widehat{\mathbb{E}[T]} \approx 4.090,$$

which results in a relative gap of only $\approx 0.049\%$. These findings are in line with Theorem 4.29, as the two expectations converge in the case when we apply a suitable layer (where "suitable" means $D \geq t^*$). Additionally, it is also worth mentioning that the layer influences the rate of growth of S and T significantly: On the one hand, this shows that the application of Per Risk XL layers leads to a relatively low coverage by the reinsurer and on the other hand, it points out that our severity has quite heavy tails, i.e. there

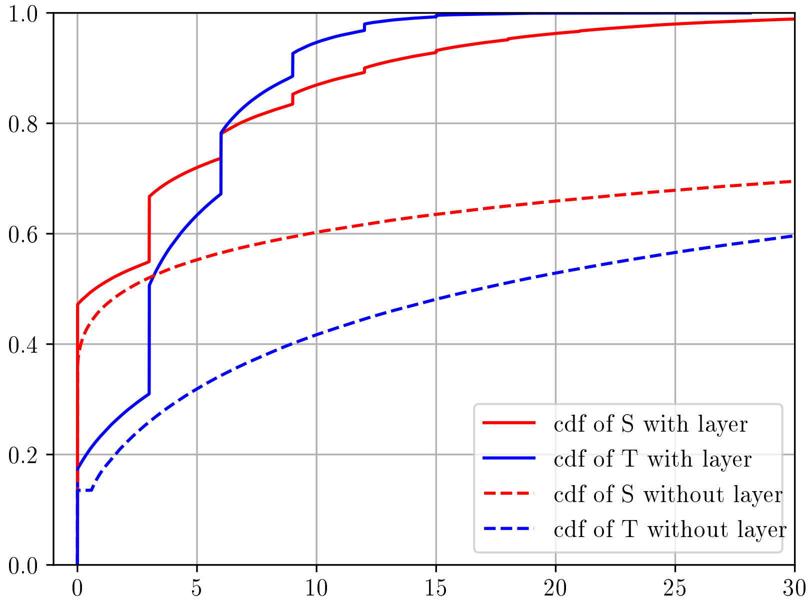


Figure 4.8: Cumulative Distribution Functions of S and$T$.

are many (and large) losses above $D + C$ which are only partially covered by the reinsurance contract.

Chapter 5

Transformation of Actuarial Loss Models into Synthetic NatCat Loss Tables

5.1 Explanation of the Transformation

In this section, we follow SCOR [9], pages 3-6.

So far, we are able to calibrate a frequency-severity model to past loss data of company A and we have seen how this model is transformed into a Poisson(2)-model. However, it is still unclear how we can procede in order to obtain a synthetic NatCat loss table compatible with the results of classical NatCat models. Let us briefly explain the procedure:

First of all, a YERT is generated for the considered peril region (e.g. Swiss floods as in the introduction). For each of the n modelled years (typically, one chooses $n = 100'000$), the number of events is drawn from a Poisson(2) distribution. As most NatCat vendor models work with Poisson frequencies, it lies at hand that we also work with this distribution type for a better compatibility. A parameter value of 2 is reasonable as most observed expected frequencies of natural catastrophes are below 2 (i.e. the YERT is able to capture all losses with high probability). The event ID's of the YERT are chosen in such a way that there is no overlap with the event ID's from used NatCat models.

Next, we model a YLT for n years with the respective Poisson(2) model for company A. The modelled losses are then ordered by rank (this sorting is independent of the year in which the different loss severities are drawn) and injected into the YERT in order to obtain a synthetic YLT, i.e. the highest modelled loss is assigned to the loss event with rank 1 in the YERT and so on. There may be two cases where we have a mismatch of modelled losses:

- The number of modelled losses in the YERT is bigger than the number of modelled losses in the YLT: In that case, the YLT is complemented with zero-loss events (see right-hand side of Figure 5.1).

- The number of modelled losses in the YERT is lower than the number of modelled losses in the YLT: In that case, we just take the highest losses of the YLT into account and discard the redundant losses (see left-hand side of Figure 5.1).

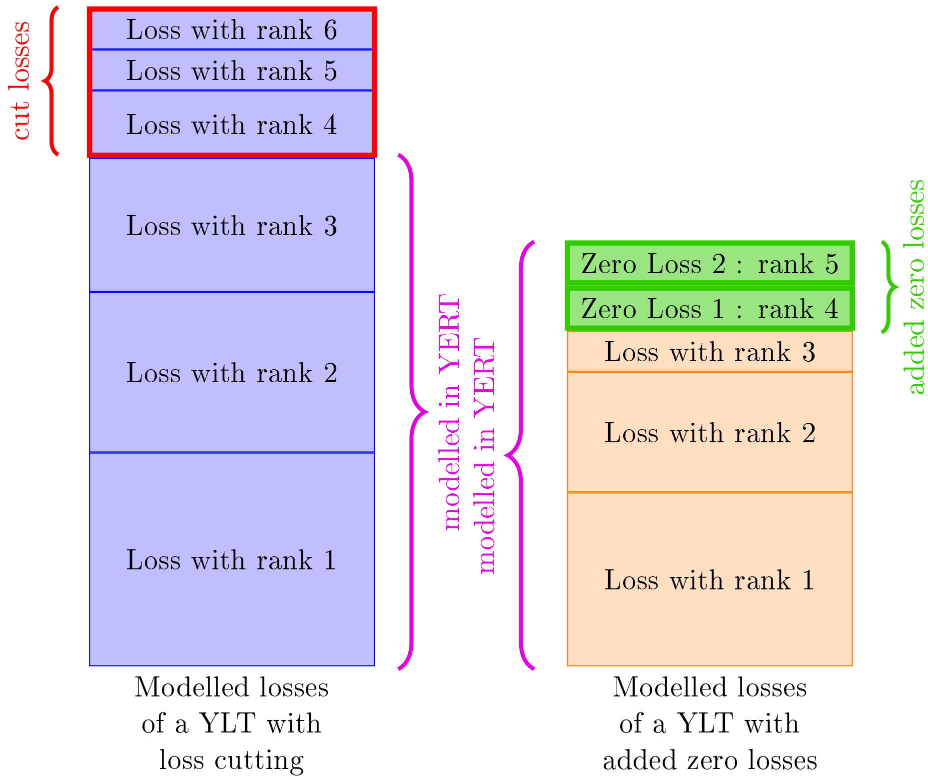


Figure 5.1: Mismatch of modelled years

It is also possible to apply this transformation to the YLT coming from the NatCat model for the losses of companies B and C in order to receive a synthetic YLT. This is only done if one wishes to obtain an aggregated YLT for all three companies, but often the YLT's from classical NatCat models are not transformed with this procedure. It is important to point out that we use *exactly* the same YERT for all YLT's that we want to transform. This is the reason why we call it *pre-generated*: we first simulate the YERT and then use it for all YLT's we want to transform. The reason for this approach lies in the compatibility: If we use the same pre-generated YERT for all YLT's, we are able to add the losses of the synthetic YLT's belonging to the same modelled loss event. This is not possible if we use different YERT's. This procedure ensures moreover correlation of big loss events from the different actuarial models. It is a very reasonable assumption that if one insurer observes big claims, then the other insurers are confronted with big losses as well, since a NatCat event usually affects the portfolios of several insurers. Hence, it is very sensible that we correlate big loss events by using the same YERT for all models which we transform.

Example 5.1. We illustrate the above explained procedure with a concrete example. Consider two loss models that we want to transform into synthetic YLT's for $n = 10$ modelled years. Figure 5.2 shows again the concept of the transformation. The "Loss Model" is in this context usually a Poisson(2) frequency-severity model as for company A.

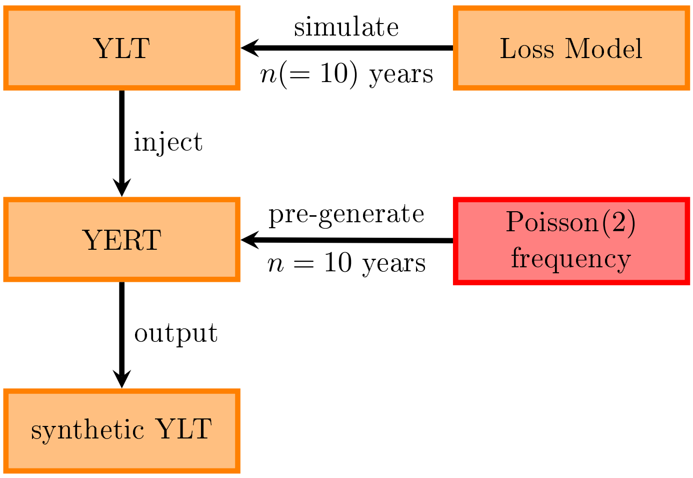


Figure 5.2: Transformation overview

Figure 5.3 shows the transformation process for loss model 1, where the pre-generated YERT models one loss more that the YLT, i.e. we need to add one zero loss (green).

Similarly, Figure 5.4 shows the transformation for loss model 2, where the YLT models one loss more than the pre-generated YERT, so we need to cut the lowest loss (red). It is important to mention that we used the *same* YERT in both examples, which ensures compatibility and correlation. Finally, Figure 5.5 shows how the two synthetic YLT's can be aggregated, if a wholistic model is desired. To do so, we just sum the losses of the two synthetic YLT's that correspond to the same rank.

YLT 1

Year	Event ID	Loss	Rank
1	5	185	3
2	1	90	5
2	9	630	1
5	1	110	4
8	3	540	2

Loss model 1

simulate

$n = 10$ years

inject

Pre-generated YERT for $n = 10$ years

Simulated Year	Common Event ID	Simulated Rank
2	16	5
3	14	1
5	21	4
5	14	6
7	19	3
9	16	2

Poisson(2) frequency

pre-generate

$n = 10$ years

output

synthetic YLT 1

Simulated Year	Common Event ID	Simulated Rank	Loss
2	16	5	90
3	14	1	630
5	21	4	110
5	14	6	0
7	19	3	185
9	16	2	540

Figure 5.3: YLT transformation with one added zero loss (green)

YLT 2

Simulated Year	Event ID	Loss (in mio)	Rank
3	2	230	4
4	1	105	6
8	3	360	2
8	4	720	1
8	5	245	3
8	2	60	7
10	1	150	5

Loss model 2

simulate

$n = 10$ years

inject

Pre-generated YERT for $n = 10$ years

Simulated Year	Common Event ID	Simulated Rank
2	16	5
3	14	1
5	21	4
5	14	6
7	19	3
9	16	2

Poisson(2) frequency

pre-generate

$n = 10$ years

output

synthetic YLT 2

Simulated Year	Common Event ID	Simulated Rank	Loss
2	16	5	150
3	14	1	720
5	21	4	230
5	14	6	105
7	19	3	245
9	16	2	360

Figure 5.4: YLT transformation with one cut loss (red)

synthetic YLT 1

Simulated Year	Common Event ID	Simulated Rank	Loss
2	16	5	90
3	14	1	630
5	21	4	110
5	14	6	0
7	19	3	185
9	16	2	540

synthetic YLT 2

Simulated Year	Common Event ID	Simulated Rank	Loss
2	16	5	150
3	14	1	720
5	21	4	230
5	14	6	105
7	19	3	245
9	16	2	360

Aggregate

aggregated synthetic YLT

Simulated Year	Common Event ID	Simulated Rank	Loss
2	16	5	90+150=240
3	14	1	630+720=1350
5	21	4	110+230=340
5	14	6	0+105= 105
7	19	3	185+245= 430
9	16	2	540+360=900

Figure 5.5: Aggregation of synthetic YLT's

5.2 The Loss Cutting Error

Let us discuss a very important issue that arises in our discussion of YLT transformations. If we procede as in Section 5.1 (assume we are working with only one YLT), then the case where the YERT models fewer losses than the YLT has always positive probability. Depending on the underlying frequencies, this probability can converge to zero if n increases, but there are also configurations where it is independent of n; If e.g. the YLT and the YERT work with exactly the same frequency distribution, then we always cut losses in the transformation process with probability 0.5. Why is this problematic? According to our algorithm, we essentially just copy the losses from the YLT to the YERT and there is no possibility to add other losses. If we always discard some of these losses with positive probability, then the total loss amount (for one year or the whole modelling period) will always be underestimated by the synthetic YLT.

