Approximating Independent Loss Distributions with an Adjusted Binomial Distribution

Dominic O’Kane
Affiliated Professor, EDHEC Business School
Abstract

We describe an approximation methodology for constructing independent loss distributions based on adjusting the binomial distribution. This method can handle both homogeneous and heterogeneous loss portfolios. We find that this simple algorithm provides an excellent fit to the exact distribution for a broad range of correlations and portfolio credit quality. For typical correlations, and homogeneous loss portfolios, the percentage error in the spread is typically 0.04% and usually less. It is at least 30 times faster than a full homogeneous loss recursion, and significantly faster for inhomogeneous loss portfolios.

EDHEC is one of the top five business schools in France owing to the high quality of its academic staff (100 permanent lecturers from France and abroad) and its privileged relationship with professionals that the school has been developing since its establishment in 1906. EDHEC Business School has decided to draw on its extensive knowledge of the professional environment and has therefore concentrated its research on themes that satisfy the needs of professionals.

EDHEC pursues an active research policy in the field of finance. Its “Risk and Asset Management Research Centre” carries out numerous research programs in the areas of asset allocation and risk management in both the traditional and alternative investment universes.

Copyright © 2007 Edhec
Approximating Independent Loss Distributions with an Adjusted Binomial Distribution

Dominic O’Kane*
EDHEC Risk and Asset Management Research Centre,
393-400 Promenade des Anglais,
0602 Nice Cedex 3, France

ABSTRACT

We describe an approximation methodology for constructing independent loss distributions based on adjusting the binomial distribution. This method can handle both homogeneous and heterogeneous loss portfolios. We find that this simple algorithm provides an excellent fit to the exact distribution for a broad range of correlations and portfolio credit quality. For typical correlations, and homogeneous loss portfolios, the percentage error in the spread is typically 0.04% and usually less. It is at least 30 times faster than a full homogeneous loss recursion, and significantly faster for inhomogeneous loss portfolios.

1. INTRODUCTION

The growth of the Single Tranche synthetic CDO (STCDO) market has been greatly assisted by the existence of pricing models which are flexible enough to fit market prices, and which are fast enough to be able to allow daily risk-management of a large correlation book. These models fall into the conditionally independent framework and capture the correlation of default times via a one-factor structure. By conditioning on the common factor, $Z_M$, the credits in the portfolio become conditionally independent and the loss distribution can be calculated via a number of well-known algorithms which include Fourier and recursion techniques. However, all of these algorithms have a computation time which is $O(N_C^2)$ where $N_C$ is the number of credits in the portfolio. Performance is also dependent on whether the credits in the portfolio incur the same loss on default. It is found that the calculation of the exact loss distribution for a portfolio with inhomogeneous losses can result in a significant increase in computing time when compared to the homogeneous loss portfolio.

However, the growing size of these correlation books and the challenge of hedging against the correlation skew have added to the volume of numbers to calculate and this has driven a search for faster approaches. One approach is to use approximate methods, provided the obtained accuracy is good enough. There is also need for a method which has a similar computation time for homogeneous and inhomogeneous losses. A portfolio which was initially homogeneous in its losses can easily become inhomogeneous if just one recovery rate is changed.

The large homogenous portfolio (LHP) approximation due to Vasicek\textsuperscript{5} is a well-known and widely used approximation. However, its accuracy is only $O(1/N_C)$ for homogeneous portfolios. The error can be much larger for portfolios where there is dispersion in the spreads - the standard case. The Gaussian approximation by Shelton\textsuperscript{6} in which the first two moments of the conditional loss distribution are fitted to a Gaussian density is a large improvement on LHP. However, its performance can vary, doing less well on high quality portfolios than on lower quality portfolios.

In this paper we set out a new approximation algorithm which is simple to implement, fast to calculate, and provides very good approximations to the exact pricing for homogenous and inhomogeneous reference portfolios. We believe that this algorithm is close to the optimal point in terms of accuracy versus speed.

