

# An Analytical Portfolio Credit Risk Model Based on the Extended Binomial Distribution

Sven Fischer

Department of Credit Management, Sparkasse Chemnitz, Chemnitz, Germany

Email: sven.fische@spk-chemnitz.de

**How to cite this paper:** Fischer, S. (2019). An Analytical Portfolio Credit Risk Model Based on the Extended Binomial Distribution. *Journal of Financial Risk Management*, 8, 177-191.

<https://doi.org/10.4236/jfrm.2019.83012>

**Received:** August 26, 2019

**Accepted:** September 23, 2019

**Published:** September 26, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

The binomial distribution describes the probability of the number of successes for a fixed number of identical independent experiments, each with binary out-put. In real life, practical applications like portfolio credit risk management trials are not identical and have different realization probabilities. In addition to the number, the quantitative impacts of the respective outputs are also important. There exist no complete model-side implementations for the expansion of the binomial distribution, especially not in the case of specific quantitative parameters up to now. Here, a solution of this issue is described by the extended binomial distribution. The key for solving the problem lies in the use of bijection between the elementary events of the binomial distribution and the digit sequences of binary numbers. Based on the extended binomial distribution, an analytical portfolio credit risk model is described. The binomial distribution approach minimizes the approximation error in modeling. In particular, the edges of the loss distribution can be determined in a realistic manner. This analytical portfolio credit risk model is especially predestined for management of risk concentrations and tail risks.

## Keywords

Credit Risk, Portfolio Management, Credit Portfolio Risk, Bernoulli Mixture Distribution, General Binomial Distribution, Weighted Distribution

## 1. Introduction

Many projects are composed of different partial projects. In general, the probabilities of success of the individual projects are not uniform. Furthermore, the individual partial projects have different weighting, in other words different values for the overall project. The aim of this study was to describe a distribution of the values of the overall project.

Abstracting specific values of the partial projects, the problem can be modeled by the general binomial distribution (Fisz, 1981), also called Poisson binomial distribution or by the Bernoulli mixture distribution (McNeil et al., 2005). In the case of further abstraction of identical probabilities of success, the experimental arrangement can be described by the binomial distribution. Also the Panjer (1981) recursion enables to determine the probability mass function and the cumulative distribution function of the binomial distribution for experimental arrangements with a large number of partial projects. However, the calculation of the Poisson binomial distribution is already very complex and only possible in the case of small experimental arrangements.

Since the turn of the millennium, the task has been taken up again in connection with the modeling of the loss estimates of loan portfolios. It is undisputable that the loss distribution of a heterogeneous loan portfolio can be described by the idea of binomial distribution. However, until now no way has been found to cope with the task directly. For solution simulation techniques were used primarily (KMV, 1997; J.P. Morgan, 1997; Wilson, 1998; McKinsey and Company, 1998). The only alternative of a direct calculation using Poisson approximations was shown by CSFB (1997) due to CreditRisk+.

This study describes for the first time an opportunity to calculate directly the weighted extended binomial distribution. The feature of this method is the exploitation of a bijection between the elementary events of the binomial distribution and the digit sequences of binary numbers. The description of a numerical implementation for the analytically exact calculation of the distribution function is unique for real qualitative and quantitative heterogeneous Bernoulli processes (see Section 2). In particular, the rigorous calculation of the range of the distribution is of practical value for the risk management in quantification of tail risks and the assessment of risk concentrations (see Sections 5 and 6).

Nevertheless, the use of the model is subjected to numerical limitations. In general, the number of sub-trials is limited. For approaches with a limited range of weights or uniform weights, the model restrictions can be significantly relativized (see Section 4).

Notations used in this paper were applied according to the general mathematical literature (see Appendix).

## 2. Extended Binomial Distribution

The simplest of discrete distributions is the Bernoulli distribution. Here, it only has to be checked whether a particular event  $X$  was successful  $X = 1$  or has failed  $X = 0$ . The probability of  $X = 1$  is  $P(X = 1) = p$  and of the complementary event  $X = 0$  it is  $P(X = 0) = 1 - p$ . In comparison to the Bernoulli distribution the binomial distribution is the hierarchical higher order distribution. Due to the binomial distribution random variables based on the so-called Bernoulli trial scheme are described. To do so,  $n$  identical independent trials of a Bernoulli distributed random variable are performed. The number of realizations of the Bernoulli distributed random variable, in which the successful event  $X$  occurs, in

each case describes a trial outcome of the binomial distributed random variable.

In this paper, an extension of the binomial distribution without the limitation of identical probabilities  $p$  of the trials is described. In order to be additionally able to weight the trials differently when required, they are enhanced by specific weighting parameters.

Bernoulli distributed random variables have only two experimental outcomes and can be represented in binary code. This enables to map the Bernoulli trial scheme for  $n$  trials one-to-one into a matrix with  $n$  columns (number of trials) and  $2^n$  different rows (number of possible combinations of the experimental outcomes) of elements zero and one. This matrix is denoted scenario matrix. The scenario matrix plays the central role in the description of the following distribution.

