Research Article

Sum of Bernoulli Mixtures: Beyond Conditional Independence

Taehan Bae¹ and Ian Iscoe²

¹ Department of Mathematics and Statistics, University of Regina, Regina, SK, Canada S4S 0A2

² Quantitative Research, Risk Analytics, IBM Corporation, 185 Spadina Avenue, Toronto, ON, Canada M5T 2C6

Correspondence should be addressed to Taehan Bae; taehan.bae@uregina.ca

Received 23 January 2014; Accepted 22 April 2014; Published 2 June 2014

Academic Editor: Aera Thavaneswaran

Copyright © 2014 T. Bae and I. Iscoe. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider the distribution of the sum of Bernoulli mixtures under a general dependence structure. The level of dependence is measured in terms of a limiting conditional correlation between two of the Bernoulli random variables. The conditioning event is that the mixing random variable is larger than a threshold and the limit is with respect to the threshold tending to one. The large-sample distribution of the empirical frequency and its use in approximating the risk measures, value at risk and conditional tail expectation, are presented for a new class of models which we call *double mixtures*. Several illustrative examples with a Beta mixing distribution, are given. As well, some data from the area of credit risk are fit with the models, and comparisons are made between the new models and also the classical Beta-binomial model.

1. Introduction

Since the U.S. subprime mortgage crisis which led the world economy into the global recession in the late 2000s, there has been serious criticism about the use of inappropriate dependent default models for portfolio credit risk (see Donnelly and Embrechts [1]). Specifically, the Gaussian copula model (Basel II Accord [2] and Li [3]) had become the industry standard model and has been widely used for both the pricing of portfolio credit derivatives such as collateralized debt obligations (CDOs) or mortgage backed securities and the credit risk management. Its inability to incorporate extreme tail dependence, however, resulted in a serious lack of default clustering, especially under stressful economic situations.

In the literature, several alternative approaches were proposed to cope with the issue of an insufficient level of dependency among defaults. As a simple extension of the Gaussian copula model, Andersen and Sidenius [4] and Burtshell et al. [5] considered stochastic asset correlation models. A more popular way of incorporating further dependence beyond the Gaussian copula model is the use of heavy tail copulas such as Archimedean copulas. Amongst many others, we mention Burtshell et al. [6] and Schönbucher and Schubert [7] for applications of copula models under the structural and the intensity-based credit risk model framework, respectively.

It is important to note that factor copula models under the typical conditional independence assumption have a close link with Bernoulli mixture models. Specifically, by de Finetti's theorem, there exists a common mixing random variable on which the default indicator random variables, in an exchangeable credit portfolio, are dependent. The mixing random variable is essentially the random probability of default which can be represented as a function of the common systematic factor in a factor copula model. Under the typical conditional independence assumption, however, the level of dependence between the default indicator random variables is controlled solely by the distributional property of the systematic factor, which can impose limitations for some applications where extreme levels of dependence are required. See Bluhm et al. [8], Cousin and Laurent [9], Frey and McNeil [10, 11], KMV [12], McNeil et al. [13], McNeil and Wendin [14], Moraux [15], and RiskMetrics Group [16] for more references on credit risk applications of the Bernoulli mixture models under the conditional independence modeling framework.

Formally, a Bernoulli mixture model is defined as follows. Let $(X_j)_{j=1}^n$ be a sequence of n identically distributed indicator random variables. We assume that each X_j , $j = 1, \dots, n$, follows the Bernoulli distribution with the common default probability P . The probability P is randomly drawn from

a distribution with cumulative distribution function (cdf) $F_p(p)$, $p \in (0, 1)$. At this point, the dependency structure for $(X_j)_{j=1}^n$ is still arbitrary.

In credit applications, the statistical properties of the sum of the n Bernoulli random variables are the main interest, and the following assumption plays an important role in alleviating computational difficulties: $\{X_j\}$ are conditionally independent given P . In this case, given P , the sum $S_n := \sum_{j=1}^n X_j$ follows the binomial distribution with mean nP and variance $nP(1 - P)$. Then, the unconditional probability mass function (pmf) is given as follows:

$$\begin{aligned} \mu_n^I(k) &:= \Pr[S_n = k] \\ &= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dF_p(p), \quad k = 0, \dots, n, \end{aligned} \quad (1)$$

where the superscript “I” on μ indicates the assumed conditional independence. The most commonly used mixing distribution in modeling the credit risk of a homogeneous portfolio such as a mortgage or credit card portfolio is the Beta distribution. This model is referred to as the Beta-binomial (β -I) model.

The present paper is concerned with the level of dependence incorporated in Bernoulli mixture models, especially under stress situations. Together with the asset correlation, the default correlation is used as the standard measure of dependence in portfolio credit risk models. In particular, we consider the behavior of the correlation between two arbitrary terms, as the random default probability P is conditioned to be larger than p , tending to 1. Specifically, we define the conditional default correlation between two terms, X_i and X_j , given $P > p$, $0 < p < 1$, as follows (Bae and Iscoe [17] used an analogous definition to study the asset correlations under stress for various single-factor credit risk models—some dynamic—while Kalkbrener and Packham [18] also studied asset correlations under stress, for static, normal variance mixture models. In both studies, in place of the conditioning on $\{P > p\}$, conditioning is done on an auxiliary risk factor which, for example, is related to a systematic factor):

$$\begin{aligned} \rho_{i,j}(p) &:= \text{Corr}[X_i, X_j \mid P > p] \\ &= \frac{\text{Cov}[X_i, X_j \mid P > p]}{\sqrt{\text{Var}[X_i \mid P > p]} \sqrt{\text{Var}[X_j \mid P > p]}} \\ &= \frac{\text{Cov}[X_i, X_j \mid P > p]}{\text{Var}[X_i \mid P > p]}, \end{aligned} \quad (2)$$

where

$$\begin{aligned} \text{Cov}[X_i, X_j \mid P > p] &= E[X_i X_j \mid P > p] - E[X_i \mid P > p] E[X_j \mid P > p] \\ &= E[X_i X_j \mid P > p] - \{E[X_i \mid P > p]\}^2, \end{aligned}$$

$$\text{Var}[X_i \mid P > p]$$

$$\begin{aligned} &= E[X_i^2 \mid P > p] - \{E[X_i \mid P > p]\}^2 \\ &= E[X_i \mid P > p] (1 - E[X_i \mid P > p]). \end{aligned} \quad (3)$$

Then, the limit of the conditional correlation as p tends to 1 (referred to as the *limiting correlation*) is defined as

$$\rho_{i,j} := \lim_{p \rightarrow 1} \rho_{i,j}(p), \quad (4)$$

provided the limit exists.