\footnotesize{This version: 3rd April 2007}
2. CONSTRUCTING THE CONDITIONAL LOSS DISTRIBUTION

In a conditionally independent framework, the default probability of credit i in the portfolio is given by \( p_i(Z_M) \). There are then a number of algorithms which can be used to calculate the full portfolio loss distribution. For homogeneous loss\(^\dagger\) portfolios, recursion is almost surely the fastest algorithm for constructing the exact portfolio loss distribution.\(^4\) However, recursion requires \( O(N_C^2) \) operations. For heterogeneous loss portfolios with the same number of credits, the computation time can be significantly increased.

2.1. The Binomial Approximation

We first approximate the conditional loss distribution using the binomial distribution. The binomial distribution has the advantage that it is skewed, and can assume a number of shapes including unimodal or bimodal. We also know that when the conditional default probabilities in the portfolio are all equal, the approximation is exact.

To fit the exact distribution with a binomial, we require a single default probability and we choose the average \( p(Z_M) = \frac{1}{N_C} \sum_{i=1}^{N_C} p_i(Z_M) \).

We assume that the portfolio losses are homogeneous. The probability of \( n \) defaults out of \( N_C \), corresponding to a loss \( nu \), is then given by

\[
   f(n) = p(Z_M)^n (1 - p(Z_M))^{N_C - n} \frac{N_C!}{(N_C - n)! n!}.
\]

Rather than use this formula, which requires calls to the expensive factorial function, it is possible to calculate the conditional loss distribution using a recursive approach. The steps in the algorithm are therefore:

1. Calculate \( p(Z_M) \) using equation 1.
2. Start the recursion at the zero loss density using
   \[
   f(0) = (1 - p(Z_M))^{N_C}.
   \]
3. Loop over \( k = 1, \ldots, N_C \) using the following recursion rule
   \[
   f(k) = f(k - 1) \times \left( \frac{p(Z_M)}{1 - p(Z_M)} \right) \times \left( \frac{N_C - k + 1}{k} \right).
   \]

We can choose to terminate the recursion at the number of losses corresponding to the tranche upper strike. Note that if \( p(Z_M) \) is large, it is more numerically stable to start the recursion from the top of the loss distribution. We do this by setting \( f(N_C) = p(Z_M)^{N_C} \) and then invert the recursion relation by writing \( f(k - 1) \) in terms of \( f(k) \). This can be controlled by a simple branching statement at the start of the algorithm.

The advantage of this approximation is that it requires \( O(N_C) \) operations. However, it is not as accurate an approximation as if we were to fit the exact distribution with a Gaussian density. The reason is because the binomial has only one parameter and so we have only fitted the first moment of the exact distribution. For a Gaussian distribution, the mean and variance are free parameters which can easily be made to fit the mean and variance of the exact distribution.

\(^\dagger\)The default of each credit results in the same loss amount.
2.2. The Adjusted Binomial Approximation

To improve the quality of this approximation, we need to find a way to fit the variance of the exact loss distribution. The exact variance of the loss distribution is given by

\[ v_E = \frac{1}{N_C} \sum_{i=1}^{N_C} p_i(Z_M) (1 - p_i(Z_M)). \]  

(2)

The variance of the binomial distribution is

\[ v_A = \frac{1}{N_C} p(Z_M) (1 - p(Z_M)). \]  

(3)

For a portfolio with dispersion in the conditional default probabilities \( v_A > v_E \). To improve our approximation, we need to find a way to decrease the variance of the approximate loss distribution without changing its mean. We propose to do this using the following scheme:

1. Scale the probability density of the binomial loss distribution by a fixed multiplier \( \alpha \leq 1 \).
2. Add the removed probability mass \((1 - \alpha)\) at the mean \( m = p(Z_M)N_C \).