**Definition 1:**

Let  $X_j, j=1, \dots, n$  be independent Bernoulli distributed random variables with probabilities  $p_j$ . Here, denote  $X_j = 1$  the occurrence of the event and  $X_j = 0$  the complementary event with each individual probabilities  $P_j(X_j = 1) = p_j$  and  $P_j(X_j = 0) = 1 - p_j$ . The random variables  $X_j$  have finite weights  $w_j; |w_j| < \infty$  for all  $j = 1, \dots, n$ . The possible combinations of events of the  $j = 1, \dots, n$  random variables  $X_j$  are represented by the scenario matrix  $S \in \mathbf{R}^{2^n \times n}$  with components  $s_{ij} \in \{0, 1\}$ . The  $2^n$  rows of the scenario matrix are the digit sequences of binary numbers from 0 to  $2^n - 1$ .

Then determine

$$f_i = \prod_{j=1}^n \left( p_j^{s_{ij}} \cdot (1 - p_j)^{1 - s_{ij}} \right), \quad i = 1, \dots, 2^n \quad (1)$$

the individual probabilities and

$$d_i = \sum_{j=1}^n w_j \cdot s_{ij} \quad (2)$$

their quantitative expressions. The distribution described by function

$$F_X(t) = P(X < t) = \sum_{d_i < t} f_i = \sum_{d_i < t} \prod_{j=1}^n \left( p_j^{s_{ij}} \cdot (1 - p_j)^{1 - s_{ij}} \right) \quad (3)$$

is called the extended binomial distribution.

The definition is meaningful only if function (3) is a distribution function. To show this, the criteria of the following theorem should be checked.

**Theorem:** (Fisz, 1981; Gnedenko, 1987)

A real-valued function  $F(x)$  is a distribution function if and only if

- 1) The two conditions  $F(-\infty) = 0$  and  $F(+\infty) = 1$  satisfy,
- 2) It is monotonically non-decreasing and
- 3) It is left-continuous.

The distribution is described by a finite number of finite weights  $w_j$ . Since  $s_{ij} \in \{0, 1\}$ , the sum of the quantitative characteristics (2) is bounded below by

the sum of all negative weights  $B = \sum_{\substack{j=1 \\ w_j < 0}}^n w_j$ . Thus summations over  $f_i$  in

Equation (3) for all  $t < B$  are summations over the empty set. That means  $F(t) = 0$  for all  $t < B$  and in particular  $F(-\infty) = 0$ . Moreover, if all weights are non-negative,  $F(t)$  is identical zero for all  $t < 0$  and the condition  $F(-\infty) = 0$  is also satisfied.

The proof of  $F(+\infty) = 1$  is carried by complete induction. Therefore, in Equation (3) all single probabilities have to be sum up  $f_i, i = 1, \dots, 2^n$ . First of all, it must be examined that the statement apply to for the base case  $n = 1$

$$F(\infty) = p_1^{s_{11}} \cdot (1 - p_1)^{1-s_{11}} + p_1^{s_{21}} \cdot (1 - p_1)^{1-s_{21}}.$$

Without restricting the generality  $s_{11} = 1$  and  $s_{21} = 0$ . This implies

$$F(\infty) = p_1^1 \cdot (1 - p_1)^0 + p_1^0 \cdot (1 - p_1)^1 = p_1 + (1 - p_1) = 1.$$

The induction assumption  $F(+\infty) = 1$  is satisfied for  $n = k$

$$F(\infty) = \sum_{i=1}^{2^k} \prod_{j=1}^k \left( p_j^{s_{ij}} \cdot (1 - p_j)^{1-s_{ij}} \right) = 1.$$

In the inductive step it is to show, that the statement apply to for  $n = k + 1$

$$F(\infty) = \sum_{i=1}^{2^{k+1}} \prod_{j=1}^{k+1} \left( p_j^{s_{ij}} \cdot (1 - p_j)^{1-s_{ij}} \right).$$

Therefore, the products under the summation sign are decomposed

$$F(\infty) = \sum_{i=1}^{2^{k+1}} \left( \left( p_{k+1}^{s_{ik+1}} \cdot (1 - p_{k+1})^{1-s_{ik+1}} \right) \cdot \prod_{j=1}^k \left( p_j^{s_{ij}} \cdot (1 - p_j)^{1-s_{ij}} \right) \right).$$

The coefficients  $s_{ik+1}$  consist of  $2^k$  coefficients that equal zero and  $2^k$  coefficients that equal one, each in pairs with identical factor below the product sign. The sum is decomposed by the expressions of the coefficients  $s_{ik+1}$

$$F(\infty) = \sum_{\substack{i=1 \\ s_{ik+1}=1}}^{2^k} \left( \left( p_{k+1}^{s_{ik+1}} \cdot (1 - p_{k+1})^{1-s_{ik+1}} \right) \cdot \prod_{j=1}^k \left( p_j^{s_{ij}} \cdot (1 - p_j)^{1-s_{ij}} \right) \right) \\ + \sum_{\substack{i=1 \\ s_{ik+1}=0}}^{2^k} \left( \left( p_{k+1}^{s_{ik+1}} \cdot (1 - p_{k+1})^{1-s_{ik+1}} \right) \cdot \prod_{j=1}^k \left( p_j^{s_{ij}} \cdot (1 - p_j)^{1-s_{ij}} \right) \right)$$