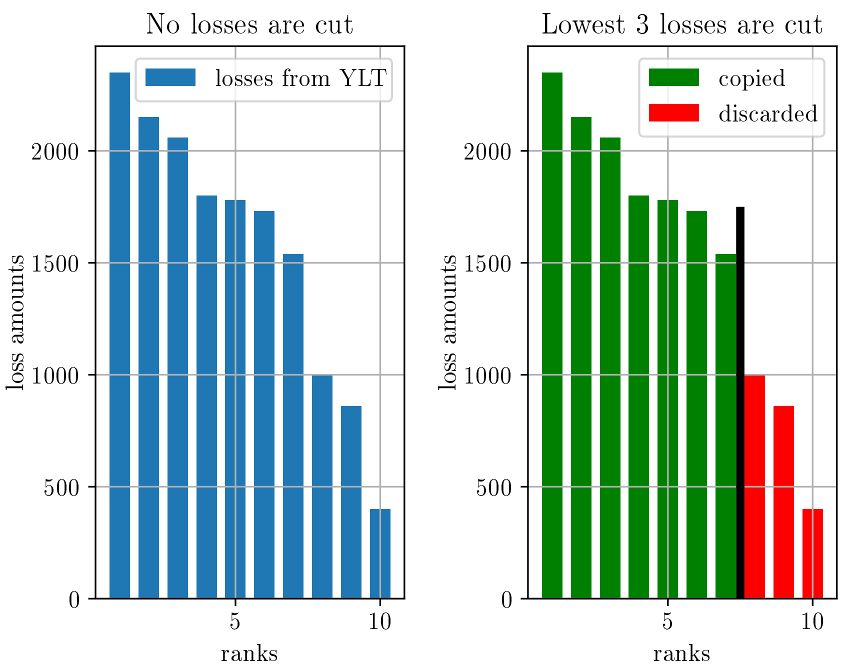


Figure 5.6: Loss cutting

The left-hand picture in Figure 5.6 shows severities generated by a loss model, ordered by ranks. The right-hand picture illustrates which losses are copied or discarded in the YLT transformation process, where the cutting threshold (black solid line) results from the difference of the modelled frequencies in the YLT and YERT.

We will investigate this error and try to limit it (or even let it converge to zero) for specific frequency distributions. It is however important to keep in mind that we cannot eliminate the error completely.

In order to formally define and investigate the transformation of Section 5.1, we introduce a new notation and use the following

Assumption 5.2. In the remainder of this Thesis, we assume that the losses for each year of all encountered YLT's are simulated from a *known* frequency-severity model. Moreover, we investigate the mathematical properties of the transformation from Section 5.1 for a single YLT. Furthermore, YERT's are assumed to have the number of losses for each year drawn from a Poisson(2) distribution.

In other words, we only look at the transformation process of loss models such as the Poisson(2) model of company A.

5.3 Terminology and Notation

Keeping Assumption 5.2 in mind, we now want to develop a mathematical framework in order to express the transformation of Section 5.1 in a mathematical language and to investigate how it behaves in terms of expectation and variance. We denote the number of modelled years by $n(= 100'000)$. By our assumption, the loss amount for year one in the YLT is simulated from a frequency-severity model. Therefore, we can write it as

$$S_1 = \sum_{i=1}^{N_1} X_i.$$

Analogously, we write the loss amount for year one in the transformed model as

$$T_1 = \sum_{j=1}^{M_1} Y_j$$

with $M_1 \sim$ Poisson(2). The total claim amounts in the two models for the second year are then given by

$$S_2 = \sum_{i=N_1+1}^{N_1+N_2} X_i, \qquad T_2 = \sum_{j=M_1+1}^{M_1+M_2} Y_j,$$

and so on for the other years. Although this notation for the different years is very heavy, this will not be an obstacle as we rarely look at single years other than the first. Furthermore, we define

$$N^n := \sum_{k=1}^{n} N_k, \qquad M^n := \sum_{k=1}^{n} M_k.$$

The total claim amount over all n simulated years can then be written as

$$S^n := \sum_{i=1}^{N^n} X_i = \sum_{k=1}^{n} S_k, \qquad T^n := \sum_{j=i}^{M^n} Y_j = \sum_{k=1}^{n} T_k.$$

By the transformation described in Section 5.1, the Y_j are actually obtained after a random permutation is applied to the X_i, so we need to extend our notation a bit further to capture this connection.

Definition 5.3. For a collection of random variables $X_1, X_2, \ldots$, we define

$$X^1 := \max(\{X_1, X_2, \ldots\}), \quad X^2 := \max(\{X_1, X_2, \ldots\} \setminus X^1) \tag{5.1}$$

and so on. Then we define

$$Y_j := X^{\Sigma(j)}, \tag{5.2}$$

where Σ is a *random permutation* of $\{1, 2, \ldots, M^n\}$ such that

$$\mathbb{P}[\Sigma = \sigma] = \frac{1}{M^n!}$$

for all $\sigma \in \mathcal{S}_{M^n}$, i.e. every element of $\mathcal{S}_{M^n}$ has the same probability to be picked by Σ. By abuse of notation, we will just write $\Sigma \in \mathcal{S}_{M^n}$ for such a random permutation.

In order to make our notation consistent, we need

Convention 5.4. For $i > N^n$, we define $X_i := 0$.

So if we sum over the severities X_i and the upper bound of the summation is bigger than N^n, then the additional severities are just zero. A priori, we could have also worked in equation (5.1) with minima instead of maxima. However, the notation with the maxima automatically implements the cutting of the lowest losses in the case where $M^n < N^n$, which is very convenient for our purposes. All these notations clearly indicate that the total claim amounts S_i and T_j have the same distribution by symmetry, more precisely

$$S_l \overset{(d)}{=} S_k, \quad T_l \overset{(d)}{=} T_k, \quad \text{for } 1 \leq l, k \leq n.$$

Therefore, we will in some situations omit the subscript and just write S (or T) instead of S_i (or T_j). Let us briefly summarize the involved random variables and their definitions:

- n $(= 100'000)$ is the number of modelled years.
- N_k are the iid frequencies of the actuarial loss model for the $1 \leq k \leq n$ modelled years.
- $M_k \overset{\text{iid}}{\sim}$ Poisson(2) model the YERT frequencies, independently of all N_l.
- X_i are the severities, which are positive and iid for $1 \leq i \leq N^n$ and $X_i := 0$ for all $i > N^n$.
- $Y_j = X^{\Sigma(j)}$ for all $1 \leq j \leq M^n$ for a random permutation $\Sigma \in \mathcal{S}_{M^n}$.

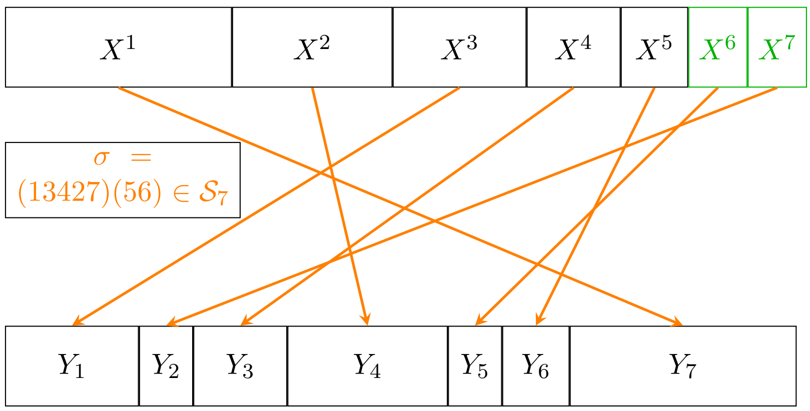


Figure 5.7: Loss permutation with $M^n > N^n$

Figure 5.7 shows the loss permutation for a specific realization $\Sigma = (13427)(56) \in \mathcal{S}_{M^n}$ of the permutation random variable in the case $7 = M^n > N^n = 5$, i.e. two zero losses $X^6 = X^7 = 0$ are added (we use Convention 5.4). On the other hand, Figure 5.8 shows the loss permutation for a specific realization $\Sigma = (134)(2) \in \mathcal{S}_{M^n}$ of the permutation random variable in the case $4 = M^n < N^n = 6$, i.e. the lowest two losses X^5 and X^6 are discarded. It is important to mention that we always consider a permutation in $\mathcal{S}_{M^n}$, and not in $\mathcal{S}_{N^n}$. We see moreover in Figure 5.8 that the cutting of the redundant losses is implemented automatically as the four severities Y_1, Y_2, Y_3 and Y_4 of the transformed model are obtained after a permutation is applied to the four biggest losses X^1, X^2, X^3 and X^4 of the initial model.

All these notations allow us to rewrite the transformation of Section 5.1 in mathematical terms:

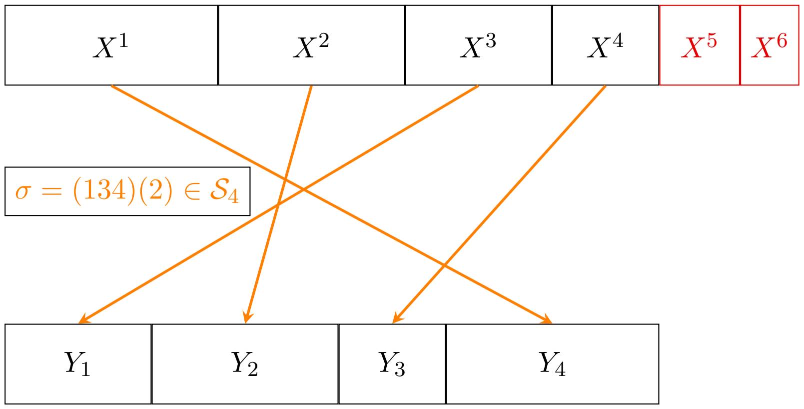


Figure 5.8: Loss permutation with $M^n < N^n$

Transformation 5.5. Let

$$S^n = \sum_{i=1}^{N^n} X_i = \sum_{k=1}^{n} S_k$$

be a frequency-severity model for the total loss amount of n simulated years. Then we approximate S^n with

$$T^n = \sum_{j=1}^{M^n} Y_j = \sum_{j=1}^{M^n} X^{\Sigma(j)} = \sum_{k=1}^{n} T_k$$

and S_k with T_k for all $1 \leq k \leq n$, for a random permutation $\Sigma \in \mathcal{S}_{M^n}$.

Remark 5.6. Let us quickly set Transformation 5.5 in context of the procedure explained at the beginning of this Chapter. S^n represents the loss modelled by the frequency-severity model of insurer A for n modelled years. A YLT can be obtained by a realization of the random variable S^n. On the other hand, our notation ensures that T^n corresponds exactly to the synthetic YLT after the transformation is applied. Here, it is a random variable; the parts which are random are M^n (the number of entries in the pre-generated YERT), Σ (the random rank assignment in the YERT) and the X_i (the losses of our original model). Consequently, if we pre-generate a YERT (i.e. consider realizations of M^n and Σ) only the losses of company A will we random. This allows us to formally define:

Definition 5.7. A *pre-generated YERT* is a realization of the random vector (M^n, Σ).