The practicalities are slightly more complicated since the discrete nature of the loss distribution means that the mean will almost always lie between two of the discrete loss probabilities. If this is the case, we share the \((1 - \alpha)\) of probability between the two loss points on either side of the mean. These are located at losses of \(lu\) and \((l + 1)u\) such that \(l \leq m \leq l + 1\). The probability of a loss \(lu\) gets an additional probability \(\epsilon_l\), while the probability of loss \((l + 1)u\) gets an additional probability \(\epsilon_{l+1}\). This presents us with 3 linear equations in 3 unknowns.

\[
\begin{align*}
\epsilon_l + \epsilon_{l+1} &= (1 - \alpha) \\
m &= am + \epsilon_l l + \epsilon_{l+1}(l + 1) \\
v_E N_C &= \alpha v_A N_C + (l - m)^2 \epsilon_l + (l + 1 - m)^2 \epsilon_{l+1}.
\end{align*}
\]

These can be solved exactly to give

\[
\begin{align*}
\alpha &= \frac{v_E N_C - (l + 1 - m)^2 - ((l - m)^2 - (l + 1 - m)^2)(l + 1 - m)}{v_A N_C - (l + 1 - m)^2 - ((l - m)^2 - (l + 1 - m)^2)(l + 1 - m)} \\
\epsilon_l &= (1 - \alpha)(l + 1 - m) \\
\epsilon_{l+1} &= 1 - \alpha - \epsilon_l.
\end{align*}
\]

(4)

The steps in the algorithm are listed below.

1. Calculate \(v_E\) and \(v_A\) using equations (2) and (3).
2. Determine \(l\) and \(l + 1\) which are the losses bracketing the mean loss \(m = p(Z_M)N_C\).
3. Calculate \(\alpha\) according to equation (4).
4. Adjust all weights such that \(f'(k) = \alpha f(k)\) for \(k = 0, \ldots, N_C\).
5. Calculate \(\epsilon_l\) and \(\epsilon_{l+1}\) according to equation (4).
6. Set \(f'(l) = f(l) + \epsilon_l\)
7. Set \(f'(l + 1) = f(l + 1) + \epsilon_{l+1}\)
Despite these additional steps, the number of operations is still $O(N_C)$.

### 2.3. Inhomogeneous Loss Portfolios

We can extend the algorithm to handle the case of inhomogeneous losses. If we define the average portfolio loss $L_{\text{Avg}} = \frac{1}{N_C} \sum_{i=1}^{N_C} (1 - R_i) F_i$

then we can write the inhomogeneous version of the binomial default probability

$$p(Z_M) = \frac{1}{N_C} \sum_{i=1}^{N_C} \frac{(1 - R_i) F_i}{L_{\text{Avg}}} p_i(Z_M).$$

This will ensure that the inhomogeneous portfolio expected loss is conserved by the binomial approximation. The approximate portfolio variance is then given as before by

$$v_E = \frac{1}{N_C^2} \sum_{i=1}^{N_C} \left( \frac{(1 - R_i) F_i}{L_{\text{Avg}}} \right)^2 p_i(Z_M) (1 - p_i(Z_M)).$$

This means that we can now apply the variance adjustment method described above to ensure that the binomial matches both the mean and variance of the exact inhomogeneous-loss loss distribution.

### 3. NUMERICAL RESULTS

For our analysis of approximate loss distribution methods we use the two most liquid Investment Grade (IG) and High Yield (HY) reference portfolios:

- A CDX North America Investment Grade portfolio of 125 credits with an average 5Y spread of 50bp.
- A CDX North-America High-Yield portfolio of 100 credits with an average 5Y spread of 549bp.

The spreads used for the individual credits in both portfolios are taken from real market data in March 2006. Both portfolios exhibit considerable spread dispersion. In the case of the HY portfolio the spread dispersion is highly skewed - we count 9 credits out of 100 with spreads in excess of 2000bp, and 7 with spreads less than 100bp. These are distributed around a mean spread of 549bp. For our tranche pricing, we choose the standard tranches linked to these respective indices. These are

- 0-3%, 3-7%, 7-10%, 10-15% and 15-30% for CDX NA IG.
- 0-10%, 10-15%, 15-25%, 25-35% and 35-100% for CDX NA HY.