It is intuitive to expect that the correlations in (2) increase monotonically as the threshold p approaches 1. Unfortunately, this is not the case for the Beta-binomial model. In fact, the conditional correlation always converges to zero regardless of the mixing distribution, as long as conditional independence is postulated (Corollary 2). This type of model failure can introduce a significant bias in the measurement and management of risk in stress situations.

The objectives of the present paper are threefold: (i) to derive the relationship between the limiting correlation and the probability structure; (ii) to construct general Bernoulli mixture models with nontrivial limiting correlations; (iii) to demonstrate implications of the constructed models, for tail risk measures such as value at risk and conditional tail expectation.

The theoretical results for these objectives are presented in Section 2. We first examine the limiting behavior of correlations for Bernoulli mixture models in Section 2.1. Then, Section 2.2 introduces a general model framework—the class of double mixture models—which allows further positive dependence between entities, beyond conditional independence (Theorem 4). Section 2.3 is devoted to the third objective, and it considers the large-sample distribution of the sum of Bernoulli mixtures under the general model framework and its application to the approximation of risk measures is discussed. As a specific case, Section 3 considers a Beta Bernoulli mixture, and Section 4 provides the results for an example with specific parameter values. Section 5 demonstrates the fitting results of the double mixture models to a dataset from the area of credit risk. Finally, Section 6 concludes this paper with a summary. Technical details of some proofs are given in the appendices.

2. Main Results

2.1. Limiting Correlation for Bernoulli Mixtures. We denote by $a_m(p)$ the regular conditional probability, given $P = p$, that all terms $(X_i)_{i=1}^m$, of a fixed sequence of Bernoulli random variables, $(X_i)_{i=1}^n$, take on the value 1. (For this definition, we allow $n = \infty$.) Formally,

$$a_m(p) = \Pr[S_m = m \mid P = p], \quad (5)$$

where $S_m = \sum_{i=1}^m X_i$. Note that, under an assumption of conditional independence, $a_m(p) = p^m$.

The following hypothesis on the properties of F_p in a deleted neighbourhood of $p = 1$ will be part of the assumptions in Theorem 1 and its Corollary 2 below.

Hypothesis H. F_P is absolutely continuous in a neighbourhood of $p = 1$, with probability density function (pdf) f_P thereon, such that

- (i) $f_P(p) > 0$;
- (ii) f_P is continuous in a deleted neighbourhood of $p = 1$;
- (iii) $(1 - p)f_P(p)/(1 - F_P(p))$ converges to a finite limit as p tends to 1.

The significance of $H(i)$ is clear: it is required for the conditioning on $\{P > p\}$ to be well defined. Parts $H(ii)$ and $H(iii)$ are technical assumptions which play a role in the proof of Theorem 1. Part (iii) of Hypothesis H is satisfied in many cases of practical interest, where f_P has a simple asymptotic behaviour near $p = 1$. For example, $H(iii)$ is satisfied if for some positive (finite) constant c and some (finite) constants $b > -1$ and d (the relation “ $x \sim y$ ” denotes asymptotic equivalence, in the sense that the ratio x/y tends to 1, as $p \rightarrow 1$)

$$f_P(p) \sim c(1 - p)^b (\log[1 - p])^d, \quad \text{as } p \rightarrow 1. \quad (6)$$

Here is the verification that (6) implies that $H(iii)$ holds, in the case that $d \neq 0$. (The case $d = 0$ is similar but easier, so the details will be omitted.) Applying (6) in the first step and L'Hospital's rule in the second step, as p tends to 1, we have

$$\begin{aligned} & \frac{(1 - p)f_P(p)}{1 - F_P(p)} \\ & \sim \frac{c(1 - p)^{b+1}(\log[1 - p])^d}{1 - F_P(p)} \\ & \sim \left(-c(b + 1)(1 - p)^b (\log[1 - p])^d \right. \\ & \quad \left. - c(1 - p)^{b+1} d (\log[1 - p])^{d-1} (1 - p)^{-1} \right) \\ & \quad \times \left(-c(1 - p)^b (\log[1 - p])^d \right)^{-1} \\ & \sim b + 1 + d(\log[1 - p])^{-1} \sim b + 1, \end{aligned} \quad (7)$$

so $H(iii)$ is satisfied with the value of the limit being $b + 1$. It will be clear that the proof of Theorem 1 is easily generalized to allow the limit in $H(iii)$ to be infinite. However, it is not apparent, given the previous example, whether such a situation can actually occur. For this reason, we have taken the limit to be finite in the formulation of $H(iii)$.

Based on the conditional joint probability $a_2(p)$, we obtain the following result for the limiting correlation.

Theorem 1. Let X_i and X_j be two Bernoulli mixture random variables with correlation, $\rho_{i,j}(p)$, as in (2). Suppose that Hypothesis H holds. One further assumes that $\lim_{p \rightarrow 1} a_2(p) = 1$, $a_2(p)$ is differentiable for p in a deleted neighbourhood of 1, and $\lim_{p \rightarrow 1} a_2'(p)$ exists. Then, the limiting correlation in (4) exists and satisfies

$$\rho_{i,j} = 2 - \lim_{p \rightarrow 1} a_2'(p). \quad (8)$$

Proof. See the appendices. $\square$

In the context of a sequence (X_k) of Bernoulli mixture random variables, the notation, “ a_2 ,” suppresses the dependence on i and j ; more precisely, it should be written as $a_{2,i,j}$. In order to have a result which does not actually depend on i or j , an extra hypothesis on the Bernoulli mixtures must be imposed. We will return to this point in Section 2.2.