5.4 General mathematical properties

Before we look at specific transformations, let us have a look at some general mathematical properties. Let us start with the mathematical version of the loss cutting error (for an intuitive explanation, see Section 5.2):

Theorem 5.8 (Loss cutting error). *We have*

$$\mathbb{E}[T^n] < \mathbb{E}[S^n].$$

Proof. We can write

$$\begin{aligned}
\mathbb{E}[T^n] &= \mathbb{E}[T^n \cdot \mathbb{1}_{\{M^n \geq N^n\}}] + \mathbb{E}[T^n \cdot \mathbb{1}_{\{M^n < N^n\}}] \\
&= \mathbb{P}[M^n \geq N^n] \cdot \underbrace{\mathbb{E}[T^n | M^n \geq N^n]}_{=\mathbb{E}[S^n]} \\
&+ \mathbb{P}[M^n < N^n] \cdot \underbrace{\mathbb{E}[T^n | M^n < N^n]}_{<\mathbb{E}[S^n]} \\
&< \underbrace{(\mathbb{P}[M^n \geq N^n] + \mathbb{P}[M^n < N^n])}_{=1} \cdot \mathbb{E}[S^n] \\
&= \mathbb{E}[S^n],
\end{aligned}$$

as desired. □

It turns out that T_1 and S_1 have different distributions. We can show that by looking at the probability to produce a loss size of zero. To simplify the argument, we look at a specific choice of the N_k.

Proposition 5.9. *Let* $N_k \overset{iid}{\sim}$ *Poisson*(2)*, then* $\mathbb{P}[T_1 = 0] > \mathbb{P}[S_1 = 0]$.

Proof. We have

$$\mathbb{P}[S_1 = 0] = \mathbb{P}\left[\sum_{i=1}^{N_1} X_i = 0\right] = \mathbb{P}[N_1 = 0] = e^{-2},$$

since $X_i > 0$ $\mathbb{P}$-a.s. On the other hand, we have

$$\begin{aligned}
\mathbb{P}[T_1 = 0] &= \mathbb{P}\left[\sum_{j=1}^{M_1} Y_j = 0\right] \\
&> \underbrace{\mathbb{P}[M_1 = 0]}_{=e^{-2}} + \underbrace{\mathbb{P}[M^n = N^n + 1, M_1 = 1, Y_1 = X^{M^n}]}_{>0} \\
&> e^{-2}.
\end{aligned}$$

□

Example 5.10. Let us have a look at Table 5.1, which visualizes the argument in the last proof. We model $n = 3$ years. In the original model, we have in total $N^n = 3$ losses (zero in year 1, two in year 2 and one in year 3) and in the synthetic model, we have $M^n = 4$ losses, i.e. one of these losses must then be an additional zero loss. Column 3 indicates the ranks of the three claims from the actuarial loss model and column 5 contains the randomly assigned ranks of the four losses in the synthetic model. We see that the loss in the synthetic model in year one has rank 4, i.e. it is the additional zero loss (marked red), so we have a total claim amount of zero for year one in the synthetic model, even though we have modelled a positive number of claims and all losses in the actuarial model are positive. Exactly this case corresponds to the second underbrace in the last proof ($M^n = 4 = N^n + 1$, $M_1 = 1$, $Y_1 = X^4 = X^{M^n} = 0$).

Table 5.1: Model transformation for $n = 3$ simulated years

<table>
<tr><th>Simulated Year</th><th>Simulated Loss original model</th><th>Rank</th><th>Simulated Loss synthetic model</th><th>random Rank</th></tr>
<tr><td>1</td><td>-</td><td>-</td><td>0</td><td>4</td></tr>
<tr><td rowspan="2">2</td><td>127</td><td>3</td><td rowspan="2">831</td><td rowspan="2">1</td></tr>
<tr><td>831</td><td>1</td></tr>
<tr><td rowspan="2">3</td><td rowspan="2">649</td><td rowspan="2">2</td><td>127</td><td>3</td></tr>
<tr><td>649</td><td>2</td></tr>
</table>

Next, let us discuss some results about (in)dependence.

Proposition 5.11 (Dependence of X_i and Y_j)**.** *The random vectors*

$$(X_1, X_2, \dots, X_{N^n}) \text{ and } (Y_1, Y_2, \dots, Y_{M^n})$$

are not stochastically independent.

Proof. Let $x > 0$ be such that $\mathbb{P}[Y_1 > x] > 0$ and $\mathbb{P}[X_1 \leq x] > 0$. Such an x exists if X_1 is a positive and non-deterministic random variable (which is a natural assumption for insurance claim severities). Then we have

$$\mathbb{P}[X_1 \leq x, X_2 \leq x, \dots, X_{N^n} \leq x] > 0$$

and

$$\mathbb{P}[Y_1 > x] > 0,$$

but

$$\mathbb{P}[X_1 \leq x, X_2 \leq x, \dots, X_{N^n} \leq x, Y_1 > x] = 0,$$

so the two random vectors cannot be independent.

□

Remark 5.12. The last proof essentially used the property that the losses Y_j are just copies of the X_i or zero losses. So if all X_i stay below a level x, then the same must hold for all $Y_j's$.

Remark 5.13. Recall that by the properties of frequency-severity models, the random variables $X_1, X_2, \dots X_{N^n}$ are jointly (and thus also pairwise) independent.

Proposition 5.14 (Dependence of Y_j on N^n and M^n)**.** *The random variables Y_j for $1 \leq j \leq M^n$ are not independent of N^n. Moreover, the random variables Y_j for $1 \leq j \leq M^n$ are as well not independent of M^n. In particular, T and T^n from Transformation 5.5 are* never *frequency-severity models.*

Proof. We have for $1 \leq j \leq M^n$

$$\mathbb{P}[Y_j > 0] > 0,$$

but

$$\mathbb{P}[Y_j > 0 | N^n = 0] = 0,$$

so the random variables Y_j and N^n are not independent. For the second statement, note that since $X_i > 0$ $\mathbb{P}$-a.s.

$$\mathbb{P}[Y_j = 0] > 0 \iff M^n > N^n,$$

so the loss sizes Y_j depend as well on M^n. In a frequency-severity model, the severities must be independent of the frequency. The argument shows that this is never the case for the transformed models T_1 (and also T^n), which implies that Transformation 5.5 produces loss models which are *never* frequency-severity models. □

5.5 Transformation of Poisson(2)-models

As we have modelled the losses of insurer A with a Poisson(2) frequency-severity model, we now want to find out how this model behaves under Transformation 5.5 in terms of the expectation. Note that both the original and the transformed model work with the same frequency distribution for every modelled year. Hence, some losses are discarded with probability 0.5 for every n. In case we have a concrete realization $N^n - M^n > 0$, then the loss amount which is cut equals

$$\sum_{i=M^n+1}^{N^n} X^i.$$

It turns out to be very hard to estimate the expectations of these error terms for general severities. We will therefore content ourselves with a simulation for this special case. Afterwards, we slightly change the setting in order to circumvent the above mentioned problem. This can be done via two different ways. Firstly, we can apply as in Chapter 4 layers to each and every loss. The deductible D must be chosen sufficiently high such that the cut losses are (in average) below D and play therefore not a relevant role. Secondly, we can work with a Poisson frequency strictly smaller than 2 in our initial model. The probability to cut some losses in the transformation will then converge to zero, which means that estimating the size of the cut losses will no longer be necessary.

Example 5.15. Let us simulate the trasformation with a concrete example. We choose $N_k \sim \text{Poisson}(2)$ and $X_i \sim \text{Pareto}(x_0 = 2, \alpha = 1.5)$.

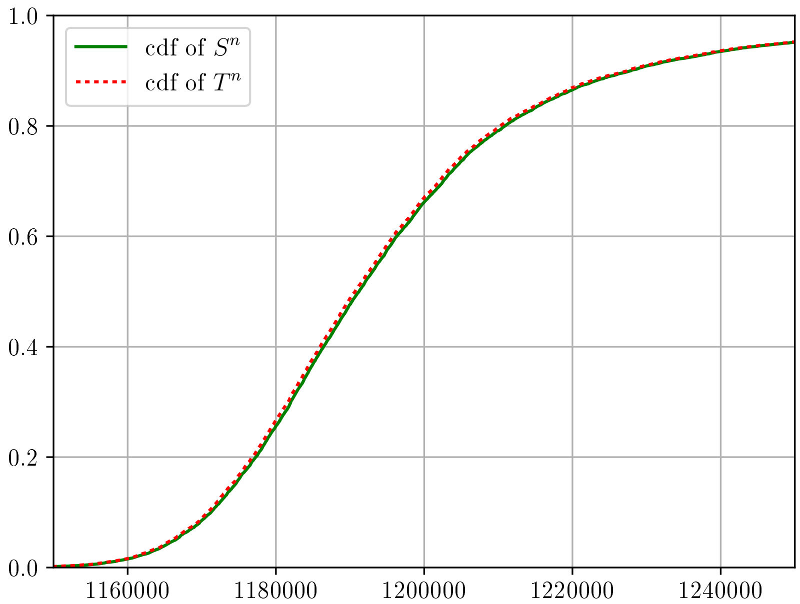


Figure 5.9: Cumulative distribution functions of S^n and T^n

Figure 5.9 shows the simulated cdf's of S^n and T^n with $n = 100'000$ modelled years. We see from the picture that the cdf of S^n is at every point smaller than the cdf of T^n. This suggests again $\mathbb{E}[T^n] < \mathbb{E}[S^n]$. We see that the values on the x-axis have a size around $1'200'000 = 200'000 \cdot 6 = \mathbb{E}[N^n] \cdot \mathbb{E}[X_1] = \mathbb{E}[S^n]$, which is reasonable. We can also have a look at the expectations of T^n and S^n. As already pointed out, $\mathbb{E}[S^n] = 1'200'000$ and the simulation used to generate Figure 5.9 yields $1'198'571$ as (simulated) expectation of T^n. In conclusion, we could reproduce the mismatch of the expectations as suggested in Theorem 5.8, but we see at the same time that the difference is very small (around 0.12%). Thus, the simulation hints that even in the limiting case where both the original and the transfomed model work with exactly the same frequency distribution, Transformation 5.5 gives very reasonable results (in terms of the expectation) and therefore an analysis of the error terms $\sum_{i=M^n+1}^{N^n} X^i$ does not lead to considerable improvements (at least in this example).

5.6 Application of Per Risk XL Layers

Let us now discuss the application of Per Risk Excess of Loss Layers. Our goal is to show that if we apply suitable layers to our losses, we can control the loss cutting error and let it converge to zero if n goes to infinity.