We decompose these tranches into their base tranche components - a base tranche is simply an equity tranche. This follows the Base Correlation\(^7\) approach currently widely adopted in the credit markets for STCDOs on the standard indices. As a result, we consider a 0-3%, 0-7%, 0-10%, 0-15% and 0-30% tranche for CDX NA IG, and 0-10%, 0-15%, 0-25% and 0-35% for CDX NA HY. The base correlation model assumes a single homogeneous level of correlation $\rho$ for each base tranche on each portfolio. These levels of so-called base correlation change over time. We therefore compare the accuracy of our approximations across a range of base correlations, from 10% to 60%. We also want to compare homogeneous loss and inhomogeneous loss portfolios. We therefore consider these two cases separately.
Table 1. Comparison of approximation methods for the standard base tranches on the CDX NA IG index portfolio. The reference portfolio is assumed to be homogeneous with a common recovery rate of 40%. We highlight in bold the exact spreads and the spreads produced by the adjusted binomial approximation.

### 3.1. Homogeneous Loss Portfolio

We assume that all credits share a common recovery rate of 40%. We compare the 5Y breakeven spreads for the base tranches of both the CDX NA IG and the CDX NA HY portfolios using the LHP, the Gaussian and the Adjusted Binomial against the exact pricing coming from a full recursion. The pricing results for CDX NA IG are shown in table 1, and those for CDX NA HY in table 2.

The results for the CDX NA IG portfolio show that the worst approximation is provided by LHP. As this approximation ignores both the spread dispersion and granularity of the portfolio, this is to be expected. The Gaussian approximation performs better in most cases. The simple Binomial approximation is even better with a percentage error in the breakeven spread less than 0.5% in all cases, and usually much better. However, the adjusted binomial outperforms all of the others with a percentage spread error always less than 0.039% and usually better.

In the case of the HY portfolio we see once again that the LHP approximation is the worst. However, the Gaussian approximation performs much better than it did for the IG portfolio. We suspect that this is because the mean of the loss distribution is close to $5Y \times 549bp = 27.45\%$, and so the shape of the loss distribution is more Gaussian i.e. it is not squashed up close to the zero loss case as it is for the IG portfolio which has a loss distribution with a mean loss at 2.5%. The binomial does worse than the Gaussian. This is because the simple
Table 2. Comparison of approximation methods for the standard base tranches on the CDX NA HY index portfolio. The reference portfolio is assumed to be homogeneous with a common recovery rate of 40%.

Binomial does not fit the variance of the loss distribution. However, the adjusted binomial does better than all other approximations with a percentage error always less than 0.064% and often much smaller.

3.2. Inhomogeneous Loss Portfolio

We repeat these numerical experiments for a non-homogeneous loss portfolio. We assume the same facevalue for all credits, but allow them to have different recovery rates. We seek a distribution of recovery rates which is realistic, but which also tests the quality of the algorithm for non-homogeneous loss portfolios. For this reason we set 10 recovery rates equal to 10%, 10 to 20%, and 10 to 30%, with the remaining credits all having a recovery rate of 40%. The algorithm used to calculate the exact result was based on the one described in 4 with the use of a loss unit which is the greatest common divisor (GCD) of all losses. For the CDX NA HY the GCD was 0.10% and the number of loss units was 660. For the CDX NA IG portfolio, the GCD was 0.08% and the number of loss units was 810.

The results for the CDX NA IG portfolio are shown in table 3. Once again we see that the LHP approximation is the worst. The Gaussian approximation is better in almost all cases. However, the adjusted binomial approximation which handles non-homogeneous losses performs much better than the Gaussian. It seems to work as well as the Gaussian when the Gaussian works very well, however, it also performs very well even when the Gaussian performs poorly. The results for the non-homogeneous HY portfolio as shown in table 4. Once again the adjusted binomial approximation outperforms all of the others.

3.3. Computational speed

Rather than provide absolute measures of computational speed which are a function of the programming language, the choice of compiler, and the processor, we provide relative speed comparisons. We compare the time taken for the calculation of a tranche price, which typically involves several hundred calls to the conditional loss distribution