An interesting example is the case that both terms take on identical values (comonotonicity) (here, *comonotonicity* refers to the case that all the components (identically distributed) of a Bernoulli random vector coincide in value with probability one, conditional on P); that is, $a_2(p) = p$. Under Hypothesis H , Theorem 1 implies that the limiting correlation for this case is 1. The following corollary, another important example, states that the limiting correlation between two arbitrary terms from an identically distributed Bernoulli mixture sequence converges to zero, under the assumption of conditional independence.

Corollary 2. Suppose that two Bernoulli mixture random variables X_i and X_j are conditionally independent given the common default probability P . Then, under Hypothesis H , the limiting correlation in (4) exists and satisfies

$$\rho_{i,j} = \lim_{p \rightarrow 1} \rho_{i,j}(p) = 0. \quad (9)$$

Proof. Under the conditional independence assumption, $a_2(p) = p^2$. The desired result follows immediately from Theorem 1:

$$\rho_{i,j} = 2 - \lim_{p \rightarrow 1} a_2'(p) = 2 - 2 = 0. \quad (10)$$

$\square$

Remark 3. Results similar to those of Theorem 1 and Corollary 2 can also be given for the other extreme, conditional on $P < p$ with p tending to 0. One simply replaces, in Hypothesis H and Theorem 1, the phrase “neighbourhood of $p = 1$ ” with “neighbourhood of $p = 0$ ” and $H(iii)$ with “ $pf_P(p)/F_P(p)$ converges to a finite limit as p tends to 0;” then, in Theorem 1 and Corollary 2, one replaces “ $\lim_{p \rightarrow 1}$ ” with “ $\lim_{p \rightarrow 0}$.” (Note that $\lim_{p \rightarrow 0} a_2(p)$ is automatically 0 because $a_2(p) \leq \Pr[X_i = 1 \mid P = p] = p$.) However, in our applications, it is the case “ $p \rightarrow 1$ ” which is important, so we will not dwell further on the case “ $p \rightarrow 0$.”

2.2. Double Mixture Models. We now turn to the construction of a more general probability structure beyond conditional independence—a structure which reveals the fundamental role of the quantities, a_k . We require a stronger assumption on the sequence $(X_j)_{j=1}^n$ than their being just conditionally identically distributed, which is a statement about the 1-dimensional marginals. We require that $(X_j)_{j=1}^n$ be conditionally exchangeable, meaning that, under each $\Pr[\cdot \mid P = p]$, all permutations of $(X_1, X_2, \dots, X_n)$ have a common joint distribution which may depend on p . (Note that a sequence of n conditionally iid random variables is conditionally exchangeable, as is a comonotonic sequence.) In particular,

conditional exchangeability is sufficient to guarantee that $a_2(p)$, $\rho_{ij}(p)$, and hence the limiting correlation in Theorem 1, are independent of i and j .

We assume, for the remainder of the paper, that our sequence of n Bernoulli mixture random variables is conditionally exchangeable. The quantities $a_k(p)$, $1 \leq k \leq n$, then do not depend on the choice of subset of k variables. The probability that any k of the Bernoulli random variables take on the value 1 is given in the following theorem which shows that the pmf of S_n is essentially determined by the quantities $a_k(p)$, $k = 1, \dots, n$, which in turn can be specified quite generally as input to the model.

Theorem 4. Let $(X_j)_{j=1}^n$ be a sequence of conditionally exchangeable Bernoulli mixture random variables. Then, for $k = 0, \dots, n$,

$$f_n(k; p) := \Pr[S_n = k \mid P = p] = \binom{n}{k} g_n(k; p), \quad (11)$$

$$g_n(k; p) := a_k(p) - \binom{n-k}{1} a_{k+1}(p) + \binom{n-k}{2} a_{k+2}(p) + \dots + (-1)^{n-k} a_n(p), \quad (12)$$

where $a_0(p) \equiv 1$ and $a_1(p) = p$.

Conversely, let $\{\nu(\cdot; p) : p \in (0, 1)\}$ be a family of probability measures on the interval $[0, 1]$, such that $\int_0^1 r \nu(dr; p) = p$ and $\nu(A; p)$ is a Borel measurable function of p , for each Borel set, A . If $a_k(p)$ is defined to be $\int_0^1 r^k \nu(dr; p)$, then f_n , as defined by the right-hand sides of (11) and (12), is a valid pmf. Moreover, suppose that $(X_j)_{j=1}^n$ follows a Bernoulli mixtures distribution with conditional distribution

$$\Pr[(X_1, \dots, X_n) = (x_1, \dots, x_n) \mid P = p] = g_n(k; p), \quad (13)$$

for all $(x_1, \dots, x_n)$ which have precisely k components equal to 1 and the remaining $n-k$ components equal to 0. Then, $(X_j)_{j=1}^n$ is conditionally exchangeable; $\Pr[S_k = k \mid P = p] = a_k(p)$ for all $k = 1, \dots, n$; hence, S_n has conditional pmf given by (11) and (12). Furthermore,

$$g_n(k; p) = \int_0^1 r^k (1-r)^{n-k} \nu(dr; p), \quad k = 0, \dots, n; \quad (14)$$

therefore, for all $(x_1, \dots, x_n)$ which have precisely k components equal to 1 and the remaining $n-k$ components equal to 0,

$$\begin{aligned} \Pr[(X_1, \dots, X_n) = (x_1, \dots, x_n)] \\ = \int_0^1 r^k (1-r)^{n-k} \nu(dr; p) dF_P(p), \end{aligned} \quad (15)$$

and hence

$$\begin{aligned} \Pr[S_n = k] &= \int_0^1 \binom{n}{k} r^k (1-r)^{n-k} \nu(dr; p) dF_P(p), \\ &k = 0, \dots, n. \end{aligned} \quad (16)$$

Proof. See the appendices. $\square$

Remark 5. For a general characterization of multivariate Bernoulli random variables, see Sharakhmetov and Ibragimov [19]. With the exchangeability assumption, (12) can be deduced from Theorem 3 of the aforementioned paper.