Lemma 5.16. *Let $X \sim Pareto(x_0, \alpha)$ with $\alpha > 1$ and let $L_{D,C}$ be a Layer function with $D \geq x_0$. Then*

$$\mathbb{E}[L_{D,C}(X)] = \frac{x_0^\alpha}{\alpha - 1}\left[D^{1-\alpha} - (D+C)^{1-\alpha}\right].$$

Proof. By Proposition 3.28 we get

$$\begin{aligned}\mathbb{E}[L_{D,C}(X)] &= \int_D^{D+C} (1 - F_X(x))\,\mathrm{d}x = \int_D^{D+C} \left(\frac{x_0}{x}\right)^\alpha \mathrm{d}x \\ &= x_0^\alpha \left[\frac{1}{1-\alpha} x^{1-\alpha}\right]_D^{D+C} = \frac{x_0^\alpha}{\alpha-1}\left[D^{1-\alpha} - (D+C)^{1-\alpha}\right].\end{aligned}$$

□

Theorem 5.17. *Let $N_k \overset{iid}{\sim} Poisson(2)$, $X_i \overset{iid}{\sim} Pareto(x_0, \alpha)$ with $\alpha > 1$ and let $L_{D,C}$ be a layer applied to each and every loss with $D > x_0$. Then*

$$\mathbb{E}[T_1] \longrightarrow \mathbb{E}[S_1], \quad \textit{for } n \longrightarrow \infty.$$

Before we start with the proof, let us briefly sketch the underlying idea: By the law of large numbers, we have $\frac{1}{2n}(N^n - M^n) \to 0$ for $n \to \infty$. If it is positive, this random variable is approximately equal to the portion of cut losses (since $\mathbb{E}[N^n] = \mathbb{E}[M^n] = 2n$). On the other hand, the number of losses below D is roughly equal to $2n \cdot F_X(D)$, if F_X denotes the severity cdf, i.e. it grows *linearly* in n, so the portion of of losses below D is approximately equal to $F_X(D)$, i.e. it stays around the same value with increasing n. Since $\frac{1}{2n}(N^n - M^n)$ is a random variable converging almost surely to zero, the probability to cut losses above D converges to zero as well. Figure 5.10 visualizes this thought: The red solid line is the deductible threshold and marks the portion of losses below D (the number of losses below D will be called W_D^n in the proof and is a random variable, however, $\frac{1}{n}W_D^n$ converges a.s. to $F_X(D)$ as we will see). The green lines show realizations of $\frac{1}{2n}(N^n - M^n)$, which get closer to zero for increasing n (attention: we consider here only realizations where $N^n > M^n$, since no losses are cut in the other case). By the argument above, the probability that the green lines lie above the red line (which means that non-zero losses are cut) becomes vanishingly small for increasing n, which is exactly what we utilize in the proof. We will also use the following result

Theorem 5.18 (Glivenko-Cantelli). *Let $(X_i)_{i=1}^n$ be an iid sequence of real valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with cumulative distribution function F. Denote by $\hat{F}_n(\cdot)$ the* empirical distribution function

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}}.$$

Then

$$\|\hat{F}_n - F\|_\infty = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \longrightarrow 0 \quad \mathbb{P}\text{-}a.s.\ for\ n \longrightarrow \infty.$$

This result including a proof can be found in Billingsley [7], Theorem 20.6.

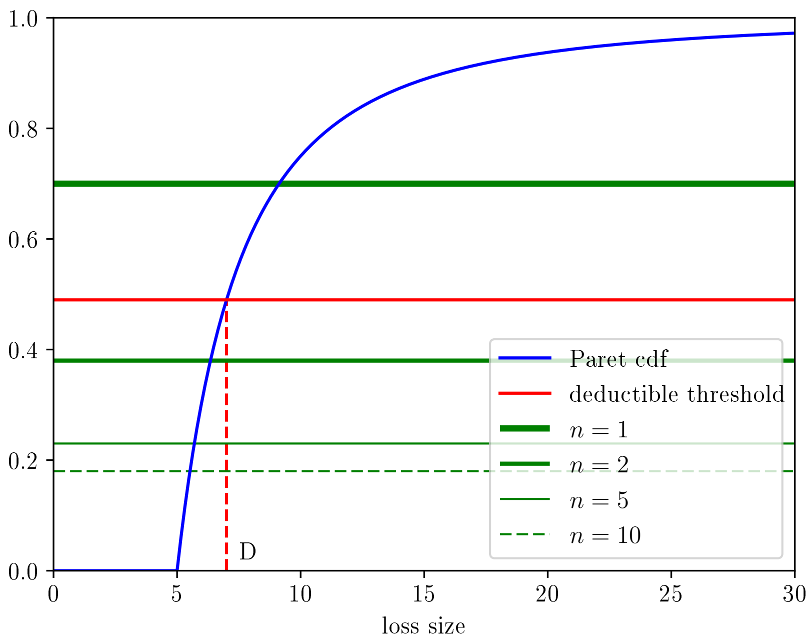


Figure 5.10: Visualization of the idea in the proof

Proof of Theorem 5.17. By the first Wald identity, we have

$$\mathbb{E}[S_1] = \mathbb{E}[N_1] \cdot \mathbb{E}[L_{D,C}(X_1)] = 2 \cdot \mathbb{E}[L_{D,C}(X_1)]. \tag{5.3}$$

To simplify the notation, let us denote by B the event that some non-zero losses are discarded in the transformation. Then

$$\begin{aligned}\mathbb{E}[T_1] &= \mathbb{E}\left[\sum_{j=1}^{M_1} L_{D,C}(Y_j)\right] = \mathbb{E}\left[\left(\sum_{j=1}^{M_1} L_{D,C}(Y_j)\right)\mathbb{1}_{\{B\}}\right] \\ &+ \mathbb{E}\left[\left(\sum_{j=1}^{M_1} L_{D,C}(Y_j)\right)\mathbb{1}_{\{B^c\}}\right] \\ &= \mathbb{P}[B]\cdot\mathbb{E}\left[\sum_{j=1}^{M_1} L_{D,C}(Y_j)\middle| B\right] + \mathbb{P}[B^c]\cdot\mathbb{E}\left[\sum_{j=1}^{M_1} L_{D,C}(Y_j)\middle| B^c\right]. \end{aligned} \tag{5.4}$$

If we define

$$W_D^n := \sum_{i=1}^{N^n} \mathbb{1}_{\{X_i \le D\}},$$

then

$$\mathbb{P}[B] = \mathbb{P}[N^n - M^n > W_D^n] = \mathbb{P}\left[\frac{1}{n}(N^n - M^n) > \frac{1}{n}W_D^n\right]. \tag{5.5}$$

By the Theorem of Glivenko-Cantelli, we have

$$\frac{1}{n}W_D^n \xrightarrow{n\to\infty} F_X(D) \quad \mathbb{P}\text{-a.s.},$$

if F_X denotes the cumulative distribution function of the severity. Since we assumed a Pareto severity with $D > x_0$, we have $F_X(D) > 0$. On the other hand, the strong law of large numbers implies

$$\frac{1}{n}(N^n - M^n) \xrightarrow[\text{SLLN}]{n\to\infty} 0 \quad \mathbb{P}\text{-a.s.}$$

Altogether, we get with equation (5.5): $\mathbb{P}[B] \xrightarrow{n\to\infty} 0$. Moreover, we have

$$0 \le \mathbb{E}\left[\sum_{j=1}^{M_1} L_{D,C}(Y_j)\middle| B\right] \le \mathbb{E}[M_1]\cdot C = 2\cdot C < \infty,$$

so

$$\mathbb{P}[B]\cdot\mathbb{E}\left[\sum_{j=1}^{M_1} L_{D,C}(Y_j)\middle| B\right] \xrightarrow{n\to\infty} 0.$$

Note that

$$\mathbb{E}\left[\sum_{j=1}^{M^n} L_{D,C}(Y_j)\middle| B^c\right] = \mathbb{E}\left[\sum_{i=1}^{N^n} L_{D,C}(X_i)\right] = 2n\cdot\mathbb{E}[L_{D,C}(X_1)]. \tag{5.6}$$

Since all the T_j have the same distribution, we have

$$n \cdot \mathbb{E}[T_1|B^c] = n \cdot \mathbb{E}\left[\sum_{j=1}^{M_1} L_{D,C}(Y_j) \middle| B^c\right] = \mathbb{E}\left[\sum_{j=1}^{M^n} L_{D,C}(Y_j) \middle| B^c\right]. \quad (5.7)$$

Combining equations (5.6), (5.7) and (5.3) we get

$$n \cdot \mathbb{E}[T_1|B^c] = 2n \cdot \mathbb{E}[L_{D,C}(X_1)] \implies \mathbb{E}[T_1|B^c] = 2 \cdot \mathbb{E}[L_{D,C}(X_1)] = \mathbb{E}[S_1].$$

In total, we receive with equation (5.4)

$$\mathbb{E}[T_1] = \underbrace{\mathbb{P}[B]}_{\to 0} \cdot \underbrace{\mathbb{E}[T_1|B]}_{\leq 2 \cdot C} + \underbrace{\mathbb{P}[B^c]}_{\to 1} \cdot \underbrace{\mathbb{E}[T_1|B^c]}_{=\mathbb{E}[S_1]} \xrightarrow{n \to \infty} \mathbb{E}[S_1],$$

as desired. □

Remark 5.19. In the proof of the last Theorem, we did not really need that the severities follow a Pareto distribution. All we used was $F_X(D) > 0$ and the proof works for any severity that satisfies this condition.

Example 5.20. Let us visualize the convergence proved in Theorem 5.17 in a concrete example. We consider a frequency-severity model with Poisson(2) frequency and Pareto($x_0 = 1, \alpha = 1.5$) severity. Moreover, we apply a layer $L_{D,C}(X)$ with $D = 1.5$ and $C = 3$ to each and every loss. Figure 5.11 shows the cdf of S_1 and T_1 for a varying number of modelled years n (note that the cdf of S_1 does not depend on n). We see that the corresponding distribution functions of T_1 get closer and closer to S_1 for increasing n. We can also see that the model T_1 gives us a higher probability for a total claim amount of zero than the model S_1, which is in line with Proposition 5.9. Note that a value of $n = 100$ modelled years already gives a very good approximation, so the default value $n = 100'000$ in a concrete application is sufficiently high.

Figure 5.12 shows the expectations of $\mathbb{E}[T_1]$ for $n = 1$ up to $n = 100$ modelled years. Note that these expected values stem from a simulation and are *not* computed analytically (more precisely, T_1 was realized N = 1'000'000 times in each of the 100 scenarios and the average of these iid realizations gives the empirical expected value). We see that already a small number of simulated years leads to a very good approximation of $\mathbb{E}[S_1]$ by $\mathbb{E}[T_1]$. We use Lemma 5.16 to compute $\mathbb{E}[S_1]$ analytically (blue line):

$$\begin{aligned}\mathbb{E}[S_1] &= \mathbb{E}[N_1] \cdot \mathbb{E}[L_{D,C}(X)] = 2 \cdot \frac{x_0^\alpha}{\alpha - 1}[D^{1-\alpha} - (D+C)^{1-\alpha}] \\ &= 4 \cdot [1.5^{-0.5} - 4.5^{-0.5}] \approx 1.38.\end{aligned}$$

