Infectious Defaults

Mark Davis and Violet Lo
Tokyo-Mitsubishi International plc
6 Broadgate
London EC2M 2AA

This version: 13 July 1999

1 Introduction and Background

Figure 1

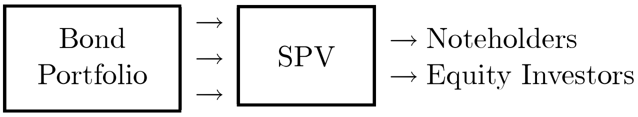


Figure 1 shows the structure of a typical CBO (Colleteralized Bond Obligation) transaction. The central box labelled SPV denotes a "special purpose vehicle", i.e. a company which is set up for this one transaction and dissolved when the transaction terminates. On the closing date of the transaction the SPV accepts principal payments from investors and purchases a portfolio of perhaps 60 high yield bonds. All subsequent payments made to investors are derived from income received from the bond portfolio. The investors are in two categories: Noteholders (contributing perhaps 85% of the total principal) and Equity Investors. The former receive a specified coupon (= interest rate) on their investment and are repaid their principal at maturity; they are first in priority of payment. Their coupon is considerably less than the average coupon of the bond portfolio. The Equity Investors are paid residual receipts, with no guaranteed coupon or guarantee of principal repayment. If the SPV actually receives all coupon and principal payments due from the bond portfolio, then the Noteholders will receive everything due to them and the Equity Investors will receive a very high return – perhaps in excess of 20%. If any of the collateral bonds defaults then the Equity Investors will suffer a loss of return, but the Noteholders are protected – up to a point – by the priority rule. (This is what justifies paying a lower coupon to them.) The Equity Investors are in effect making a leveraged investment in the high-yield portfolio.

Clearly, the performance of a CBO depends entirely on the default performance of the underlying high yield portfolio. The rating agencies (S&P, Moody's *et al.*) give credit ratings to companies or individual debt issues, and publish comprehensive statistics of default performance in the various rating categories. Reassuringly, it is indeed true that the historically experienced default rates are a monotone function of the rating category. The yield on lower-rated bonds is higher, to compensate investors for increased default risk. The idea of a CBO is to concentrate this risk, shifting the equity investors to the high-risk end of the risk/return curve. At the same time, the overall risk is reduced by diversification, in that the SPV invests in a medium-size portfolio of bonds rather than just one high-yield issue.

The most difficult question to answer in analysing CBOs is: how effective is this diversification? Obviously, this depends on "correlation". If the 60 bonds default independently, then it is an easy matter to predict the default performance of the portfolio from the rating agencies' default statistics. However, there is clearly the potential for "concentration risk": default of a whole subset of the bond portfolio might be triggered by some common event which affects all the issuers in that subset. The rating agencies deal with this question in various ways. The one we shall be investigating here is the "diversity score" approach used by Moody's. More specifically, we introduce a simple model designed to add correlation effects to the Binomial Expansion Technique (BET) used by Moody's [1] to evaluate CBO structures. In the BET, we assume we have a portfolio of n identical bonds, each of which defaults (over some fixed time horizon) with probability p, independently of the others. Thus if N denotes the number defaulting, N has the binomial distribution

$$P[N=k] = C_k^n p^k (1-p)^{n-k}, \tag{1}$$

where C_k^n is the binomial coefficient, $C_k^n = n!/k!(n-k)!$. The expected number of defaults is of course

$$EN = np. \tag{2}$$

In the Moody's analysis the number n is the "diversity score" for a portfolio containing $n' \geq n$ real bonds (see Section 4 below). The idea is that reducing the size to n is a way of accounting for correlation between defaults of the original bonds. In this paper we introduce an explicit correlation model, with a view to determining what the diversity score ought to be.

2 The Model

As above, we consider n identical bonds, but they are no longer independent. Let $(Z_i, i = 1, \dots, n)$ be random variables such that $Z_i = 1$ if bond i defaults and $Z_i = 0$ otherwise. Thus the number defaulting is

$$N = Z_1 + Z_2 + \cdots + Z_n. \tag{3}$$

In our model, the value of Z_i is determined as follows. For $i = 1, \ldots, n$ and $j = 1, \ldots, n$ with $j \neq i$ let X_i, Y_{ij} be independent Bernoulli random variables with

$$\begin{aligned} P[X_i = 1] &= p \\ P[Y_{ij} = 1] &= q. \end{aligned}$$

Then

$$Z_i = X_i + (1 - X_i)\left(1 - \prod_{j \neq i}(1 - X_j Y_{ji})\right). \tag{4}$$

The idea here is that bond i may default 'directly' ($X_i = 1$), or may be 'infected' by default of bond j; if $X_j = 1$ then random variable Y_{ji} determines whether infection takes place or not. Indeed, the second term on the right of (4) is equal to 1 when bond i does not default directly but is infected by default of bond j, for at least one $j \neq i$.