The coefficients  $s_{ik+1}$  are replaced by their concrete values one or zero. Then follows

$$F(\infty) = \sum_{i=1}^{2^k} \left( p_{k+1} \cdot \prod_{j=1}^k \left( p_j^{s_{ij}} \cdot (1 - p_j)^{1-s_{ij}} \right) \right) \\ + \sum_{i=1}^{2^k} \left( (1 - p_{k+1}) \cdot \prod_{j=1}^k \left( p_j^{s_{ij}} \cdot (1 - p_j)^{1-s_{ij}} \right) \right) \\ = \sum_{i=1}^{2^k} \left( (p_{k+1} + 1 - p_{k+1}) \cdot \prod_{j=1}^k \left( p_j^{s_{ij}} \cdot (1 - p_j)^{1-s_{ij}} \right) \right)$$

From the induction assumption it follows, that  $F(\infty) = 1$  for  $n = k + 1$ . So it was verified that the condition (1) of the theorem is accomplished.

By increasing  $t$  the number of summands in function (3) increases. To demonstrate the monotony it has to be shown, that all summands are non-negative. The summands are products themselves. The factors have the structure  $p_j^{s_{ij}} \cdot (1 - p_j)^{1-s_{ij}}$ . Distinction has to be made between two cases:

$s_{ij} = 0$ : then the factor reduces to  $(1 - p_j)$ , which is non-negative since  $0 \leq p_j \leq 1$ .

$s_{ij} = 1$ : then the factor only remaining  $p_j$ , which is non-negative.

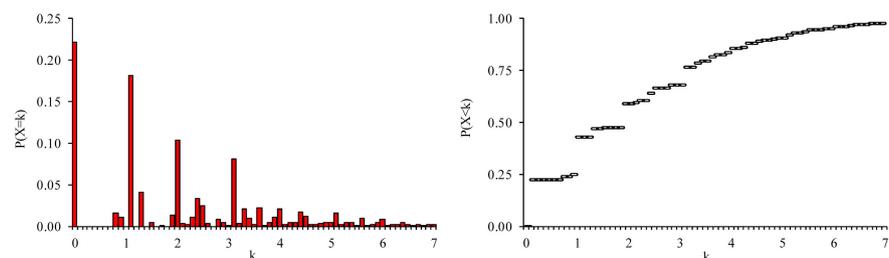
So the function (3) also satisfies the condition (2) of the theorem.

The function (3) is a jump continuous function and because of the strict inequality in  $F_X(t) = P(X < t)$  it is a left-continuous function. So the function (3) also satisfies the condition (3) of the sentence and it is a distribution function.

To illustrate the **Definition 1**, the probability mass function and the cumulative distribution function are considered for the following example in **Table 1** and **Figure 1**.

**Table 1.** Fictional example.

Trial-No.	Weight	Probability
$X_j$	$w_j$	$p_j$
1	4.0	0.067
2	3.9	0.020
3	3.1	0.013
4	2.5	0.100
5	2.3	0.044
6	2.0	0.300
7	1.5	0.013
8	1.3	0.009
9	1.5	0.004
10	1.3	0.150
11	1.1	0.450
12	0.9	0.045
13	0.8	0.067



**Figure 1.** Probability mass function and cumulative distribution function.

### 3. Moments and Characteristics of the Extended Binomial Distribution

The extended binomial distribution of a random variable  $X$  by **Definition 1** is a linear combination of independent Bernoulli distributed random variables  $X_j, j = 1, \dots, n$  with probabilities  $p_j$  and with weights  $w_j$  as linear coefficients. Based on these considerations, the moments of the extended binomial distribution are determined in the following.

The **expected value** of a random variable is generated by the expected value operator. The expected value operator is also linear (Fisz, 1981; Rényi, 1971). It means, for the sum of a finite number of random variables  $X_j, j = 1, \dots, n$  and real numbers  $\alpha_j, j = 1, \dots, n$  applies

$E(\alpha_1 \cdot X_1 + \dots + \alpha_n \cdot X_n) = \alpha_1 \cdot E(X_1) + \dots + \alpha_n \cdot E(X_n)$ . Bernoulli distributed random variables  $X_j$  have expected values  $E(X_j) = p_j$ . For the expected value of the extended binomial distributed random variable  $X = \sum_{j=1}^n w_j \cdot X_j$  follows

$$E(X) = \sum_{j=1}^n w_j \cdot p_j.$$

The **variance** of a sum of independent random variables  $X_j$  and real numbers  $\alpha_j, j = 1, \dots, n$  is given by

$D^2(\alpha_1 \cdot X_1 + \dots + \alpha_n \cdot X_n) = \alpha_1^2 \cdot D^2(X_1) + \dots + \alpha_n^2 \cdot D^2(X_n)$  (Rényi, 1971). Bernoulli distributed random variables  $X_j$  have the variances

$D^2(X_j) = p_j \cdot (1 - p_j)$ . For the variance of the extended binomial distributed random variable  $X = \sum_{j=1}^n w_j \cdot X_j$  follows  $D^2(X) = \sum_{j=1}^n w_j^2 \cdot p_j \cdot (1 - p_j)$ .