<table>
<thead>
<tr>
<th>$K_1-K_2$</th>
<th>$\rho$</th>
<th>Exact (bp)</th>
<th>Gauss (bp)</th>
<th>Error</th>
<th>LHP (bp)</th>
<th>Error</th>
<th>Binomial (bp)</th>
<th>Error</th>
<th>AdjBinom (bp)</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10%</td>
<td>10.00%</td>
<td>7553.52</td>
<td>7533.571</td>
<td>0.109%</td>
<td>8899.44</td>
<td>17.63%</td>
<td>7422.838</td>
<td>-1.883%</td>
<td></td>
<td>-0.014%</td>
</tr>
<tr>
<td>20.00%</td>
<td>6402.20</td>
<td>6411.820</td>
<td>0.150%</td>
<td>7006.44</td>
<td>9.44%</td>
<td></td>
<td>6293.798</td>
<td>-1.693%</td>
<td></td>
<td>-0.019%</td>
</tr>
<tr>
<td>30.00%</td>
<td>5551.58</td>
<td>5561.313</td>
<td>0.175%</td>
<td>5676.54</td>
<td>2.25%</td>
<td></td>
<td>5465.397</td>
<td>-1.552%</td>
<td></td>
<td>-0.022%</td>
</tr>
<tr>
<td>40.00%</td>
<td>4895.47</td>
<td>4904.906</td>
<td>0.193%</td>
<td>4686.63</td>
<td>-4.27%</td>
<td></td>
<td>4824.041</td>
<td>-1.495%</td>
<td></td>
<td>0.001%</td>
</tr>
<tr>
<td>50.00%</td>
<td>4368.69</td>
<td>4377.624</td>
<td>0.204%</td>
<td>3908.34</td>
<td>-10.54%</td>
<td></td>
<td>4307.554</td>
<td>-1.399%</td>
<td></td>
<td>-0.019%</td>
</tr>
<tr>
<td>60.00%</td>
<td>3932.70</td>
<td>3940.991</td>
<td>0.211%</td>
<td>3266.76</td>
<td>-19.93%</td>
<td></td>
<td>3879.063</td>
<td>-1.364%</td>
<td></td>
<td>-0.036%</td>
</tr>
<tr>
<td>0-15%</td>
<td>10.00%</td>
<td>4462.35</td>
<td>4463.952</td>
<td>0.036%</td>
<td>5756.27</td>
<td>29.00%</td>
<td>4417.419</td>
<td>-1.007%</td>
<td></td>
<td>0.019%</td>
</tr>
<tr>
<td>20.00%</td>
<td>3965.25</td>
<td>3967.466</td>
<td>0.056%</td>
<td>4802.29</td>
<td>21.11%</td>
<td></td>
<td>3932.662</td>
<td>-0.822%</td>
<td></td>
<td>-0.018%</td>
</tr>
<tr>
<td>30.00%</td>
<td>3577.99</td>
<td>3580.580</td>
<td>0.072%</td>
<td>497.97</td>
<td>14.53%</td>
<td></td>
<td>3552.130</td>
<td>-0.723%</td>
<td></td>
<td>-0.008%</td>
</tr>
<tr>
<td>40.00%</td>
<td>3256.21</td>
<td>3259.017</td>
<td>0.086%</td>
<td>3538.18</td>
<td>8.66%</td>
<td></td>
<td>3234.761</td>
<td>-1.693%</td>
<td></td>
<td>-0.028%</td>
</tr>
<tr>
<td>50.00%</td>
<td>2978.45</td>
<td>2980.347</td>
<td>0.097%</td>
<td>3069.35</td>
<td>3.05%</td>
<td></td>
<td>2960.213</td>
<td>-0.612%</td>
<td></td>
<td>-0.015%</td>
</tr>
<tr>
<td>60.00%</td>
<td>2731.88</td>
<td>2734.746</td>
<td>0.105%</td>
<td>2659.80</td>
<td>-2.64%</td>
<td></td>
<td>2716.102</td>
<td>-0.578%</td>
<td></td>
<td>0.064%</td>
</tr>
<tr>
<td>0-25%</td>
<td>10.00%</td>
<td>2113.94</td>
<td>2114.604</td>
<td>0.031%</td>
<td>3086.31</td>
<td>46.00%</td>
<td>2111.209</td>
<td>-0.129%</td>
<td></td>
<td>-0.010%</td>
</tr>
<tr>
<td>20.00%</td>
<td>2043.46</td>
<td>2044.269</td>
<td>0.040%</td>
<td>2812.85</td>
<td>37.65%</td>
<td></td>
<td>2040.044</td>
<td>-0.167%</td>
<td></td>
<td>0.001%</td>
</tr>
<tr>
<td>30.00%</td>
<td>1960.64</td>
<td>1961.569</td>
<td>0.056%</td>
<td>2576.61</td>