Note that the random variables Z_i are *exchangeable* [2] in that the joint distribution of $(Z_{\kappa(1)}, \ldots, Z_{\kappa(n)})$ is the same as that of $(Z_1, \ldots, Z_n)$, for any permutation $\kappa(\cdot)$ of the indices.

3 Default Distribution

3.1 The Main Result

Let $F(n, k, p, q)$ denote the probability mass distribution of the random variable N, defined by (3) with Z_i given by (4), i.e.

$$F(n, k, p, q) = P[N = k]. \tag{5}$$

Theorem 1 *The distribution function F defined by (5) is given by*

$$F(n, k, p, q) = C^n_k \alpha^{pq}_{nk}, \tag{6}$$

where

$$\begin{aligned} \alpha^{pq}_{nk} &= p^k(1-p)^{n-k}(1-q)^{k(n-k)} \\ &\quad + \sum_{i=1}^{k-1} C^k_i p^i (1-p)^{n-i}(1-(1-q)^i)^{k-i}(1-q)^{i(n-k)}. \end{aligned}$$

The expected value is

$$E[N] = n\left(1 - (1-p)(1-pq)^{n-1}\right), \tag{7}$$

and the variance is

$$\mathrm{var}[N] = E[N] + n(n-1)\beta_n^{pq} - (E[N])^2 \tag{8}$$

where

$$\begin{aligned} \beta_n^{pq} &= p^2 + 2p(1-p)\left(1-(1-q)(1-pq)^{n-2}\right) \\ &\quad +(1-p)^2\left(1-2(1-pq)^{n-2}+(1-2pq+pq^2)^{n-2}\right). \end{aligned}$$

PROOF: The expression for α_{nk}^{pq} arises as follows. The event $(Z_1 = 1, \ldots, Z_k = 1, Z_{k+1} = 0, \ldots, Z_n = 0)$ can be achieved in various disjoint ways. Firstly, we may have $(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots X_n = 0$, and $Y_{ij} = 0, i = 1, \ldots, k, j = k+1, \ldots, n)$, i.e. bonds 1 to k default directly and do not infect bonds $k+1$ to n. On the other hand, bonds 1 to i (for some $i < k$) may default directly and infect the other bonds of the first k but none of the remainder. There are C_i^k ways in which i bonds out of the first k may default, each, by exchangeability, having the same probability. α_{nk}^{pq} is the sum of the probabilities of these disjoint events. The k bonds defaulting can be chosen in C_k^n ways, and again each arrangement has the same default probability. Hence (6). The expectation (7) follows directly from (3) and (4) since all the random variables on the right of (4) are independent. For the variance, note that $E[Z_iZ_j]$ does not depend on i or j; it is given by β_n^{pq} above, as can be calculated from (4). Now

$$\begin{aligned} E[N^2] &= E(Z_1 + Z_2 + \ldots + Z_n)^2 \\ &= nEZ_1^2 + n(n-1)EZ_1Z_2 \\ &= nEZ_1 + n(n-1)\beta_n^{pq}, \end{aligned}$$

from which (8) follows.◇

Note that expressions (6),(7) reduce to the corresponding expressions (1),(2) for the binomial distribution when $q = 0$. When $q = 1$, either nobody defaults (probability $(1-p)^n$) or everybody defaults (probability $(1-(1-p)^n)$). The expected value is $n(1-(1-p)^n)$.