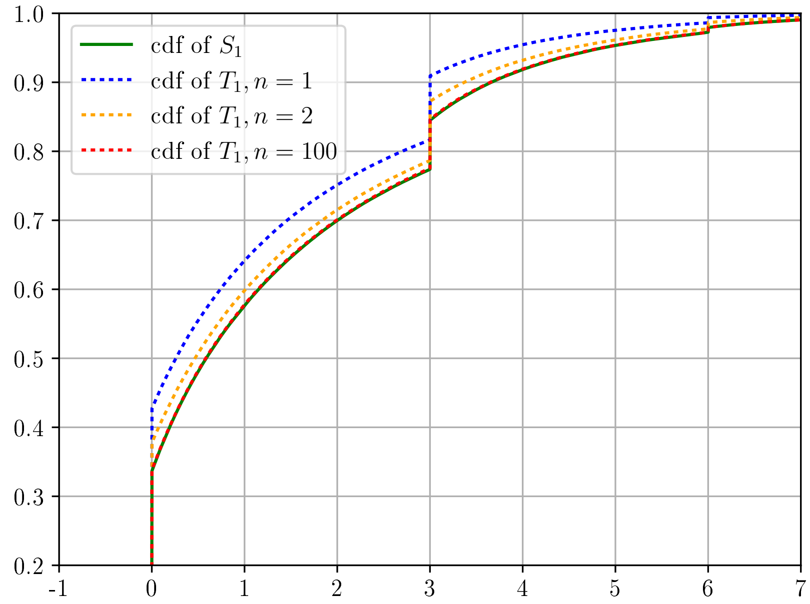


Figure 5.11: Cumulative distribution functions of S_1 and T_1 for different n

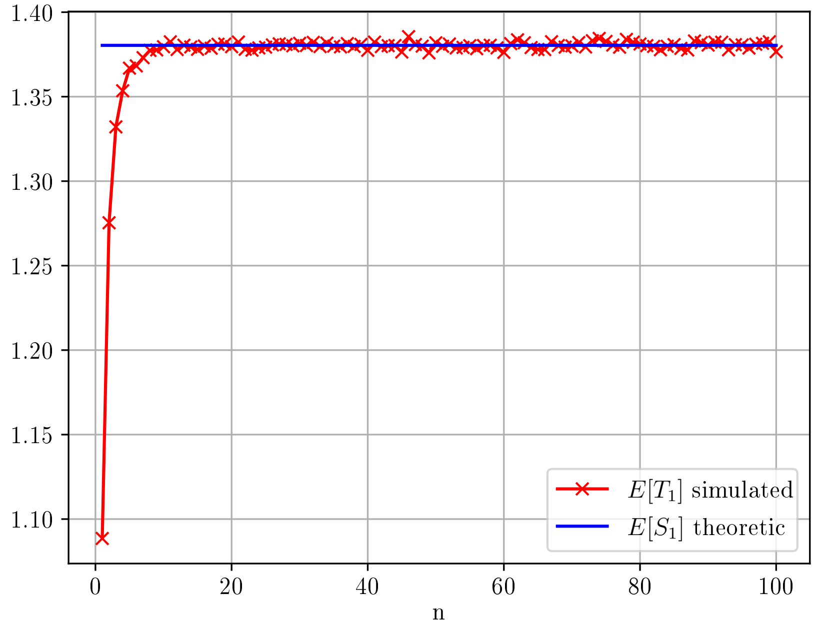


Figure 5.12: $\mathbb{E}[S_1]$ and $\mathbb{E}[T_1]$ for different n

5.7 Transformation of Poisson(2-ϵ) models

In a next step, we apply Transformation 5.5 to Poisson frequency-severity models with parameter value $2-\epsilon$ for $0<\epsilon<2$. There are two reasons for this course of action. First of all, most natural catastrophes occur with expected frequencies smaller than 2, so reducing the Poisson parameter brings us - informally speaking - closer to the reality. On the other hand, we also get some mathematical advantages: Since we leave the expected frequency for the YERT at 2, the probability to cut losses converges to zero if n goes to infinity. In other words, the loss cutting error will converge to zero (as we will see). Instead of reducing the frequency parameter of the frequency-severity model, one could also increase the frequency parameter of the YERT, as it is only relevant that the two Poisson parameters are distinct. Formally, we write for a (small) positive value $\epsilon>0$

$$\mu_n := \frac{1}{n}N^n - (2-\epsilon), \quad \tau_n := \frac{1}{n}M^n - 2, \quad \rho_n := \mu_n - \tau_n.$$

Since

$$N^n = \sum_{i=1}^{n} N_i, \quad M^n = \sum_{i=1}^{n} M_i,$$

we get by the strong law of large numbers

$$\frac{1}{n}N^n \xrightarrow{n\to\infty} (2-\epsilon) \quad \mathbb{P}\text{-a.s.}, \quad \frac{1}{n}M^n \xrightarrow{n\to\infty} 2 \quad \mathbb{P}\text{-a.s.},$$

i.e.

$$\mu_n, \tau_n, \rho_n \xrightarrow{n\to\infty} 0, \ \mathbb{P}\text{-a.s.}$$

Therefore,

$$\begin{aligned}\mathbb{P}[M^n < N^n] &= \mathbb{P}[N^n - M^n > 0] = \mathbb{P}[-n\epsilon + n\mu_n - n\tau_n > 0] \\ &= \mathbb{P}[\rho_n > \epsilon] \xrightarrow{n\to\infty} 0 \quad \text{for all } \epsilon > 0,\end{aligned}$$

i.e. the probability to cut losses converges to zero for $n\to\infty$. This convergence result is not sufficient, as we need more precise information how fast $\mathbb{P}[M^n < N^n] = \mathbb{P}[N^n - M^n > 0]$ converges to zero. The apposite tool for this question is the so called *Skellam distribution*, which we introduce in the next section.

5.8 The Skellam distribution and its properties

Definition 5.21. Let N_1 and N_2 be two independent Poisson random variables with parameters λ_1 and λ_2. Define $K = N_1 - N_2$. The distribution of K is called the *Skellam distribution* related to N_1 and N_2.

Theorem 5.22 (Skellam). *Let N_1 and N_2 be two independent Poisson random variables with parameters λ_1 and λ_2. Then the Skellam distribution related to N_1 and N_2 has the following properties for any $k \in \mathbb{Z}$*

1. $\mathbb{P}[K = k] = e^{-(\lambda_1+\lambda_2)} \left(\frac{\lambda_1}{\lambda_2}\right)^{k/2} I_{|k|}(2\sqrt{\lambda_1\lambda_2})$.
2. *If* $\lambda_1 < \lambda_2$, *then* $\mathbb{P}[K \geq 0] \leq e^{-(\sqrt{\lambda_1}-\sqrt{\lambda_2})^2}$,

where I_k denotes the modified Bessel function of the first kind.

Proof. For the first point, see Skellam [3]. For the second part, note that

$$\mathbb{P}[K \geq 0] = \mathbb{P}[N_1 - N_2 \geq 0] = \mathbb{P}[e^{t(N_1-N_2)} \geq 1] \quad \text{for all } t > 0.$$

Using Markov's inequality, we get

$$\mathbb{P}[K \geq 0] \leq \frac{\mathbb{E}[e^{t(N_1-N_2)}]}{1} = M_K(t) \quad \text{for all } t > 0.$$

Note that

$$M_K(t) = M_{N_1}(t) \cdot M_{N_2}(-t) = e^{-(\lambda_1+\lambda_2)+\lambda_1 e^t+\lambda_2 e^{-t}}$$

due to independence of N_1 and N_2. The moment generating function of a Poisson random variable can be found in Wüthrich [5], page 31. Choosing

$$t = \log\left(\sqrt{\frac{\lambda_2}{\lambda_1}}\right) > 0$$

yields

$$M_K(t) = e^{-(\lambda_1+\lambda_2)+2\sqrt{\lambda_1\lambda_2}} = e^{-(\sqrt{\lambda_1}-\sqrt{\lambda_2})^2},$$

i.e.

$$\mathbb{P}[K \geq 0] \leq e^{-(\sqrt{\lambda_1}-\sqrt{\lambda_2})^2},$$

as desired. □

Figure 5.13 shows the probability mass functions of two Poisson random variables N_1 and N_2 with parameters 5 and 2 respectively. The red curve shows the probability mass function of $N_1 - N_2$ (i.e. the Skellam distribution of N_1 and N_2). Note that the Poisson random variables can only attain non-negative values, while the Skellam random variable can attain every integer with positive probability. We also see from the picture that the Skellam tails decay quite fast.

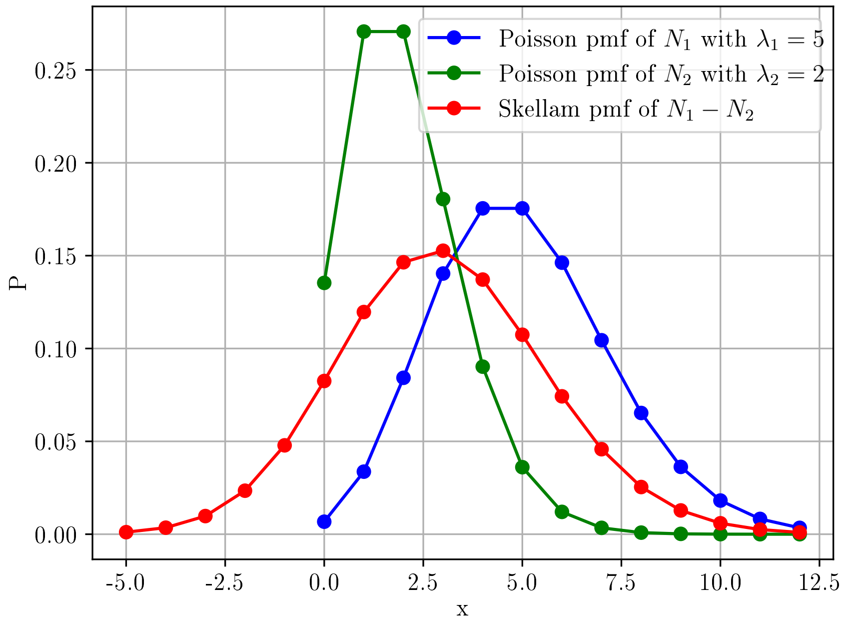


Figure 5.13: Skellam and Poisson probability mass functions

5.9 Convergence for Poisson$(2-\epsilon)$ models

We are now able to investigate if and how $\mathbb{E}[T^n]$ approaches $\mathbb{E}[S^n]$ for n going to infinity if we work with Poisson$(2-\epsilon)$ frequencies in the initial model. We must take into consideration that $\mathbb{E}[S^n]$ itself goes to infinity for $n \to \infty$, since $\mathbb{E}[N^n]$ is growing linearly in n. Therefore, we cannot expect a result of the form $\mathbb{E}[T^n] \to \mathbb{E}[S^n]$, for $n \to \infty$, since this does not make sense mathematically. However, we can still answer the question of convergence by looking at the difference $\mathbb{E}[S^n] - \mathbb{E}[T^n]$ and we will see that this expression converges to zero for $n \to \infty$.

Theorem 5.23. *Let $N_k \overset{iid}{\sim}$ Poisson$(2-\epsilon)$ for $0 < \epsilon < 2$ and let S^n and T^n be as in Transformation 5.5, then we have*

$$|\mathbb{E}[S^n] - \mathbb{E}[T^n]| \longrightarrow 0 \text{ for } n \longrightarrow \infty.$$

More precisely, we obtain the following rate of convergence:

$$|\mathbb{E}[S^n] - \mathbb{E}[T^n]| \leq 2(2-\epsilon)n \cdot \mathbb{E}[X_1] \cdot e^{-nk},$$

for

$$k = \left(\sqrt{2} - \sqrt{2-\epsilon}\right)^2.$$

Proof. Note that by Theorem 3.21

$$\mathbb{E}[S^n] = \mathbb{E}[N^n] \cdot \mathbb{E}[X_1] = (2-\epsilon)n \cdot \mathbb{E}[X_1].$$

Next, we can write

$$\begin{aligned}\mathbb{E}[T^n] &= \mathbb{E}\left[\sum_{j=1}^{M^n} Y_i\right] \\ &= \underbrace{\mathbb{E}\left[\left(\sum_{j=1}^{M^n} Y_j\right) \mathbb{1}_{\{M^n \le N^n\}}\right]}_{\mathrm{I}} + \underbrace{\mathbb{E}\left[\left(\sum_{j=1}^{M^n} Y_j\right) \mathbb{1}_{\{M^n > N^n\}}\right]}_{\mathrm{II}}.\end{aligned}$$

For I, we have with $k = \left(\sqrt{2} - \sqrt{2-\epsilon}\right)^2$, $\Sigma \in \mathcal{S}_{M^n}$ and part 2 of Theorem 5.22