The **probability generating function**  $G_X(z)$  of a discrete random variable  $X$  is defined by  $G_X(z) = \sum_{k=0}^{\infty} P(X=k) \cdot z^k = \sum_{k=0}^{\infty} p_k \cdot z^k$ . The probability generating

function for a  $\alpha$ -multiples of a random variable  $X$  is  $G_{\alpha \cdot X}(z) = G_X(z^\alpha)$  (Gribakin, 2002). Furthermore, the probability generating function of the sum of independent random variables  $X_j, j = 1, \dots, n$  applies to be equal to the product of their probability generating functions  $G_{X_1 + \dots + X_n}(z) = G_{X_1}(z) \cdot \dots \cdot G_{X_n}(z)$ . From both follows the probability generating function of a linear combination of independent random variables  $\alpha_1 \cdot X_1 + \dots + \alpha_n \cdot X_n$  is

$G_{\alpha_1 \cdot X_1 + \dots + \alpha_n \cdot X_n}(z) = G_{X_1}(z^{\alpha_1}) \cdot \dots \cdot G_{X_n}(z^{\alpha_n})$  (Gribakin, 2002). The probability generating function of a Bernoulli distributed random variable  $X_j$  is given by  $G_{X_j}(z) = 1 - p_j + p_j \cdot z$  (Rényi, 1971). For the probability generating function

of the extended binomial distributed random variable  $X = \sum_{j=1}^n w_j \cdot X_j$  follows

$$G_X(z) = \prod_{j=1}^n (1 - p_j + p_j \cdot z^{w_j}).$$

The **characteristic function**  $\varphi_X(t)$  of a random variable  $X$  is defined by the expected value  $\varphi_X(t) = E(e^{itX})$ . The characteristic function of the linear expression  $\alpha \cdot X + \beta$  is  $\varphi_{\alpha \cdot X + \beta}(t) = e^{i\beta t} \cdot \varphi(\alpha \cdot t)$ . Furthermore, the characteristic

function of the sum of independent random variables  $X_j, j = 1, \dots, n$  applies to be equal to the product of their characteristic functions

$\varphi_{X_1+\dots+X_n}(t) = \varphi_{X_1}(t) \cdots \varphi_{X_n}(t)$ . From both follows the characteristic function of a linear combination of independent random variables  $\alpha_1 \cdot X_1 + \dots + \alpha_n \cdot X_n$  is

$\varphi_{\alpha_1 \cdot X_1 + \dots + \alpha_n \cdot X_n}(t) = \varphi_{X_1}(\alpha_1 \cdot t) \cdots \varphi_{X_n}(\alpha_n \cdot t)$  (Lukacs, 1960). The characteristic function of Bernoulli distributed random variables  $X_j$  is given by

$\varphi_{X_j}(t) = 1 - p_j + p_j \cdot e^{it}$ . For the characteristic function of the extended binomial distributed random variable  $X = \sum_{j=1}^n w_j \cdot X_j$  follows

$$\varphi_X(t) = \prod_{j=1}^n (1 - p_j + p_j \cdot e^{i \cdot w_j \cdot t}).$$

#### 4. A Numerical Approach to Apply the Extended Binomial Distribution on Higher Number of Independent Trials

Computational effort for calculating the extended binomial distribution will be doubled for an additional trial. The limits of the computational feasibility are quickly achieved. Under the additional condition, that the weights are in approximately the same order of size, the computational effort can be reduced significantly. Then the extended binomial distribution also is numerically approximately applicable for problems with a large number of trials.

In **Definition 1** it is assumed, that the trials  $X_j$  are independent from each other. Under this assumption, the extended binomial distribution can be calculated for problems with a large number of trials, by:

- Splitting larger trials into partial tests,
- Determining the distribution functions and the probability for the partial tests separately and
- Finally, aggregating the distributions of the partial tests to the distribution of the complete problem successively.

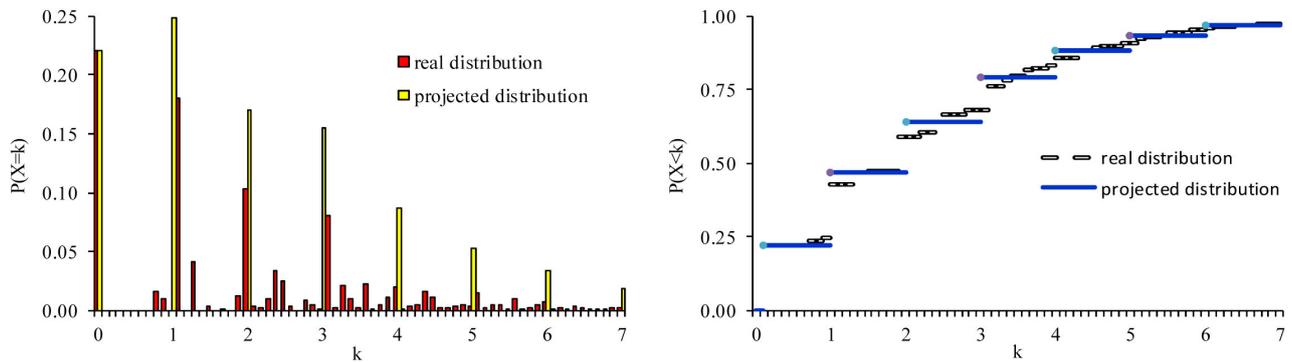
For aggregation the following calculus is used:

**Definition 2** (Smirnow & Dunin-Barkowski, 1969):

Let  $D \subseteq \mathbf{Z}$  be a discrete subset of the integers and  $P_1$  and  $P_2$  be two functions with  $P_i: D \rightarrow \mathbf{R}$  for  $i = 1, 2$ . Then  $(P_1 * P_2)(X = n) = \sum_{k \in D} P_1(X = k) \cdot P_2(X = n - k)$  is the **discrete convolution** of  $P_1$  and  $P_2$ .

In the computational implementation of the convolution of the probability functions of extended binomial distributed random variables is a practical problem. This results from the potentially large number of different quantitative manifestations (3). Moreover, in definition 1 it was not required that the weights  $w_j$  are integers. To apply the calculus of convolution computationally efficiently, the probability functions are approximated for aggregation. For this purpose the quantitative manifestations of the extended binomial distributions have to be projected onto reference points (**Figure 2**). The projection is done by rounding the quantitative manifestations of integer multiples of a given unit discretization  $U$ :

$$\tilde{d}_i = \text{round}(d_i; U) = [d_i/U] \cdot U = \left[ \frac{1}{U} \cdot \sum_{j=1}^n w_j \cdot s_{ij} \right] \cdot U$$



**Figure 2.** Real and projected extended binomial distribution.

Because of the projection, the discretized probability functions take positive values only at identical points with the same distance  $U$ .

Now, each of the two probability functions of two partial tests should be aggregated by convolution. The resulting probability functions again take positive values only on multiples of  $U$ . With the resulting probability functions the aggregation can be continued successively until the probability function for the complete problem was calculated.

The approximation error caused by the projection is low if the weights  $w_j$  are approximately in the same order of size. In this way the model of the extended binomial distribution is applicable for problems with a large number of trials. Up to this point, the meaning of “approximately the same order of size” was not specified. The key role is played by the discretization unit  $U$ .

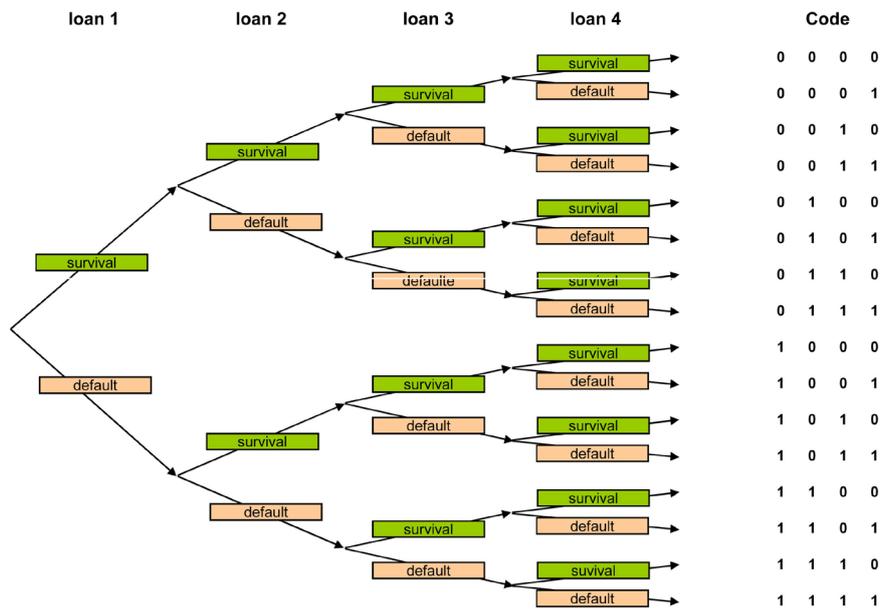
On the one hand the discretization unit  $U$  should not be larger than the smallest weight, since otherwise in the approximation the impact of the trials with smaller weights is neutralized. On the other hand the discretization unit  $U$  should not be smaller than a fraction of the greatest weights, since this would have negative effects on the performance of the convolution. Experience shows, that there are no significant performance impairments, if one percent of the largest weight is chosen as discretization unit. From both restrictions it can be derived that in the context above the weights are in approximately the same order of size, when the greatest weight is not significantly higher than one hundred times of the smallest.

## 5. Application of the Extended Binomial Distribution

A project in this sense is the investment in a loan portfolio. A loan portfolio consists of a certain number of loans. Each loan has a specific exposure and its own probability of default. An estimation of the expected portfolio loss and the loss distribution is required for the management of the portfolio.

To illustrate the problem, a portfolio of four loans is considered. Usually a tree structure is used to represent the elementary events. The characteristic of the extended binomial distribution, the bijection between the tree structure and the scenario matrix is shown in **Figure 3**.

Hereinafter concrete values are used in the example (**Table 2**).