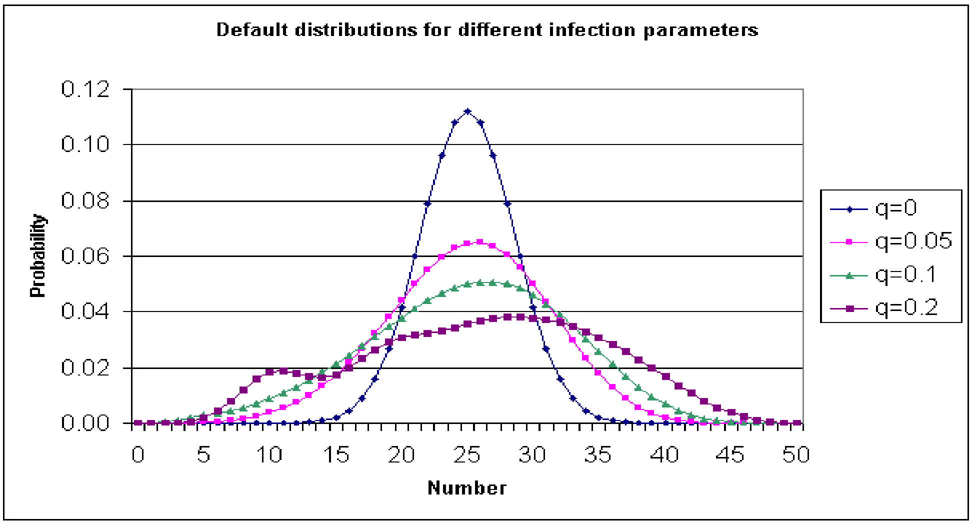
3.2 Example

As an example, consider a portfolio of $n = 50$ bonds with, initially, $p = 0.5$, so that the expected number of defaults if the bonds are independent (i.e. $q = 0$) is 25. To study the impact of infection, we compute the distribution for $q > 0$, *keeping the expected number of defaults constant*, which can be done by reducing p as q increases. Chart 1 shows the results for $q = 0, 0.05, 0.1, 0.2$. As can be seen, even a small infection probability has a dramatic effect in increasing the weight in the tails of the distribution. $q = 0$ corresponds to the binomial distribution. The distribution is still smooth and unimodal for small positive values of q but loses unimodality for larger q. Table 1 gives the standard deviations, and the values of p needed to maintain an expected value of 25.

Table 1

q	Implied p	Standard Deviation
0	0.5	3.54
0.05	0.194	6.05
0.1	0.116	7.70
0.2	0.064	10.32

Chart 1



4 Bonds in Different Industry Sectors

4.1 Calculating the Default Probabilities

Now suppose we have a portfolio of $n = n_1 + n_2 + \cdots + n_m$ bonds in m different industry sectors, n_i being the number of bonds in sector i. It is assumed that bonds in different industry sectors default independently, but there is infection within each industry sector. The probability of k_i defaults in sector i for $i = 1, \ldots, m$ is

$$\prod_{i=1}^{m} F(k_i, n_i, p_i, q_i), \tag{9}$$

where p_i, q_i are the infection model parameters for sector i. The probability of exactly k defaults for the portfolio as a whole is therefore

$$\sum_{a \in \mathcal{A}_m(k)} \prod_{i=1}^{m} F(k_i, n_i, p_i, q_i), \tag{10}$$

where $\mathcal{A}_m(k)$ is the set of arrangements $a = \{k_1, \ldots, k_m\}$, of k defaults in the m industry sectors. To calculate this, first note that

$$\mathcal{A}_m(k) = \bigcup_{j=j_{\min}}^{j_{\max}} \{\mathcal{A}_{m-1}(k-j), j\}. \tag{11}$$

The term in braces on the right denotes the set of all concatenations $\{k_1, \ldots, k_{m-1}, j\}$ for $\{k_1, \ldots, k_{m-1}\} \in \mathcal{A}_{m-1}(k-j)$. Denoting by

$$t_i = \sum_{j=1}^{i} n_j$$

the total number of bonds in sectors 1 to i, we have in (11),

$$\begin{aligned} j_{\min} &= \max(k - t_{m-1}, 0) \\ j_{\max} &= \min(k, n_m). \end{aligned}$$

Equation (11) shows that the number of elements $\text{card}(m,k)$ in $\mathcal{A}_m(k)$ is given recursively by

$$\text{card}(m,k) = \sum_{j=j_{\min}}^{j_{\max}} \text{card}(m-1, k-j) \tag{12}$$

with the boundary condition

$$\text{card}(1,k) = \text{card}(m,0) = 1. \tag{13}$$

A recursive program based on (12) and (13) will list all elements in $\mathcal{A}_m(k)$ and hence enable us to calculate the overall probability of exactly k defaults, using formula (10).