Remark 6. Under the assumption of conditional independence, $f_n(k; p)$ reduces to the pmf of a binomial distribution (i.e., $a_k(p) = p^k$).

Remark 7. The converse part of the theorem is connected with the easy part of the classical Hausdorff moment problem and its application to the proof of de Finetti's theorem for sequences of exchangeable random variables (see, e.g., Chapter VII in Feller [20]).

Definition 8. Let $\nu(\cdot; p)$, $0 < p < 1$ be as in Theorem 4, and let $a_k(p) := \int_0^1 r^k \nu(dr; p)$, for $k = 0, 1, 2, \dots, n$. A Bernoulli mixture model with conditional distribution as specified in (14) (or equivalently, (12)) and (13) is called a *Bernoulli double mixture* model. It is determined by the following distributions: the family of measures $\nu(\cdot; p)$, $0 < p < 1$ and the mixing distribution, F_P , and it can be formally denoted as a F_P - $\nu(\cdot; p)$ Bernoulli double mixture model. Similarly, model (16), for the sum, S_n , of the F_P - $\nu(\cdot; p)$ Bernoulli double mixture random variables, can be formally referred to as a F_P - $\nu(\cdot; p)$ binomial double mixture model.

In concrete examples, we will employ some meaningful, worded description or an acronym, rather than through the formal notation of Definition 8. In addition, for the remainder of the paper, all considered models will be Bernoulli (or binomial) double mixture models; therefore, we may omit those words from the names of models, simply retaining the description of the building blocks, $\nu(\cdot; p)$ and F_P .

The following simple example illustrates the converse of the theorem.

Example 9 (ICM). We consider a weighted average of independence and comonotonicity. Specifically, for a specified weight $\omega \in [0, 1]$, the random n -vector $X \equiv (X_j)_{j=1}^n$ is a two-point mixture of a conditionally comonotonic n -vector of Bernoulli mixtures and a conditionally independent n -vector of Bernoulli mixtures (with common mixing distribution, F_P), with respective weights ω and $1 - \omega$. For such a model, the (conditional) joint probability, $a_k(p)$, $k = 2, \dots, n$, is expressed as the weighted average of the joint probability of the comonotonic case and that of independence case:

$$a_k(p) = \omega p + (1 - \omega) p^k, \quad k = 2, \dots, n. \quad (17)$$

In practice, the weight parameter $\omega \in [0, 1]$ must be estimated from data or be chosen exogenously.

This model (with F_P unspecified) will be denoted by the acronym ICM standing for independent comonotonic mixture (with everything being implicitly conditional). It

corresponds to the following choice for $\nu(\cdot; p)$, which is a double mixture of point masses:

$$\nu(\cdot; p) = \omega [p\delta_1 + (1-p)\delta_0] + (1-\omega)\delta_p. \quad (18)$$

Note that $\int_0^1 r^0 \nu(dr; p) = 1$ follows from the usual understanding that $0^0 = 1$.

Then, the limiting correlation is (cf. (4))

$$\rho_{i,j} = 2 - \lim_{p \rightarrow 1} a'_2(p) = \omega, \quad (19)$$

and the unconditional pmf of the sum, $\mu_n(k) := \Pr[S_n = k]$, is

$$\mu_n(k) = \begin{cases} \omega(1 - E[P]) + (1-\omega)\mu_n^I(0) \\ = \omega(1 - E[P]) \\ + (1-\omega)E[(1-P)^n], & k = 0 \\ (1-\omega)\mu_n^I(k), & k = 1, \dots, n-1 \\ \omega E[P] + (1-\omega)\mu_n^I(k) \\ = \omega E[P] + (1-\omega)E[P^n], & k = n, \end{cases} \quad (20)$$

with $\mu_n^I(k)$ given in (1). The mean and variance of the sum S_n can be easily obtained as follows:

$$E[S_n] = nE[P],$$

$$\begin{aligned} \text{Var}[S_n] &= \omega \{n^2 E[P] - n^2 (E[P])^2\} + (1-\omega) \\ &\times \{nE[P] - nE[P^2] + n^2 E[P^2] - n^2 (E[P])^2\}. \end{aligned} \quad (21)$$

Note that both the mean and variance are the weighted averages of those for the independent and comonotone cases. (In general, the k th moment is the weighted average of the k th moment of the independent and comonotone cases.)

2.3. Large-Sample Distribution of Empirical Frequency. The proportion of observed defaults to the total number of entities in a credit portfolio is of interest in practice. For example, the historical default rate for a homogenous credit portfolio is used to estimate the probability of default for a generic counterparty in the portfolio.

For a fixed n , let $L_n = S_n/n$ denote the empirical frequency, which can be considered as the percentage gross loss of a portfolio of n loans in equal dollar amounts.

The probability distribution of L_n is

$$F_{L_n}(x) := \Pr[L_n \leq x] = \Pr[S_n \leq nx], \quad x \geq 0. \quad (22)$$

For the general binomial double mixture model, based on a family of probability measures, $\nu(\cdot; p)$, and mixing distribution, F_P , we have the following result:

$$\begin{aligned} F_{L_\infty}(x) &\equiv \lim_{n \rightarrow \infty} F_{L_n}(x) \\ &= \begin{cases} \int_0^1 \nu(\{0\}; p) dF_P(p), & x = 0 \\ \int_0^1 \left[\nu([0, x]; p) + \frac{1}{2} \nu(\{x\}; p) \right] dF_P(p), & 0 < x < 1 \\ 1, & x = 1. \end{cases} \end{aligned} \quad (23)$$