**Figure 3.** Bijection between the tree structure and the scenario matrix.

**Table 2.** Example for a small portfolio.

Loan-No. $j$	1	2	3	4		
Exposure $w_j$	1234	9750	4698	2135	Probability	Loss
Probability of Default $p_j$	0.10	0.03	0.02	0.05	$f_i$	$d_i$
Scenario Matrix	0	0	0	0	0.8128	0
	0	0	0	1	0.0428	2135
	0	0	1	0	0.0166	4698
	0	0	1	1	0.0009	6833
	0	1	0	0	0.0251	9750
	0	1	0	1	0.0013	11,885
	0	1	1	0	0.0005	14,448
	0	1	1	1	0.0000	16,583
	1	0	0	0	0.0903	1234
	1	0	0	1	0.0048	3369
	1	0	1	0	0.0018	5932
	1	0	1	1	0.0001	8067
	1	1	0	0	0.0028	10,984
	1	1	0	1	0.0001	13,119
	1	1	1	0	0.0001	15,682
	1	1	1	1	0.0000	17,817

The loss distribution is determined according to Equation (3) with  $F_X(t) = P(X < t) = \sum_{d_i < t} f_i$ .

For portfolios with a larger number of loans, the portfolio is divided into partial portfolios as described in section 4. The loss distributions are computed for the sub-portfolios. While doing so, the losses are rounded to integer multiples of a discretization unit  $U$ . Next, the loss distributions of the partial portfolios are successively aggregated by convolution until the loss distribution of the complete portfolio is determined, see Figure 4.

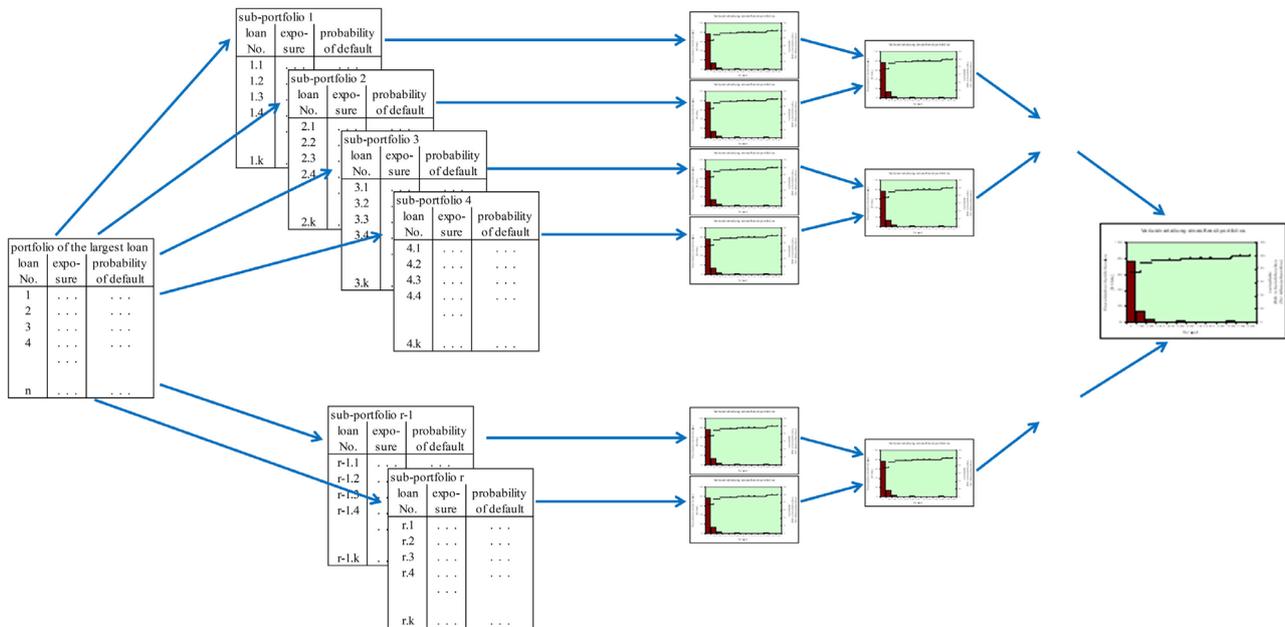


Figure 4. Strategy disassembling—calculation—aggregation.

By disassembling the problem and aggregating the partial results as described above it is possible to determine the loss distribution for portfolios consisting of a few hundred of loans. In this way, the numerical restrictions are relativized, but not completely eliminated. What does it mean in practical realization?

A partial portfolio of the largest loans is taken from the complete portfolio. For this sub-portfolio the loss distribution is determined by the extended binomial distribution. What should be done with the remaining portfolio?

In the remaining portfolio the largest loan account for only a few per thousand of the complete portfolio. If the remaining portfolio consists of a few loans, its influence on loss distribution of the complete portfolio is marginal. This is not the case if the remaining portfolio consists of many loans. Then because of the large number of small loans, the remaining portfolio is well diversified and heterogeneous in general. The loss distribution in such a portfolio can be well approximated by a Gaussian normal distribution. The parameters  $\mu$  and  $\sigma$  for the normal distribution approximation are (Fischer, 2012)

$$\mu = \sum_{j=1}^n w_j \cdot \mu_j = \sum_{j=1}^n w_j \cdot p_j$$

this correspond to expected loss and

$$\sigma^2 = \sum_{j=1}^n w_j^2 \cdot \sigma_j^2 = \sum_{j=1}^n w_j^2 \cdot p_j \cdot (1 - p_j).$$

The  $w_j$  are the exposures,  $p_j$  are the probabilities of default and  $r$  is the number of loans of the remaining portfolio.