4.2 Example

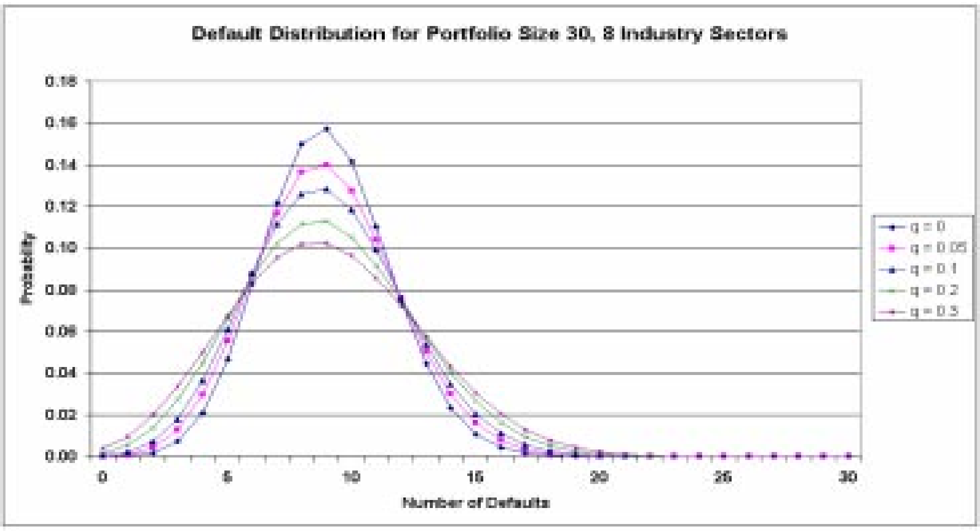
As an example, consider a portfolio of 30 bonds belonging to 8 different industry sectors, the numbers in each sector being $(1, 2, 2, 3, 4, 5, 6, 7)$. Suppose that, by considering the credit rating of the bonds, we have determined that the individual default probability for each bond is $p = 0.3$. We will assume that the infection parameter q is the same for each sector – this is just to simplify the presentation. We can now compute the implied value of "p" for each industry sector so that the marginal default probability is 0.3 for every bond. These values are shown in Table 2.

Table 2: Implied "p" for $q = 0.1, 0.2$

Sector Size	q = 0.1	q = 0.2
1	0.300	0.300
2	0.280	0.261
3	0.262	0.231
4	0.246	0.206
5	0.231	0.186
6	0.218	0.169
7	0.217	0.155

Having obtained these values, the distribution of the number of defaults can be obtained from formula (10). The resulting distributions are shown in Chart 2. As can be seen, the effect of infection is similar to the previous example: the variance and the weight in the tails are increased. Because of the "diversity", however, there is still a considerable amount of independence even at high infection rates.

Chart 2



In a typical CBO structure, the senior tranche experiences no losses up to some default threshold and suffers increasing loss as defaults increase beyond the threshold. To get some comparison between our technique and Moody's, let us consider the expected value of an option-type loss function as shown in Chart 3 (this captures the tail of the distribution, since the expected number of defaults is $9 = 30 \times 0.3$ whereas the loss is zero up to 15 defaults). Chart 4 shows the expected loss as infection increases.