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j=1}^{M^n} Y_j\right) \mathbb{1}_{\{M^n \le N^n\}}\right] &= \underbrace{\mathbb{P}[M^n \le N^n]}_{\le e^{-nk}} \cdot \mathbb{E}\left[\sum_{j=1}^{M^n} Y_j \middle| M^n \le N^n\right] \\ &\le e^{-nk} \cdot \mathbb{E}\left[\sum_{j=1}^{M^n} X^{\Sigma(j)} \middle| M^n \le N^n\right] \\ &\le e^{-nk} \cdot \mathbb{E}\left[\sum_{i=1}^{N^n} X_i\right] \\ &= e^{-nk} \cdot (2-\epsilon)n \cdot \mathbb{E}[X_1].\end{aligned}$$

On the other hand, we get for II:

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{j=1}^{M^n} Y_j\right) \mathbb{1}_{\{M^n > N^n\}}\right] &= \mathbb{P}[M^n > N^n] \cdot \mathbb{E}\left[\sum_{j=1}^{M^n} Y_j \middle| M^n > N^n\right] \\ &= \mathbb{P}[M^n > N^n] \cdot \mathbb{E}[N^n] \cdot \mathbb{E}[X_1] \\ &= \mathbb{P}[M^n > N^n] \cdot (2-\epsilon)n \cdot \mathbb{E}[X_1].\end{aligned}$$

If we write informally $\mathbb{E}[T^n] = \mathrm{I} + \mathrm{II}$, we get with the triangle inequality

$$\begin{aligned}|\mathbb{E}[S^n] - \mathbb{E}[T^n]| = |\mathbb{E}[S^n] - (\mathrm{I}+\mathrm{II})| &\le \mathrm{I} + |\mathbb{E}[S^n] - \mathrm{II}| \\ &\le (2-\epsilon)n \cdot \mathbb{E}[X_1] \cdot \left(e^{-nk} + \underbrace{1 - \mathbb{P}[M^n > N^n]}_{=\mathbb{P}[M^n \le N^n] \le e^{-nk}}\right) \\ &\le 2(2-\epsilon)n \cdot \mathbb{E}[X_1] \cdot e^{-nk} \xrightarrow{n\to\infty} 0,\end{aligned}$$

as desired. □

Corollary 5.24. *With the same setting as in Theorem 5.23, we have*

$$\mathbb{E}[T] \to \mathbb{E}[S], \text{ for } n \to \infty.$$

Proof. By definition and the property that all T_j have the same distribution, we have

$$T^n = \sum_{j=1}^{n} T_k, \quad \text{and} \quad \mathbb{E}[T^n] = n \cdot \mathbb{E}[T].$$

Note that the same holds for S^n and S. Using the previous result, we get

$$|\mathbb{E}[T] - \mathbb{E}[S]| = \frac{1}{n}|\mathbb{E}[T^n] - \mathbb{E}[S^n]| \xrightarrow{n\to\infty} 0.$$

Since $\mathbb{E}[S]$ and $\mathbb{E}[T]$ are finite, this implies that $\mathbb{E}[T]$ converges to $\mathbb{E}[S]$ for $n \to \infty$. □

Corollary 5.25. *With the same setting as in Theorem 5.23, we have*

$$\mathbb{E}[T] \to \mathbb{E}[S], \text{ for } n \to \infty$$

if we apply a layer $L_{D,C}(\cdot)$ to each and every loss.

Proof. In the proof of Theorem 5.23, we did not assume a specific severity distribution. Therefore, we can replace X_i by $L_{D,C}(X_i)$ and the arguments of the last two results still work. □

Remark 5.26. Corollary 5.25 is essentially a generalization of Theorem 5.17 to the case where the expected frequencies are below 2. However, we do not need additional assumptions on D in this case.

Example 5.27. Let us have a look at a concrete example to visualize Theorem 5.23. We choose

- $\epsilon = 0.1$,
- $X_i \overset{\text{iid}}{\sim} \text{Pareto}(\alpha = 2, x_0 = 1)$,
- N_k are iid and have a Negative Binomial distribution with expectation $2 - \epsilon$ (thus, N^n has a Negative Binomial distribution with expectation $n(2-\epsilon)$ by Propositon 3.9).

Then, we have

$$\mathbb{E}[S^n] = \mathbb{E}[N^n] \cdot \mathbb{E}[X_1] = n(2-\epsilon) \cdot \frac{\alpha x_0}{\alpha - 1} = 2n(2-\epsilon) = 3.8n. \tag{5.8}$$

Moreover, we have

$$k = (\sqrt{2} - \sqrt{2-\epsilon})^2 \approx 0.0013,$$

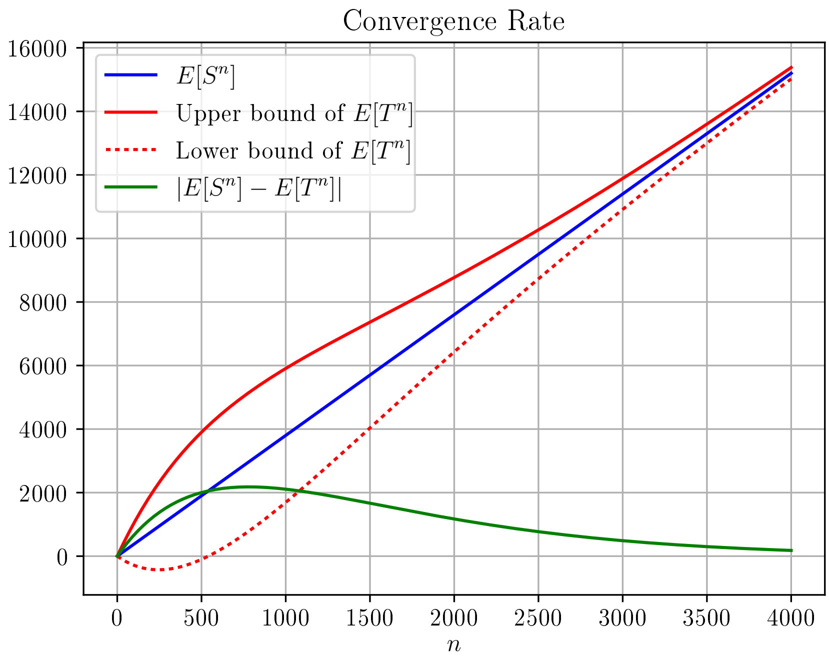


Figure 5.14: Convergence of $|\mathbb{E}[S^n] - \mathbb{E}[T^n]|$

i.e. the convergence of $|\mathbb{E}[S^n] - \mathbb{E}[T^n]|$ is not very fast, as the exponential decay factor is quite small.

Figure 5.14 shows $\mathbb{E}[S^n]$ depending on n, which grows linearly in n as already computed in equation (5.8) (blue line). The red lines show the boundary term $2(2-\epsilon)n \cdot \mathbb{E}[X_1] \cdot e^{-nk}$ added and subtracted to $\mathbb{E}[S^n]$, i.e. $\mathbb{E}[T^n]$ must lie between the dotted and solid red line. Finally, the green curve plots $2(2-\epsilon)n \cdot \mathbb{E}[X_1] \cdot e^{-nk}$. We see that a few thousand modelled years n are necessary such that this term can be neglected. As the image suggests, the default number of $n = 100'000$ is sufficient. In fact, one can compute explicitly for $n = 100'000$

$$2(2-\epsilon)n \cdot \mathbb{E}[X_1] \cdot e^{-nk} \approx 1.56 \cdot 10^{-50},$$

which is more than enough precision for any application.

Theorem 5.23 gives us a very strong convergence result for the transformation of Poisson frequency-severity models with expected frequency smaller than 2. The Poisson assumption was crucial as it enabled us to use the properties of the Skellam distribution (especially the second part of Theorem 5.22).

The whole modelling process for company A always relied on Poisson frequencies and we have seen good reasons why this assumption is reasonable.

However, there might still be cases where this approach is not suitable: If the past loss data of company A shows e.g. significant overdispersion, the calibration of Poisson models leads to mediocre results. Hence, it can be useful in certain situations to transform models with other frequencies as well. Moreover, it would be also interesting to generalize Theorem 5.23 form a purely theoretical point of view, and this is exactly what we are going to do in the next sections. To do so, we need another tool called *Large Deviations Theory*, as the Skellam distribution is only defined for Poisson random variables.

5.10 Large Deviations Theory

We follow Dembo and Zeitouni [1], pages 4-5 and 26-27.

5.10.1 Notation and Definitions

In the following, let $\{\mu_\epsilon\}$ for $\epsilon > 0$ be a family of probability measures on the measurable space $(\mathcal{X}, \mathcal{B})$, where we assume that $\mathcal{X}$ is a topological space and $\mathcal{B}$ the corresponding Borel-σ-Algebra.

Definition 5.28. A *rate function* I is a lower semicontinuous mapping $I\colon \mathcal{X} \to [0,\infty]$ (such that for all $\alpha \in [0,\infty)$, the level set $\Psi_I(\alpha) = \{x : I(x) \leq \alpha\}$ is a closed subset of $\mathcal{X}$). A *good rate function* is a rate function for which all level sets $\Psi_I(\alpha)$ are compact sets. The *effective domain* of I, denoted $\mathcal{D}_I$, is the set of points in $\mathcal{X}$ of finite rate (i.e. $\mathcal{D}_I = \{x : I(x) < \infty\}$). When no confusion occurs, we refer to $\mathcal{D}_I$ as the domain of I.

Definition 5.29. We say that $\{\mu_\epsilon\}$ satisfies the *Large Deviations Principle* with rate function I, if for all $\Gamma \in \mathcal{B}$,

$$-\inf_{x\in\Gamma^o} I(x) \leq \liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon(\Gamma) \leq \limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(\Gamma) \leq -\inf_{x\in\overline{\Gamma}} I(x)$$

For iid real valued random variables $X_1, X_2, \ldots$, we write $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ and we denote the moment generating function (of X_i) by $M(t)$. Let μ denote the law of X_1 and let μ_n denote the law of S_n.

Definition 5.30. The *logarithmic moment generating function* associated with the law μ is defined as

$$\Lambda(\lambda) = \log M(\lambda) = \log \mathbb{E}\left[e^{\lambda X_1}\right].$$

Definition 5.31. The *Fenchel-Legendre transform* of $\Lambda(\lambda)$ is

$$\Lambda^*(x) = \sup_{\lambda\in\mathbb{R}}\{\lambda x - \Lambda(\lambda)\}.$$

5.10.2 Cramér's Theorem

Theorem 5.32 (Cramér). *When $X_i \in \mathbb{R}$, the sequence of measures $\{\mu_n\}$ satisfies the Large Deviations Principle with the convex rate function $\Lambda^*(\cdot)$, namely*

1. *For any closed set $F \subset \mathbb{R}$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x\in F} \Lambda^*(x). \tag{5.9}$$

2. *For any open set $G \subset \mathbb{R}$,*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_{x\in G} \Lambda^*(x).$$

Proof. See Dembo and Zeitouni [1], Theorem 2.2.3. □

Remark 5.33. One can show that (5.9) can be strengthened to the statement, that for all n

$$\mu_n(F) \leq 2e^{-n\inf_{x\in F}\Lambda^*(x)}. \tag{5.10}$$

Equation (5.10) will be the crucial result that we use later to estimate exceedance probabilities of frequency distributions. Concretely, we get exponential bounds if $\inf_{x\in F}\Lambda^*(x) > 0$. We will therefore further investigate the properties of this expression:

Proposition 5.34. *The rate function $\Lambda^*(\cdot)$ is a convex, non-negative function satisfying $\Lambda^*(\mathbb{E}[X_1]) = 0$. Furthermore, it is an increasing function on $[\mathbb{E}[X_1], \infty)$, and a decreasing function on $(-\infty, \mathbb{E}[X_1]]$.*

Proof. This result including a proof can be found in MIT [4], Proposition 1 a), notation slightly adapted. □

5.11 Convergence for arbitrary frequencies with expectation $2 - \epsilon$

The crucial part in the proof of Theorem 5.23 was the estimation of the probability $\mathbb{P}[M^n \leq N^n]$ by the exponential term e^{-nk}. We achieved this by using the properties of the Skellam distribution. It would be very convenient if we could show exponential bounds as well for other distributions. It turns out that Cramér's Theorem is the required tool to achieve our goal: Suppose N_i has a frequency distribution with expectation $2 - \epsilon$ and finite moment generating function in a non-empty interval (this condition is always satisfied for Poisson, Binomial and Negative Binomial frequencies). Then we can write

$$\mathbb{P}[M^n \leq N^n] = \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}(M_i - N_i) \leq 0\right].$$

If we define $X_i := M_i - N_i$, $F = \mathbb{R}^-$ and μ_n the law of $\frac{1}{n}\sum_{i=1}^n X_i$ as in Theorem 5.32, we have $\mathbb{E}[X_i] = 2-(2-\epsilon) = \epsilon > 0$ and $\mathbb{P}[M^n \leq N^n] = \mu_n(F)$. In particular, note that $\mathbb{E}[X_i] = \epsilon \notin F$. By Proposition 5.34, we have $\Lambda^*(\epsilon) = 0$ and Λ^* is convex, so it is enough to show that Λ^* is not zero on an interval containing ϵ and 0, i.e. we need to check that the infimum of Λ^* on $F = \mathbb{R}^-$ is positive. To do so, one can compute Λ^* explicitly, which is in most cases a very cumbersome task, why numerical methods are a good option.