Finally, the loss distribution determined by the generalized binomial distribution and the loss distribution of the remaining portfolio approximated by the Gaussian normal distribution are aggregated by convolution to the loss distribution of the complete portfolio.

## 6. Peripheral Modeling of Dependencies

In **Definition 1** it is assumed, that the trials  $X_j$  are independent from each other. But the default behaviors of the loans of a portfolio are not independent in reality. There are common dependencies on external risk factors for example, the economic situation. Because the dependencies cannot be mapped directly in the extended binomial distribution, the modeling of the dependencies of risk factors is transferred to the processing of input data.

The common dependencies of external risk factors are modeled on the basis of construction from [Gordy \(2002\)](#). First, the probabilities of default  $p_j$  are decomposed as follows:

$$p_j = [(1 - s_j) + s_j] \cdot p_j, \quad 0 \leq s_j \leq 1 \quad (4)$$

The parameters  $s_j$  are specific sensitivity factors. Subsequently, a risk factor  $M$ , for example for mapping the macroeconomic economic situation, is integrated in the formula (4). The decomposition of the probabilities of default is indicated by the index ( $M$ ):

$$p_j^{(M)} = [(1 - s_j) + M \cdot s_j] \cdot p_j, \quad 0 \leq M \leq \frac{1 + p_j \cdot (s_j - 1)}{p_j \cdot s_j} \quad (5)$$

Under normal conditions the parameter  $M$  has the value 1 (or 100%). That means under normal conditions apply  $p_j^{(M)} = p_j$  for all  $j$ . The effects of systematic changes are implemented in the model by modifying the parameter  $M$ , especially by  $M > 1$  for events of economic downturn and by  $M < 1$  for economic upturn phases.

Systematic changes affect the individual debtors individually ([Basel Committee on Bank Supervision, 2004](#)), Article 415, Sentence 3. The parameter  $s_j$  takes over the control of the individual sensitivity. If the creditworthiness of a debtor is above-average dependent on systematic influences, then  $s_j$  is greater 1 (or 100%). In the opposite case, if the creditworthiness of a debtor is lower-average dependent on systematic influences, then  $s_j$  is less than 1 (or 100%). In principle, the sensitivities can be individually assigned to the debtors. For practical reasons of data availability and from a calibration point of view a sectoral assignment will be the more practical, for example by industry sectors.

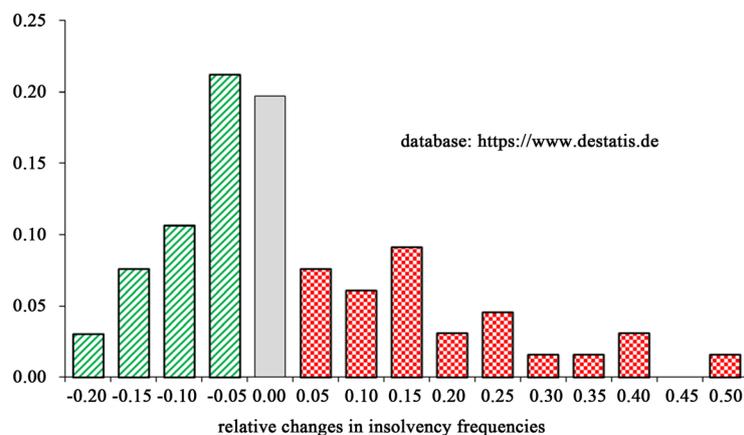
Adjusted decomposed probabilities of default for debtors reacting differently sensitive to changes in the systematic factor are shown schematically in the following **Table 3** for different scenarios of the systematic factor  $M$ . In this case, an a priori probability of default of 0.03 is assumed.

**Table 3.** Adjusted probabilities of default.

Scenario	Systematic parameter	Sensitivity factor	Adjusted probability of default
	$M$	$s_j$	$p_j^{(M)}$
Economic upturn	0.70	0.80	0.023
		1.00	0.021
		1.25	0.019
Normal case	1.00	0.80	0.030
		1.00	0.030
		1.25	0.030
Economic downturn	1.50	0.80	0.042
		1.00	0.045
		1.25	0.049

Similar to the one-factor model, the approach of decomposed probability of default can be further differentiated extended to several risk factors  $R_i, i = 1, \dots, k$  (Gordy, 2002). Through the synchronization of probabilities of default implied correlations respectively dependencies arise, which are observable in the real world.

Not yet considered is the parameter  $M$ . The  $M$  parameter is used for mapping clusters for relative changes in the economic situation. The economy itself is not measurable. Representative for changes in economy, changes in insolvency frequencies are used as a measurable quantity, see **Figure 5**. For risk considerations, this substitution should be opportune.



**Figure 5.** Distribution of the one year relative changes in insolvency frequencies from 1950 to 2016 in Germany (Statistisches Bundesamt, 2017).