Chart 3

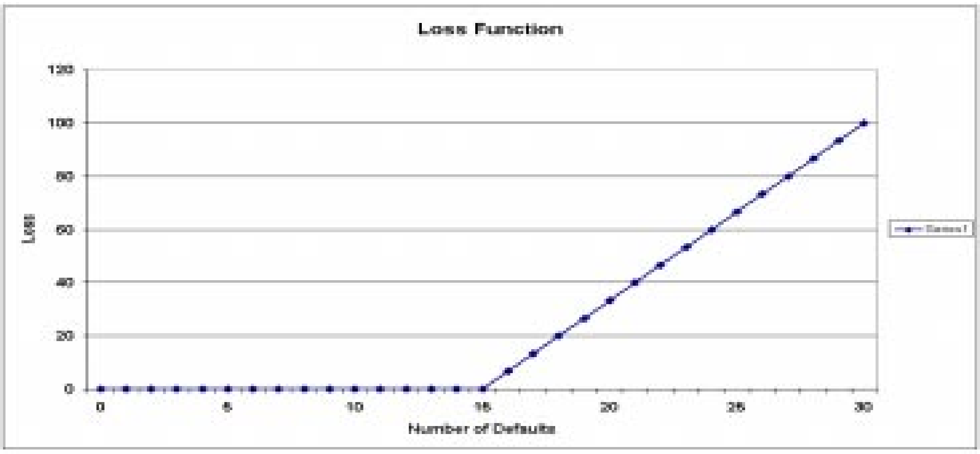


Chart 4

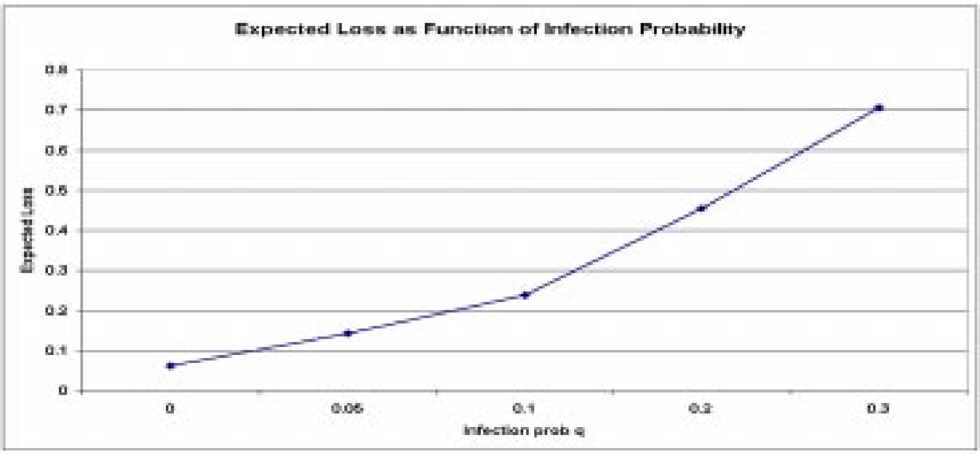
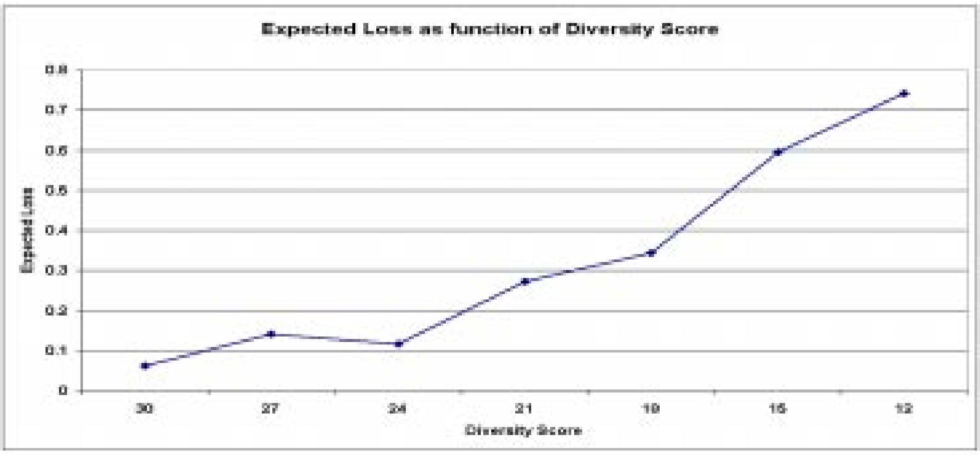


Chart 5



The Moody's technique is to replace the portfolio of 30 bonds by a portfolio of $d < 30$ independent bonds, each defaulting with probability 0.3 and having a notional amount $30/d$ (so that the expected defaulting amount is still 9). The number d is the diversity score. It is derived from the following table. For the specific portfolio in this example the diversity score is 17.

Table 3: Moody's Diversity Score Table [1]

Number of Firms in the Same Industry	Diversity Score
1	1.0
2	1.5
3	2.0
4	2.3
5	2.6
6	3.0
7	3.2
8	3.5
9	3.7
10	4.0
10 or more	Evaluated case by case

Chart 5 shows the expected loss as a function of diversity score: as anticipated, losses increase as the score is reduced. Charts 4 and 5 enable us to determine what infection parameter corresponds to a particular diversity score on an "equivalent expected loss" basis. In our example, the diversity score of 17 gives an expected loss of 0.470, which corresponds, in the infection model, to an infection parameter $q = 0.206$. The main outstanding problem, of course, is how to estimate q objectively from historical or market data.