Here is the proof of (22). The case $x = 1$ is trivial. For $0 \leq x < 1$, first note that by (14)

$$\begin{aligned} \Pr[S_n \leq nx \mid P = p] &= \sum_{k=0}^{[nx]} \binom{n}{k} g_n(k; p) \\ &= \int_0^1 \sum_{k=0}^{[nx]} \binom{n}{k} r^k (1-r)^{n-k} \nu(dr; p), \end{aligned} \quad (24)$$

where, for a nonnegative number y , $[y]$ represents the greatest integer less than or equal to y . The integrand is the cdf, at x , of S'_n/n where $S'_n \sim \text{Bin}(n, r)$; so, by the LLN, as $n \rightarrow \infty$, $S'_n/n \rightarrow r$, a.s. Therefore, for $0 < x < 1$,

$$\lim_{n \rightarrow \infty} \Pr \left[\frac{S'_n}{n} \leq x \right] = \mathbb{1}_{[0, x)}(r) + \frac{1}{2} \mathbb{1}_{\{x\}}(r), \quad (25)$$

where $\mathbb{1}_A$ denotes the indicator function of a set, A . (The second term, with the factor “1/2,” comes from an application of the CLT to $\Pr[S'_n - nr \leq 0]$, when $x = r$.) For $x = 0$,

$$\begin{aligned} \Pr[S_n = 0 \mid P = p] \\ = \int_0^1 (1-r)^n \nu(dr; p) \longrightarrow \nu(\{0\}; p), \quad \text{as } n \longrightarrow \infty. \end{aligned} \quad (26)$$

Integrating these two limiting results with respect to $\nu(dr; p) dF_P(p)$ yields the first two cases of (22). The interchange of limit and integrals is justified by the bounded convergence theorem.

Example 10 (ICM reprise). For the ω -weighted average of conditionally comonotonic and conditionally independent Bernoulli mixtures described in Example 9, we can easily obtain the following limiting distribution (this is the weighted average of the obvious limit for the conditionally comonotonic component and the convergence result in Vasicek

[21] for the conditionally independent component of the mixture):

$$\begin{aligned} & \lim_{n \rightarrow \infty} F_{L_n}(x) \\ & \equiv F_{L_\infty}(x) := \begin{cases} \omega(1 - E[P]) + (1 - \omega)F_P(x), & 0 \leq x < 1 \\ 1, & x = 1. \end{cases} \end{aligned} \quad (27)$$

Then, the mean and variance of the limiting distribution are

$$\begin{aligned} E[L_\infty] &= E[P] = \lim_{n \rightarrow \infty} E\left[\frac{S_n}{n}\right], \\ \text{Var}[L_\infty] &= \omega\{E[P] - (E[P])^2\} + (1 - \omega)\text{Var}[P] \\ &= \lim_{n \rightarrow \infty} \text{Var}\left[\frac{S_n}{n}\right]. \end{aligned} \quad (28)$$

In general, the k th moment of the limiting distribution is

$$E[L_\infty^k] = (1 - \omega)E[P^k] + \omega E[P], \quad k = 1, 2, \dots \quad (29)$$

The limiting result (27) can be applied to the area of financial credit risk management. The tail risk measures of a large-sample portfolio credit loss distribution are of particular interest for financial risk managers. The two well-known risk measures, value at risk (VaR) and conditional tail expectation (CTE), can be approximated by using the limiting distribution, (27). Formally, for a loss random variable X with the cumulative distribution function $F(x)$, VaR and CTE at confidence level $\alpha \in [0, 1]$ are defined as follows:

$$\begin{aligned} \text{VaR}_\alpha(X) &= F^{-1}(\alpha) = \inf\{x : F(x) \geq \alpha\}, \\ \text{CTE}_\alpha(X) &= \text{VaR}_\alpha(X) + \frac{\Pr[X > \text{VaR}_\alpha(X)]}{(1 - \alpha)} \\ &\quad \times E[X - \text{VaR}_\alpha(X) \mid X > \text{VaR}_\alpha(X)]. \end{aligned} \quad (30)$$

Note that this definition reduces to

$$\text{CTE}_\alpha(X) = E[X \mid X > \text{VaR}_\alpha(X)], \quad (31)$$

if X is a continuous random variable (see Hardy [22] or Acerbi and Tasche [23]).

Then, by the scaling property of these risk measures and the limiting result (27), for a large n , the VaR of the sum of n Bernoulli mixtures at level α can be approximated as follows:

$$\begin{aligned} & \text{VaR}_\alpha(S_n) \\ &= n\text{VaR}_\alpha(L_n) \approx n\text{VaR}_\alpha(L_\infty) = nF_{L_\infty}^{-1}(\alpha) \\ &= \begin{cases} 0, & 0 \leq \alpha \leq \omega(1 - E[P]) \\ nF_P^{-1}(\alpha^*), & \omega(1 - E[P]) < \alpha < 1 - \omega E[P] \\ n, & 1 - \omega E[P] \leq \alpha \leq 1, \end{cases} \end{aligned} \quad (32)$$

where $\alpha^* = \{\alpha - \omega(1 - E[P])\}/(1 - \omega)$.

Similarly, the CTE at level α can be approximated as

$$\begin{aligned} & \text{CTE}_\alpha(S_n) \\ &= n\text{CTE}_\alpha(L_n) \\ &\approx n\text{CTE}_\alpha(L_\infty) \\ &= \begin{cases} \left(\frac{n}{1 - \alpha}\right)E[P], & 0 \leq \alpha \leq \omega(1 - E[P]) \\ \frac{n}{1 - \alpha}\{(1 - \alpha^*)(1 - \omega)\text{CTE}_{\alpha^*}(P) + \omega E[P]\}, & \omega(1 - E[P]) < \alpha < 1 - \omega E[P] \\ n, & 1 - \omega E[P] \leq \alpha \leq 1. \end{cases} \end{aligned} \quad (33)$$

3. Beta Mixing Distribution

We now specialize the results of the preceding section to the case where the mixing distribution, F_P , is the Beta distribution with parameters $a > 0$ and $b > 0$. Specifically, the density $f_P(p)$ is given by

$$\begin{aligned} f_P(p) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \\ &= \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1}, \quad 0 < p < 1, \end{aligned} \quad (34)$$

where $\Gamma(\cdot)$ and $B(\cdot, \cdot)$ are the Gamma and Beta functions, respectively.