For the individual clusters  $c$  (from  $M_1 = (1 - 0.50) = 0.50$  to  $M_a = (1 + 0.20) = 1.20$ ), the distributions of loss  $P_c$  are determined using the extended binomial distribution. Finally, the loss distributions are summarized weighted by the frequencies  $h_c$  of the clusters  $c$

$$P(X = x) = \sum_{c=1}^a h_c \cdot P_c(X = x).$$

From the aggregated distribution function and from the aggregated probability mass function of the loss of the complete loan portfolio, the known risk ratios value at risk or expected shortfall are determined (Albrecht, 2004).

## 7. Conclusion

A technique was developed to determine the probability mass function and the cumulative distribution function for the extended binomial distribution or for trials consisting of independent heterogeneous Bernoulli distributed single trials. Additionally, a numerical approach was described for approximating solutions of tasks with a larger number of trials.

The extended binomial distribution provides the foundation for a new analytical portfolio credit risk model. The new model expanded the set of analytical portfolio credit risk models, which were previously essentially represented only by the family of CreditRisk+ models. The analytical approach enables identical reproducibility of results. This in turn allows separate analysis with regard to individual risk factors or risk positions.

The approach of the extended binomial distribution allows the reduction of the approximation error. This has practical benefits in particular in determining the edges of the loss distribution. Hence the model is predestined for the identification of tail phenomena and for the management of risk concentrations.

## Conflicts of Interest

The author reports no conflicts of interest. The author alone is responsible for the content and writing of the paper.

## References

- Albrecht, P. (2004). Risk Measures. In *Encyclopedia of Actuarial Science*. New York: John Wiley & Sons. <https://doi.org/10.1002/9780470012505.tar039>
- Basel Committee on Bank Supervision (2004). *International Convergence of Capital Measurement and Capital Standards*. Bank for International Settlements. <https://www.bis.org/publ/bcbs107.pdf>
- CSFB (1997). *CreditRisk + A Credit Risk Management Framework*. London: Credit Suisse First Boston International.
- Fischer, S. (2012). *Ratio calculandi periculi-ein analytischer Ansatz zur Bestimmung der Verlustverteilung eines Kreditportfolios*. Dresden, Saxony, Germany: Technical University Dresden.
- Fisz, M. (1981). *Wahrscheinlichkeitsrechnung und mathematische Statistik*. Berlin: VEB Deutscher Verlag der Wissenschaften.

- Gnedenko, B. W. (1987). *Lehrbuch der Wahrscheinlichkeitsrechnung*. Berlin: Akademie Verlag.
- Gordy, M. B. (2002). *A Risk-Factor Model Foundation for Rating-Based Bank Capital Rules*. Federal Reserve System. <https://doi.org/10.2139/ssrn.361302>  
<https://www.federalreserve.gov/pubs/feds/2002/200255/200255pap.pdf>
- Gribakin, G. (2002). Chapter 3: Probability Generating Functions. In *Probability and Distribution Theory* (pp. 39-49). Belfast, Northern Ireland: Queen's University Belfast. <http://web.am.qub.ac.uk/users/g.gribakin/sor/Chap3.pdf>
- KMV (1997). *Modeling Default Risk*. San Francisco, CA: KMV LLC.
- Lukacs, E. (1960). *Characteristic Functions*. London: Griffin.
- McKinsey and Company (1998). *CreditPortfolioView™ Approach Documentation and User's Documentation*. Zurich: McKinsey and Company.
- McNeil, A., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: Princeton University Press.
- J.P. Morgan (1997). *CreditMetrics™—Technical Document*. New York: J.P. Morgan & Co. Incorporated.
- Panjer, H. (1981). Recursive Evaluation of a Family of Compound Distributions. *ASTIN Bulletin*, 12, 22-26. <https://doi.org/10.1017/S0515036100006796>  
<http://www.casact.org/library/astin/vol12no1/22.pdf>
- Rényi, A. (1971). *Wahrscheinlichkeitsrechnung*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Smirnow, N., & Dunin-Barkowski, I. W. (1969). *Mathematische Statistik in der Technik*. Berlin: VEB Deutscher Verlag der Wissenschaften.
- Statistisches Bundesamt (2017). *Insolvenzen*. Wiesbaden.
- Wilson, T. (1998). Portfolio Credit Risk. *FRBNY Economic Policy Review*, 4, 71-82. <https://doi.org/10.2139/ssrn.1028756>

---

## Appendix

### Notations

$X$	random variable
$P(X)$	probability mass function
$F_X(t)$	cumulative distribution function
$E(X)$	expected value of the random variable $X$
$D^2(X)$	variance of the random variable $X$
$G_X(z)$	probability generating function of the random variable $X$
$\varphi_X(t)$	characteristic function of the random variable $X$
$(P^*R)$	convolution on functions $P$ and $R$
$\mathbf{R}$	set of all real numbers
$\mathbf{Z}$	set of all integers
$\mathbf{R}^n$	$n$ -dimensional euclidean space
$\mathbf{R}^{m \times m}$	set of all real $m$ -by- $n$ matrices
$\Sigma$	sum sign for the summation operator
$\Pi$	product sign for the product operator
$\infty$	sign for infinity