Finally, we get by Theorem 5.32 and Remark 5.33 an exponential bound

$$\mathbb{P}[M^n \leq N^n] \leq 2e^{-nk}, \tag{5.11}$$

for $k = \inf_{x\in\mathbb{R}^-} \Lambda^*(x) > 0$, and therefore the proof of Theorem 5.23 works for all frequency distributions which have finite moment generating function on a non-empty open interval. We can summarize these findings in a

Theorem 5.35. *Let N_k be iid frequencies with $\mathbb{E}[N_k] = 2 - \epsilon$ for $0 < \epsilon < 2$ and finite moment generating function on a non-empty interval. Then*

$$|\mathbb{E}[S^n] - \mathbb{E}[T^n]| \to 0 \text{ for } n \to \infty.$$

More precisely, we obtain the following rate of convergence:

$$|\mathbb{E}[S^n] - \mathbb{E}[T^n]| \leq 4(2-\epsilon)n \cdot \mathbb{E}[X_1] \cdot e^{-nk},$$

for

$$k = \inf_{x\in\mathbb{R}^-} \Lambda^*(x),$$

with Λ^* as defined above.

Proof. The inequality $\mathbb{P}[M^n \leq N^n] \leq 2e^{-nk}$ was established in equation (5.11). The remainder of the proof is completely analogous to the proof of Theorem 5.23. Note that we get a slightly different rate of convergence, since we estimate $\mathbb{P}[M^n \leq N^n]$ with $2e^{-nk}$. □

As before, convergence of T^n implies convergence of T:

Corollary 5.36. *With the same setting as in Theorem 5.35, we have*

$$\mathbb{E}[T] \to \mathbb{E}[S], \text{ for } n \to \infty.$$

Proof. The argument is exactly the same as in Corollary 5.24. □

Example 5.37. Let $N_i \sim \text{NegBin}(\beta = 2, r = 0.8)$, i.e. $\mathbb{E}[N_i] = 1.6$ and $\epsilon = 0.4$ with our notation. Moreover, we set as above $X_i := M_i - N_i$ with $M_i \overset{\text{iid}}{\sim} \text{Poisson}(2)$ independent of all N_i. Then

$$\begin{aligned} M_X(\lambda) &= M_{M-N}(\lambda) = M_M(\lambda) \cdot M_N(-\lambda) \\ &= e^{2(e^\lambda - 1)} \cdot \left(\frac{1}{1 + \beta(1 - e^{-\lambda})}\right)^r \end{aligned}$$

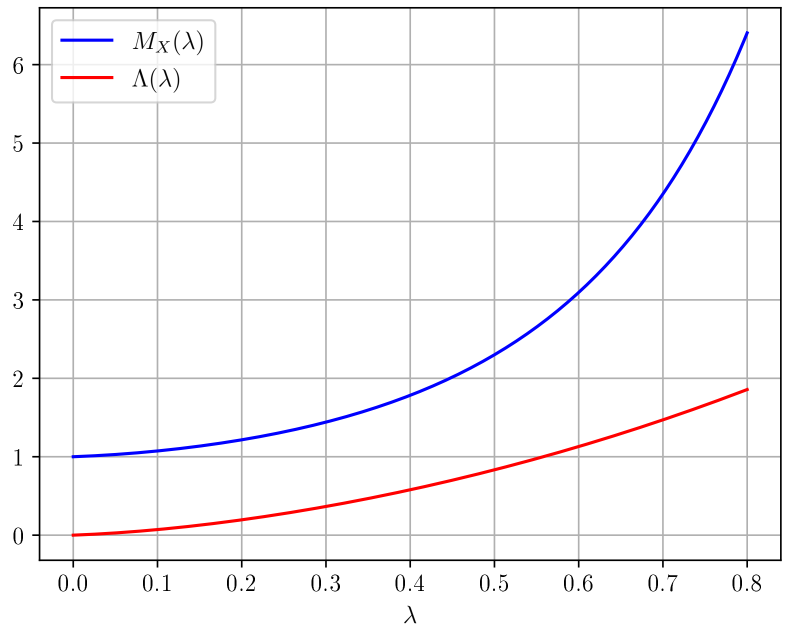


Figure 5.15: M_X and Λ

In Figure 5.15, we see plots of M_X and Λ. In order to compute $\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}}\{\lambda x - \Lambda(\lambda)\}$, we need to solve an optimization problem for every x of interest. We used the Nelder-Mead algorithm and obtained the curve in Figure 5.16 for $0 \leq x \leq 0.8$.

First of all, the picture suggests that Λ^* is indeed convex and attains its minimum in $\mathbb{E}[N_i] = \epsilon \overset{\text{here}}{=} 0.4$ as Proposition 5.34 showed. In particular, Nelder-Mead yields

$$\Lambda^*(0) \approx 0.011$$

and thus again by Proposition 5.34:

$$k = \inf_{x \in F} \Lambda^*(x) = \inf_{x \in \mathbb{R}^-} \Lambda^*(x) > 0.$$

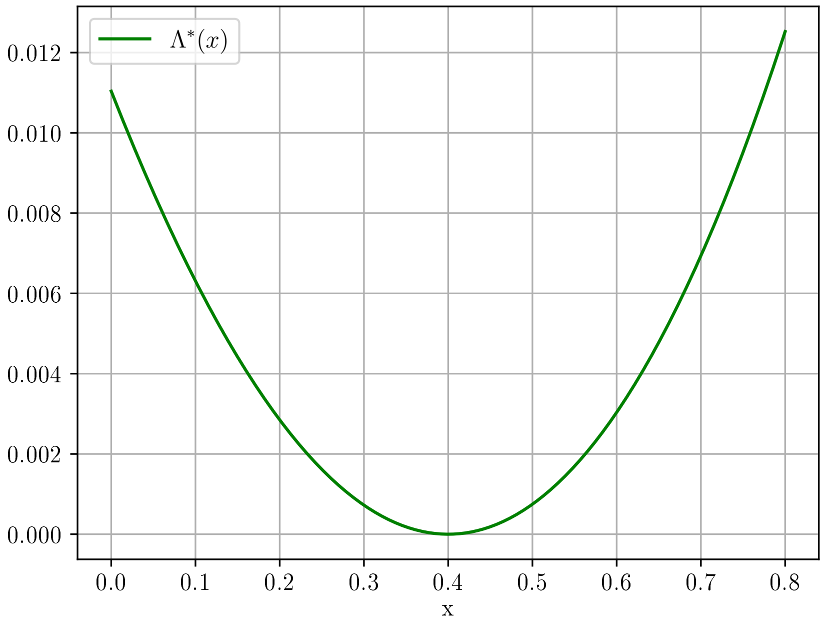


Figure 5.16: Λ^*

5.12 Variance analysis of $(2-\epsilon)$ models

So far, we investigated expectations under Transformation 5.5 with different assumptions on the initial frequencies N_k. In Chapter 4, we measured the quality of transformations with the variance and that is what we want to do here as well. We stay in the setting of expected frequencies smaller than 2.

Theorem 5.38. *Let S^n and T^n be as in Transformation 5.5 and assume that $\mathbb{E}[N_1] = 2-\epsilon$ for $0 < \epsilon < 2$. Then*

1. $Var(S_1) = \mathbb{E}[N_1] \cdot Var(X_1) + Var(N_1) \cdot \mathbb{E}[X_1]^2$.
2. $Var(T_1) \longrightarrow \mathbb{E}[N_1] \cdot (Var(X_1) + \mathbb{E}[X_1]^2)$ *for* $n \to \infty$.

In particular, we have

$$Var(T_1) \longrightarrow Var(S_1) + \mathbb{E}[X_1]^2 \cdot (\mathbb{E}[N_1] - Var(N_1)) \quad \text{for } n \to \infty.$$

Remark 5.39. This statement is very similar to the results which we entountered in Chapter 4 (see e.g. Theorem 4.14). Here, we do not have an equality in the second point, but a convergence. This finding suggests that Transfomation 4.13 can be seen as limiting case of Transformation 5.5.

Proof. The first point is just the second part of Wald's Theorem 3.21. For the second point, we write (assuming the variance exists)

$$\mathrm{Var}(T_1) = \mathbb{E}[T_1^2] - \mathbb{E}[T_1]^2. \tag{5.12}$$

We already know the limiting behaviour of $\mathbb{E}[T_1]^2$ from Corollary 5.24, so we can focus on the other term. In the next step, we change to conditional expectations via the tower property. We use again Cramér's Theorem and equation (5.11).

$$\begin{aligned}\mathbb{E}[T_1^2] = &\underbrace{\mathbb{P}[M^n \leq N^n]}_{\leq 2e^{-nk}} \cdot \underbrace{\mathbb{E}[T_1^2|M^n \leq N^n]}_{=\mathcal{O}(n^2)} \\ &+ \underbrace{\mathbb{P}[M^n > N^n]}_{\to 1} \cdot \mathbb{E}[T_1^2|M^n > N^n],\end{aligned} \tag{5.13}$$

so we are left to investigate the expression $\mathbb{E}[T_1^2|M^n > N^n]$. Recall our convention $X_i = 0$ for $i > N^n$ and define B_j to be the event that $\Sigma(j) \leq N^n$, i.e. B_j means that Y_j is *not* an added zero loss. Analogously, let $B_{jk} = B_j \cap B_k$ for $j \neq k$. Note that

$$\mathbb{P}[B_j|N^n, M^n > 0] = \frac{N^n}{M^n} = \frac{\frac{1}{n}N^n}{\frac{1}{n}M^n} \xrightarrow{\text{a.s.}} \frac{2-\epsilon}{2} = 1 - \epsilon/2 \quad \text{for } n \to \infty.$$

Conditioning by $M^n > 0$ is not a big issue, as $\mathbb{P}[M^n = 0] \xrightarrow{n\to\infty} 0$. One can also show that $\mathbb{P}[B_{jk}] \xrightarrow{\text{a.s.}} (1-\epsilon/2)^2$ for $n \to \infty$. This is however a bit more involved, since $\Sigma(j)$ and $\Sigma(k)$ are not independent. We leave the details to the interested reader. For a better readability, we will occasionally omit conditioning in the formulas which follow.