For a Beta Bernoulli mixture, several computationally convenient expressions are available. We first introduce a notation which is convenient for the cumulative distribution function:

$$F_P(p) = I_p(a, b) := \frac{B_p(a, b)}{B(a, b)}, \quad (35)$$

where $B_x(a, b) = \int_0^x s^{a-1}(1-s)^{b-1}ds$ is the incomplete Beta function. Equation (1), the pmf of the sum under the conditional independence assumption, can be evaluated as

$$\mu_n^I(k) = \binom{n}{k} \frac{B(k+a, n-k+b)}{B(a, b)}. \quad (36)$$

The k th moments for P and $(1 - P)$ in this case are

$$\begin{aligned} E[P^k] &= \frac{B(a+k, b)}{B(a, b)}, \\ E[(1-P)^k] &= \frac{B(a, b+k)}{B(a, b)}. \end{aligned} \quad (37)$$

The following lists several results for the ICM model with the Beta mixing distribution (β -ICM model).

Example 11 (β -ICM). The pmf of the binomial double mixture is given from (20) as

$$\mu_n(k) = \begin{cases} \omega \left(\frac{b}{a+b} \right) + (1-\omega) \frac{B(a, b+n)}{B(a, b)}, & k=0 \\ (1-\omega) \binom{n}{k} \frac{B(k+a, n-k+b)}{B(a, b)}, & k=1, \dots, n-1 \\ \omega \left(\frac{a}{a+b} \right) + (1-\omega) \frac{B(a+n, b)}{B(a, b)}, & k=n. \end{cases} \quad (38)$$

The limiting distribution of the empirical frequency, (27), is

$$F_{L_\infty}(x) = \begin{cases} \omega \left(\frac{b}{a+b} \right) + (1-\omega) I_x(a, b), & 0 \leq x < 1, \\ 1, & x = 1. \end{cases} \quad (39)$$

Since the inverse cumulative Beta distribution function does not admit a closed-form expression, a numerical method is required to approximate the VaR of a large-sample credit portfolio. Given the VaR at level α , the CTE can be written in a concise form. Specifically,

$$\text{VaR}_\alpha(S_n) \approx \begin{cases} 0, & 0 \leq \alpha \leq \frac{\omega b}{a+b} \\ n I_{\alpha^*}^{-1}(a, b), & \frac{\omega b}{a+b} < \alpha < 1 - \frac{\omega a}{a+b} \\ n, & 1 - \frac{\omega a}{a+b} \leq \alpha \leq 1, \end{cases} \quad (40)$$

where $\alpha^* := \{\alpha - \omega(b/(a+b))\}/(1-\omega)$, and

$$\text{CTE}_\alpha(S_n) \approx \begin{cases} \left(\frac{n}{1-\alpha} \right) \left(\frac{a}{a+b} \right), & 0 \leq \alpha \leq \frac{\omega b}{a+b} \\ \left(\frac{n}{1-\alpha} \right) \left(\frac{a}{a+b} \right) \left[1 - (1-\omega) I_{p^*}^{-1}(a+1, b) \right], & \frac{\omega b}{a+b} < \alpha < 1 - \frac{\omega a}{a+b} \\ n, & 1 - \frac{\omega a}{a+b} \leq \alpha \leq 1, \end{cases} \quad (41)$$

where $p^* = \text{VaR}_{\alpha^*}(P) = I_{\alpha^*}^{-1}(a, b)$.

As can be seen from (27), the limiting distribution of the empirical frequency under the β -ICM model has point masses at both end points, 0 and 1 (when $\omega \neq 0$). This may restrict the use of a certain parameter estimation method in the case that all observations are strictly inside the distributional range, which is often the case. In the following example, we consider a Beta distribution for the measures $\nu(\cdot; p)$, $0 < p < 1$, as well as a Beta mixing distribution for F_p . We call the model a *double Beta* model with acronym β - β . The limiting empirical frequency for the β - β model does not have point masses at the boundary points, 0 and 1.

Example 12 (β - β). In this example, the choice for $\nu(\cdot; p)$ is the Beta distribution with the two shape parameters θp and $\theta(1-p)$:

$$\begin{aligned} \nu(dr; p) &= f_\nu(r; p) dr \\ &= \frac{1}{B(\theta p, \theta(1-p))} r^{\theta p-1} (1-r)^{\theta(1-p)-1} dr, \quad 0 < r < 1, \end{aligned} \quad (42)$$

where parameter θ satisfies $0 < \theta < \infty$. Then, using the recursive property of the Beta function ($B(\alpha+1, \beta) = [\alpha/(\alpha+\beta)]B(\alpha, \beta)$),

$$a_k(p) = \int_0^1 r^k \nu(dr; p) = \prod_{j=0}^{k-1} \frac{\theta p + j}{\theta + j}, \quad (43)$$

with the usual convention that a product over an empty set of indices equals 1. In particular,

$$a_2(p) = \frac{\theta}{1+\theta} p^2 + \frac{1}{1+\theta} p, \quad (44)$$

and thus the limiting correlation for this model is, by Theorem 1,

$$\rho_{i,j} = 2 - \lim_{p \rightarrow 1} a'_2(p) = \frac{1}{1+\theta}. \quad (45)$$

(Note that this is identical to the result for an ICM model, with $\omega = 1/(1+\theta)$; cf. (19) in Example 9.)

The pmf of the binomial double mixture is given, by Theorem 4, as

$$\begin{aligned} \mu_n(k) &= \int_0^1 f_n(k; p) dF_p(p) \\ &= \binom{n}{k} \int_0^1 \sum_{i=0}^{n-k} \binom{n-k}{i} (-1)^i a_{k+i}(p) dF_p(p) \\ &= \binom{n}{k} \sum_{i=0}^{n-k} \binom{n-k}{i} (-1)^i \\ &\quad \times \int_0^1 \prod_{j=0}^{k+i-1} \frac{\theta p + j}{\theta + j} dF_p(p), \quad k=0, \dots, n. \end{aligned} \quad (46)$$

By (22), the limiting distribution of the empirical frequency is absolutely continuous, and its pdf is

$$f_{L_\infty}(x) = \int_0^1 f_\nu(x; p) dF_p(p), \quad 0 < x < 1. \quad (47)$$

Since (47) does not admit a closed-form expression, a numerical method is required to calculate the approximate VaR and CTE of a large-sample credit portfolio.

FIGURE 1: The cumulative distribution functions of S_{20} , for a scale of limiting correlations, ω , under the β -ICM model.