<td>31.42%</td>
<td></td>
<td>1957.151</td>
<td>-0.178%</td>
<td></td>
<td>0.006%</td>
</tr>
<tr>
<td>40.00%</td>
<td>1873.40</td>
<td>1874.419</td>
<td>0.054%</td>
<td>2362.56</td>
<td>26.11%</td>
<td></td>
<td>1870.007</td>
<td>-0.181%</td>
<td></td>
<td>-0.010%</td>
</tr>
<tr>
<td>50.00%</td>
<td>1784.16</td>
<td>1785.233</td>
<td>0.060%</td>
<td>2162.42</td>
<td>21.20%</td>
<td></td>
<td>1780.912</td>
<td>-0.182%</td>
<td></td>
<td>-0.021%</td>
</tr>
<tr>
<td>60.00%</td>
<td>1693.43</td>
<td>1694.530</td>
<td>0.065%</td>
<td>1970.02</td>
<td>16.33%</td>
<td></td>
<td>1690.300</td>
<td>-0.185%</td>
<td></td>
<td>-0.008%</td>
</tr>
<tr>
<td>0-35%</td>
<td>10.00%</td>
<td>1303.41</td>
<td>1303.607</td>
<td>0.016%</td>
<td>1917.66</td>
<td>47.13%</td>
<td>1303.351</td>
<td>-0.004%</td>
<td></td>
<td>-0.001%</td>
</tr>
<tr>
<td>20.00%</td>
<td>1298.44</td>
<td>1298.739</td>
<td>0.022%</td>
<td>1867.82</td>
<td>43.85%</td>
<td></td>
<td>1298.175</td>
<td>-0.021%</td>
<td></td>
<td>-0.001%</td>
</tr>
<tr>
<td>30.00%</td>
<td>1285.376</td>
<td>1285.948</td>
<td>0.029%</td>
<td>1803.39</td>
<td>40.05%</td>
<td></td>
<td>1285.090</td>
<td>-0.038%</td>
<td></td>
<td>-0.003%</td>
</tr>
<tr>
<td>40.00%</td>
<td>1265.398</td>
<td>1265.839</td>
<td>0.035%</td>
<td>1724.04</td>
<td>36.25%</td>
<td></td>
<td>1264.755</td>
<td>-0.051%</td>
<td></td>
<td>-0.004%</td>
</tr>
<tr>
<td>50.00%</td>
<td>1239.003</td>
<td>1239.498</td>
<td>0.040%</td>
<td>1641.38</td>
<td>32.48%</td>
<td></td>
<td>1238.243</td>
<td>-0.061%</td>
<td></td>
<td>-0.008%</td>
</tr>
<tr>
<td>60.00%</td>
<td>1207.077</td>
<td>1207.606</td>
<td>0.044%</td>
<td>1552.54</td>
<td>28.62%</td>
<td></td>
<td>1206.181</td>
<td>-0.074%</td>
<td></td>
<td>0.007%</td>
</tr>
</tbody>
</table>
Table 3. Comparison of approximation methods for the standard base tranches on the CDX NA IG index portfolio. The reference portfolio is assumed to be inhomogeneous. See text for details.

4. CONCLUSIONS

The Adjusted Binomial approximation provides an excellent fit to the pricing of standard tranches across portfolios with different average risk and across a broad range of correlations. In every case examined here it has been within the bid-offer of the tranche spread calculated using an exact model. It is as fast, and significantly more accurate than other simple approximations and we believe it is a realistic alternative to full recursion. It is also very simple to implement and is able to handle heterogeneous and homogeneous loss portfolios.

5. ACKNOWLEDGEMENTS

I would like to thank Lutz Schloegl for useful comments.
Table 4. Comparison of approximation methods for the standard base tranches on the CDX NA HY index portfolio. The reference portfolio is assumed to be inhomogeneous. See text for details.

REFERENCES

1. Finger, C., Conditional Approaches for CreditMetrics Portfolio Distributions, CreditMetrics Monitor, April 1999