Conditionally, given $M^n > N^n$ and M_1, we have

$$\begin{aligned}\mathbb{E}[T_1^2] &= \mathbb{E}\left[\left(\sum_{j=1}^{M_1} Y_j\right)^2\right] = \mathbb{E}\left[\sum_{j=1}^{M_1}\sum_{k=1}^{M_1} Y_j Y_k\right] = \mathbb{E}\left[\sum_{j=1}^{M_1}\sum_{k=1}^{M_1} X^{\Sigma(j)} X^{\Sigma(k)}\right] \\ &= \mathbb{E}\left[\sum_{j=1}^{M_1}\left(X^{\Sigma(j)}\right)^2\right] + \mathbb{E}\left[\sum_{j\neq k} X^{\Sigma(j)} X^{\Sigma(k)}\right] \\ &= \mathbb{E}\left[\sum_{j=1}^{M_1}\left(X^{\Sigma(j)}\right)^2 \mathbb{1}_{\{B_j\}}\right] + \mathbb{E}\left[\sum_{j\neq k} X^{\Sigma(j)} X^{\Sigma(k)} \mathbb{1}_{\{B_{jk}\}}\right] \\ &\overset{\text{a.s.}}{\to} \sum_{j=1}^{M_1} \underbrace{\mathbb{P}[B_j]}_{=(1-\epsilon/2)} \underbrace{\mathbb{E}\left[\left(X^{\Sigma(j)}\right)^2 \middle| B_j\right]}_{=\mathrm{Var}(X_1)+\mathbb{E}[X_1]^2} + \sum_{j\neq k} \underbrace{\mathbb{P}[B_{jk}]}_{(1-\epsilon/2)^2} \mathbb{E}\left[X^{\Sigma(j)} X^{\Sigma(k)} \middle| B_{jk}\right].\end{aligned} \tag{5.14}$$

Moreover, we compute

$$\mathbb{E}\left[X^{\Sigma(j)} X^{\Sigma(k)} \middle| B_{jk}\right] = \underbrace{\mathrm{Cov}\left(X^{\Sigma(j)} X^{\Sigma(k)} \middle| B_{jk}\right)}_{=0} + \mathbb{E}[X_1]^2;$$

Since we consider two different indices j and k, such that both $\Sigma(j)$ and $\Sigma(k)$ are $\leq N^n$, the two random variables are independent and thus their covariance is zero. Note that the first sum in equation (5.14) has M_1 summands, while the second one has $M_1^2 - M_1$ summands. So we have in total, since $M_1 \sim \text{Poisson}(2)$, conditionally given $M^n > N^n$

$$\begin{aligned}\mathbb{E}[T_1^2|M^n > N^n] &\overset{\text{a.s.}}{\to} \underbrace{\mathbb{E}[M_1]}_{=2}(1-\epsilon/2)\left(\text{Var}(X_1) + \mathbb{E}[X_1]^2\right) \\ &+ \underbrace{\mathbb{E}[M_1^2 - M_1]}_{=4}(1-\epsilon/2)^2 \cdot \mathbb{E}[X_1]^2 \\ &= (2-\epsilon) \cdot [\text{Var}(X_1) + (3-\epsilon) \cdot \mathbb{E}[X_1]^2]. \end{aligned} \tag{5.15}$$

As we have seen earlier in Corollary 5.24, $\mathbb{E}[T_1]$ converges to $\mathbb{E}[S_1]$ and therefore

$$E[T_1]^2 \to \mathbb{E}[S_1]^2 = (2-\epsilon)^2 \cdot \mathbb{E}[X_1]^2, \quad \text{for } n \to \infty. \tag{5.16}$$

If we put equations (5.12), (5.13), (5.15) and (5.16) together, we receive

$$\text{Var}(T_1) \to (2-\epsilon) \cdot (\text{Var}(X_1) + \mathbb{E}[X_1]^2) = \mathbb{E}[N_1] \cdot (\text{Var}(X_1) + \mathbb{E}[X_1]^2)$$

and

$$\text{Var}(T_1) \to \text{Var}(S_1) + \mathbb{E}[X_1]^2 \cdot (\mathbb{E}[N_1] - \text{Var}(N_1)) \quad \text{for } n \to \infty,$$

as desired. □

5.13 Connection between Chapter 4 and 5

As already pointed out in Remark 5.39 the results of Theorem 5.38 and Theorem 4.14 are very similar. This is actually no coincidence, so let us elaborate why there is a connection. We stay in the setting where $\mathbb{E}[N_1] \leq \mathbb{E}[M_1] = 2$.

First of all, we have a look at the cdf of Y_j. Let $t \in \mathbb{R}$. Since $X^{\Sigma(j)}$ has the same distribution as X_1 conditionally given $\Sigma(j) \leq N^n$ and $X^{\Sigma(j)} = 0$ conditionally given $\Sigma(j) > N^n$ (Convention 5.4), we receive

$$\begin{aligned}\mathbb{P}[Y_j \leq t] &= \mathbb{P}[X^{\Sigma(j)} \leq t] \\ &= \mathbb{P}[\Sigma(j) \leq N^n] \cdot \mathbb{P}[X_1 \leq t] + \mathbb{P}[\Sigma(j) > N^n] \cdot \mathbb{P}[0 \leq t] \\ &= \frac{N^n}{M^n} \cdot F_X(t) + \frac{M^n - N^n}{M^n} \cdot \mathbb{1}_{\{t \geq 0\}}, \end{aligned} \tag{5.17}$$

since the probability $\Sigma(j) \leq N^n$ is equal to the proportion of non-zero losses $\frac{N^n}{M^n}$.

This is however not a very clean notation, as the expression for the probability $\mathbb{P}[Y_j \leq t]$ includes other random variables. One can interpret it as follows: Once M^n and N^n are known, i.e. when we consider concrete realizations of these two random variables, then $\mathbb{P}[Y_j \leq t]$ can be expressed with the (known) values of these two random variables. Alternatively, one can also write

$$\mathbb{P}[Y_j \leq t | M^n, N^n] = \frac{N^n}{M^n} \cdot F_X(t) + \frac{M^n - N^n}{M^n} \cdot \mathbb{1}_{\{t \geq 0\}} \tag{5.18}$$

to obtain more mathematical clarity. Another issue is a possible division by zero, as M^n can be zero. However, we will study equation (5.18) only for big values n and therefore, this special case is not relevant for us (since $\mathbb{P}[M^n = 0] \xrightarrow{n\to\infty} 0$). Although equation (5.18) is the mathematically correct version, we work with equation (5.17) as it is more convenient for our purposes.

If we write again $\mathbb{E}[N_1] = 2 - \epsilon$ for $0 < \epsilon < 2$, then the expression of (5.17) has the following limiting behaviour

$$\begin{aligned} \mathbb{P}[Y_j \leq t] &= \frac{N^n}{M^n} \cdot F_X(t) + \frac{M^n - N^n}{M^n} \cdot \mathbb{1}_{\{t \geq 0\}} \\ &= \frac{\frac{1}{n} N^n}{\frac{1}{n} M^n} \cdot F_X(t) + \frac{\frac{1}{n}(M^n - N^n)}{\frac{1}{n} M^n} \cdot \mathbb{1}_{\{t \geq 0\}} \\ &\overset{\text{a.s.}}{\to} \frac{2-\epsilon}{2} \cdot F_X(t) + \frac{\epsilon}{2} \mathbb{1}_{\{t \geq 0\}} \\ &= \frac{\mathbb{E}[N_1]}{2} \cdot F_X(t) + \frac{2 - \mathbb{E}[N_1]}{2} \mathbb{1}_{\{t \geq 0\}} \\ &= G_X(t) \quad \text{for } n \to \infty. \end{aligned} \tag{5.19}$$

In summary, we see that the cdf of Y_j depends on M^n, which reproduces Proposition 5.14. However, as n increases, this dependence gradually vanishes due to the convergence of the involved random variables to constant values. Hence, the limit of the cdf of Y_j is completely independent of M^n, which means that the loss amount (for one year) in the synthetic model

$$T_1 = \sum_{j=1}^{M_1} Y_j$$

is *not* a frequency severity model, but it *converges* to a frequency severity model for $n \to \infty$ as the dependence of M^n and the severities dissolves.

Additionally, the cdf of Y_j converges (pointwise) to G_X. As $M_1 \sim \text{Poisson}(2)$, $\sum_{j=1}^{M_1} Y_j$ corresponds exactly to the transformed frequency-severity model of Transformation 4.13, so if n goes to infitity, Transformation 4.13 can be seen as the "limiting transformation" of Transformation 5.5. These arguments can be summarized in the following

Theorem 5.40. *Let*

$$S = \sum_{i=1}^{N} X_i$$

be a frequency-severity model with $\mathbb{E}[N] < 2$. *Then its approximation*

$$T_1 = \sum_{j=1}^{M_1} Y_j$$

as defined in Transformation 5.5 converges to the following frequency-severity model (as defined in Transformation 4.13)

$$R = \sum_{j=1}^{K} Z_j$$

for $K \sim \mathit{Poisson}(2)$, $Z_j \overset{iid}{\sim} G_X$ *and* $n \to \infty$ *in the sense that*

- $M_1,\ K \overset{iid}{\sim} \mathit{Poisson}(2)$.
- *The severities* Y_j *converge in distribution to* Z_j.

Moreover, we obtain

1. $\mathbb{E}[R] = \mathbb{E}[S]$,
2. $\mathit{Var}(R) = \mathit{Var}(S) + \mathbb{E}[X_1]^2 \cdot (\mathbb{E}[N] - \mathit{Var}(N)$,
3. $\mathbb{E}[T_1] \to \mathbb{E}[S]$ *for* $n \to \infty$,
4. $\mathit{Var}(T_1) \to \mathit{Var}(S) + \mathbb{E}[X_1]^2 \cdot (\mathbb{E}[N] - \mathit{Var}(N)$ *for* $n \to \infty$.

Proof. The convergence was established in equation (5.19). The first two points were shown in Theorem 4.14. The third point was proved in Corollary 5.36 and the last point stems from Theorem 5.38. □

Bibliography

[1] Dembo A. and Zeitouni O. *Large Deviations Techniques and Applications.* Springer, New York, second edition, 1998.

[2] Föllmer H., Künsch H., and Teichmann J. Skript zur vorlesung wahrscheinlichkeitsrechnung und statistik. Lecture notes, Version February, 2013.

[3] Skellam J.G. The frequency distribution of the difference between two poisson variates belonging to different populations. *Journal of the Royal Statistical Society*, 109(3):296, 1946.

[4] MIT. Large deviations theory. cramér's theorem. Lecture notes, available at https://ocw.mit.edu/search/ocwsearch.htm?q=cramers theorem, 2013.

[5] Wüthrich M.V. Non-life insurance: Mathematics and statistics. Lecture notes, available at SSRN: https://ssrn.com/ abstract=2319328, Version January 7, 2020.

[6] Arbenz P. and Antal P. Reinsurance analytics. Lecture notes, Version Decenber 22, 2020.

[7] Billingsley P. *Probability and Measure.* Wiley, New York, 3rd edition, 1995.

[8] Muldowney P., Ostaszewski K., and Wojdowski W. The darth vader rule. *Tatra Mountains Mathematical Publications*, page 5, 2012.

[9] SCOR. Natcat non-modeled methodology revision. Impact Analysis for Solvency II Model Change, 2021.

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

On the Transformation of Actuarial Loss Models into Synthetic NatCat Loss Tables

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	First name(s):
Müller	Simon Lukas

With my signature I confirm that
- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	Signature(s)
Zurich, 13.08.2021	S. Müller

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.