4. Numerical Examples

As an illustrative example, we set $a = 1.5$ and $b = 13.5$ for the Beta mixing distribution for both the β -ICM and β - β models. (These prescribed parameter values are chosen from the authors' previous experience in portfolio credit risk modeling.) This gives $E[P] = a/(a + b) = 0.1$ and $\text{Var}[P] = ab/[(a + b)^2(a + b + 1)] = 0.005625$.

4.1. β -ICM Model. In Figure 1, we plot the cumulative distribution functions of the β -ICM binomial double mixture, S_{20} , for several different values of ω (the limiting correlation). The plot shows that both the left and right tails get thicker as the level of limiting correlation, ω , increases to 1.

In Figures 2 and 3, we display the approximate risk measures, VaR and CTE, for a scale of confidence levels $\alpha \in \{0.8, 0.85, 0.90, 0.95, 0.99\}$ and by the level of the limiting correlation, ω . The results show that both VaR_α and CTE_α are increasing in the level of limiting correlation, for $\alpha \geq 0.90$. This is not the case for VaR_α when the confidence level is less than $1 - E[P] = b/(a + b)$, because

$$\begin{aligned} \alpha^* &= \frac{1}{1 - \omega} \left\{ \alpha - \omega \left(\frac{b}{a + b} \right) \right\} \\ &= \frac{1}{1 - \omega} \left\{ \alpha - \frac{b}{a + b} \right\} + \frac{b}{a + b} \end{aligned} \quad (48)$$

is decreasing in ω when $\alpha < b/(a + b)$.

More specifically, Table 1 compares the approximated VaR and CTE at the level $\alpha = 0.95$.

4.2. β - β Model. Recall that the limiting correlation of the β - β model is $1/(1 + \theta)$. For the purpose of comparison with the β -ICM model, we reparametrize the β - β mixture model as follows: $\theta = (1 - \omega)/\omega$, such that the limiting correlation becomes ω .

Figure 4 displays the cumulative distribution functions of the β - β binomial double mixture, S_{20} , for several different

FIGURE 2: The approximated VaRs of S_{1000} , by confidence level α and the limiting correlation ω , under the β -ICM model.

FIGURE 3: The approximated CTEs of S_{1000} , by confidence level α and the limiting correlation ω , under the β -ICM model.

TABLE 1: The approximated VaR and CTE of S_{1000} at confidence level $\alpha = 0.95$, under the β -ICM model.

ω	VaR_α	CTE_α
0.0	247.51	303.14
0.1	254.67	447.72
0.2	264.83	591.35
0.3	280.73	733.38
0.4	311.00	872.23
0.5	1000.00	1000.00

values of ω (the limiting correlation). The plot shows that S_{20} has a heavier left tail under the β - β model than under the β -ICM model, for each limiting correlation level. Note that the β - β model does not admit the two extreme cases: $\omega = 0$ and $\omega = 1$.

Figures 5 and 6 show the approximate risk measures, VaR and CTE, for a scale of confidence levels $\alpha \in \{0.8, 0.85, 0.90, 0.95, 0.99\}$ and by the level of the limiting

FIGURE 4: The cumulative distribution functions of S_{20} , for a scale of limiting correlations, ω , under the β - β model.

FIGURE 5: The approximated VaRs of S_{1000} , by confidence level α and the limiting correlation ω , under the β - β model.

correlation, ω . The approximate risk measures are smaller than those of the β -ICM model for each confidence level and limiting correlation; compare, for example, Figures 2 and 3. This indicates that the right tail of the β - β model is thinner than that of the β -ICM model for the same value of ω .

5. Models Fit to Real Data

In addition to the previous examples with prescribed parameter values, we illustrate the two binomial double mixture models with real data: Bloomberg mortgage delinquency rate index. U.S. residential mortgage loans are segmented into three buckets, Prime, Alt-A, and Subprime, based upon

FIGURE 6: The approximated CTEs of S_{1000} , by confidence level α and the limiting correlation ω , under the β - β model.

TABLE 2: The method of moments estimates of the shape parameters, a and b , of Beta mixing distribution and the limiting correlation ω .

Model	Parameter		
	a	b	ω
β -ICM	4.038	14.563	0.009
β -I	3.381	12.191	

the loan type. Several delinquency rates for each bucket are calculated as the percentage of loans that are reported to be delinquent beyond a certain number of days (30, 60, or 90 days) or are classified as real estate owned (REO) or foreclosure. In this section, we only use the Alt-A 60+ days (60+90+REO+foreclosure) delinquency rate (Ticker: BBMDA60P) for the purpose of illustration. The same approach can be taken for other delinquency indexes such as the Prime (Tickers: BBMDP60P, BBMDPDLQ) or Subprime (Tickers: BBMDS60P, BBMDSDLQ) segments. Figure 7 is taken from the Bloomberg interactive chart, and it plots the monthly history of the delinquency rate during May 2007–March 2012.

5.1. β -ICM Model. Note that none of the observations has the value 0 or 1. In this case, the maximum likelihood estimate of the weight parameter ω is 0 because, for $0 < p < 1$, $\log f_{L\infty}(p) = \log(1 - \omega) + \log f_p(p)$, where $f_p(p)$ is the density of the Beta distribution, (34). To avoid this issue, we estimate the three parameters in the β -ICM model by means of the method of moments (MMEs) which uses the first three moments, (29) and (37), of the limiting distribution of the empirical frequency.

For comparison, we also estimate the two shape parameters of the Beta distribution—the limiting distribution of the classical Beta-binomial (β -I) model—by matching the first and second sample moments with the theoretical ones, using (28) and (37). Table 2 gives the parameter estimates.

FIGURE 7: Bloomberg (60+ days) delinquency rate for Alt-A mortgage segment (May 2007–March 2012). Source: Bloomberg Finance L.P. used with permission of Bloomberg L.P. Copyright 2012. All rights reserved.