5 Computation and Superboxes

In deference to traditional probabilistic thinking, let us call the bonds "balls" and the industry sectors "boxes". The number card (m, k) introduced in Section 3.1 is the number of ways of distributing k balls in m boxes of given capacities $(n_1, n_2, .., n_m)$. The probability of each one of these arrangements is then calculated by formula (9). The problem with this computation is that $\mathrm{card}(m, k)$ can be colossal even for moderate m. If the infection parameters (p_i, q_i) are different for each box, nothing much can be done to reduce this number, but if there are boxes of equal size and with the same (p_i, q_i) then considerable gains in efficiency can be made. For example, consider distributing 6 balls in 3 boxes of size 3. The 10 possible arrangements are

$$\begin{array}{ccc}
3 & 3 & 0 \\
3 & 2 & 1 \\
3 & 1 & 2 \\
3 & 0 & 3 \\
2 & 3 & 1 \\
2 & 2 & 2 \\
2 & 1 & 3 \\
1 & 3 & 2 \\
1 & 2 & 3 \\
0 & 3 & 3
\end{array}$$

There are however only 3 basic arrangements, namely

$$\begin{array}{ccclllll}
3 & 3 & 0 & \quad C_2^3 & C_1^1 & & = 3 \\
3 & 2 & 1 & \quad C_1^3 & C_1^2 & C_1^1 & = 6 \\
2 & 2 & 2 & \quad C_3^3 & & & = 1
\end{array}$$

All other arrangements are just permutations of the basic ones, the number of such permutations being given by the expressions on the right. All permutations of one basic arrangement have the same probability if the box parameters are equal. We can thus reduce the number of function calls from 10 to 3.

This suggests that instead of listing all arrangements $\mathcal{A}_m(k)$ we should think in terms of "superboxes", a superbox being a collection of boxes of equal size. For example a given portfolio might be as follows

size of box	1	2	3	4
number of boxes	2	7	5	3

The total capacity here is 43. We can list the arrangements of $k = k_1 + k_2 + k_3 + k_4$ balls in the 4 superboxes as in Section 3, and then compute the superbox probabilities using only the basic arrangements as described above.

The remaining problem is how to list the basic arrangements. These are simply the monotonic ones: if we have k balls to be distributed in m boxes of size n then a basic arrangement is a set of integers $a_1, .., a_m$ such that

$$\begin{array}{l}
n \geq a_1 \geq a_2 ... \geq a_m \geq 0 \\
\sum_1^m a_i = k
\end{array}$$

This is related to the classical problem in number theory of "partitions of an integer". A great deal of information can be found[1] in Chapter 6 of

---

[1] thanks to Jim Massey for this reference

Riordan [3] but there is no simple formula for the number of arrangements. However, we have an algorithm that demonstrably lists all of them. To describe this we need the following notation:

`LeftPack` (i, j, k) is the arrangement of k balls in boxes $(i, .., m)$, with at most j balls in each box, obtained by putting the maximum number of the remainder in box i, then the maximum number in box $i + 1$, etc. (i.e. "loading from the left"). For an arrangement a,

$$\texttt{Index}(a) = \begin{cases} 0 \text{ if } a_1 \leq a_m + 1 \\ \max\{i : a_i > a_{m+1}\} \text{ otherwise} \end{cases}$$

The algorithm for printing out all the basic arrangements is now as follows:

```
$a =$ LeftPack $(1, n, k)$
Print $a$
While (Index $(a) > 0$)
    $i =$ Index$(a)$
    $j = a_{i+1} + ..a_m$
    $a = \{a_1, \ldots, a_{i-1}, a_i - 1,$ LeftPack $(i + 1, a_i - 1, j + 1)$ $\}$
    Print $a$
```

As an example, for $m = 4, n = 3, k = 7$ we have

				Arrangement	**Index**
3	3	1	0	`LeftPack`(1,3,7)	2
3	2	2	0	{3,2, `LeftPack`(3,2,2) }	3
3	2	1	1	{3,2,1, `LeftPack`(4,1,1) }	1
2	2	2	1	{2, `LeftPack`(2,2,5) }	0

It is clear that the first column (i.e. the sequence of first elements a_1) decreases from the maximum to the minimum number possible for nay arrangement. For any fixed a_1, the sequence of elements a_2 again decreases from maximum to minimum. And so on. It follows that any feasible monotonic arrangement must appear in the list generated by the algorithm.

References

[1] Moody's Investment Services. The Binomial expansion method applied to CBO/CLO analysis, 1999.

[2] S.C. Port. *Theoretical Probability for Applications.* Wiley, New York, 1994.

[3] J. Riordan. *An Introduction to Combinatorial Analysis.* Wiley, New York,1958.