The table shows that the MMEs for the two shape parameters of the Beta mixing distribution of the general model ($\omega \neq 0$) are larger than those of the reduced model ($\omega = 0$). As a result, the right tail of the fitted Beta distribution in the general model is slightly thinner than that of the fitted Beta distribution in the reduced model. This is explained as follows: in the general model, the tail thickness in the data is explained by the estimates of both the shape parameters and ω . On the other hand, the reduced model is less flexible, and the two shape parameter estimates attempt to capture the observed tail.

In order to identify the implications of the two fitted models (with and without ω), for the risk measures, we plot approximate VaRs and CTEs of S_n with $n = 10^6$ at various confidence levels. (The chosen number of entities is purely nominal since we are using the large-sample results. According to the National Delinquency Survey issued by Mortgage Bankers Association, the average number of conventional subprime mortgage loans during 2011 is over four million.)

Figure 8 shows that the risk measures calculated under the conditional independence (β -I) model are larger than those of the more general model, the β -ICM model (17), at confidence levels up to 0.996 and 0.945 for VaR and CTE, respectively. Note that the estimated ω is small to the extent that the effect is not well recognized in VaR until the confidence level takes an extremely high value. The conditional tail expectation, however, of the more general model becomes larger than that under the conditional independence assumption at confidence levels which are typically used in the calculation of economic capital for a credit portfolio. The nonzero ω allows further dependence between entities,

and thus this result suggests that the use of the conditional independence model may result in the underestimation of extreme tail risk.

5.2. β - β Model. Here, we illustrate the β - β model of Example 12. Note that the large-sample distributions (of the empirical frequency) are continuous, for both the β - β and β -I models, and none of the observations are either 0 or 1. Thus, the maximum likelihood estimation (MLE) method can be applied by using the limiting distribution of the empirical frequency, (47). However, the number of observations, 57 data points, may not be sufficient to yield statistically reliable MLE fitting results. For the sake of statistical precision and model comparison, we reduce the number of parameters in both the β - β model and the β -I model, which corresponds to the case of conditional independence. Specifically, we use the following theoretical relationship between the first two moments of the Beta mixing distribution and the two shape parameters (a, b):

$$b = \left(\frac{\sigma^2}{\mu^3} \right) a(a + \mu), \quad (49)$$

where μ and σ^2 denote the mean and variance of the Beta mixing distribution, which we estimate with the sample mean and variance, $\bar{x} = 0.2171$ and $s^2 = 0.0103$, respectively, of the delinquency rate, x . Thus, we use the following form for b in the Beta density:

$$b = \left(\frac{0.0103}{0.2171^3} \right) a(a + 0.2171), \quad (50)$$

to obtain a single-parameter family of Beta distributions.

FIGURE 8: The approximated VaR_α and CTE_α of S_n , $n = 10^6$, the number of delinquencies in a month, at confidence levels, $0.9 \leq \alpha \leq 0.999$, based on β -ICM and β -I models.

TABLE 3: The maximum likelihood estimates of the model parameters under the β - β model and the β -I model.

Model	a	b	ω	Log-likelihood
β - β	3.481	12.900	0.0231	48.248
β -I	3.161	10.703		45.607

Table 3 gives the maximum likelihood estimates for the parameter a of the Beta mixing distribution, the limiting correlation parameter $\omega = 1/(1+\theta)$, and the values of the log-likelihood functions under the β - β model and the β -I model, respectively.

Figure 9 displays the differences in estimates of each of the two risk measures, VaR and CTE, under the β - β model and the β -I model. For illustration, the approximate risk measures are calculated assuming that there are one million entities in the mortgage portfolio.

The result shows that both the VaRs and the CTEs under the β - β model are larger than those under the β -I model, especially at high confidence levels, and the differences increase rapidly as the confidence level tends to 1.

This result demonstrates the role of the limiting correlation parameter ω in explaining the tail risk resulting from an interdependency among names within a credit portfolio.

6. Conclusion

In this paper, we study the conditional correlation and the distribution of the sum of Bernoulli mixtures under a general dependence structure. We show, in particular, under

FIGURE 9: The differences between approximated risk measures, under the β - β model and the β -I model, of S_n , $n = 10^6$, the number of delinquencies in a month for the Alt-A residential mortgage segment with (nominally) one million entities, at confidence levels, $0.9 \leq \alpha \leq 0.995$.

the typical conditional independence assumption, that the conditional correlation between two Bernoulli mixtures converges to 0, given that the mixing random variable is larger than a threshold tending to 1. We propose a method to construct a general dependence structure in the form of

a double mixture model, in which the conditional iid assumption is replaced by the more general assumption of conditional exchangeability. As a simple illustration, we consider a weighted average of two cases: conditional independence and comonotonicity, for which the limiting correlation is included as a model parameter.

The large-sample distribution of the empirical frequency and its use in approximating the risk measures VaR and CTE are presented. Several tractable results for two Bernoulli double mixture models with Beta mixing distribution, the β -ICM model and the β - β model, are given as illustrative numerical examples and also applied to real data.

Note that there is a strong demand for a credit risk model with an appropriate level of dependence in a stressed economic environment. The most popular model, the Beta-binomial mixture, however, cannot properly explain the empirically observed default correlation under stress. On the other hand, the model framework presented in this paper is simple but flexible enough to accommodate the required level of limiting correlation and thus can be effectively applied to portfolio credit risk models.

Future directions of research include the application of double mixture models to pricing of credit derivatives such as CDOs, on a completely homogeneous pool (e.g., Bae et al. [24]).

Appendices

A. Proof of Theorem 1

Let

$$\bar{F}_P(p) = 1 - F_P(p) = \int_p^1 f_P(s) ds \quad (\text{A.1})$$

be the survival probability function. Then, from (2),

$$\begin{aligned} \text{Cov}[X_i, X_j | P > p] &= \frac{1}{\bar{F}_P(p)} \int_p^1 a_2(s) f_P(s) ds \\ &\quad - \left(\frac{1}{\bar{F}_P(p)} \int_p^1 sf_P(s) ds \right)^2, \\ \text{Var}[X_i | P > p] &= \frac{1}{\bar{F}_P(p)} \int_p^1 sf_P(s) ds \\ &\quad - \left(\frac{1}{\bar{F}_P(p)} \int_p^1 sf_P(s) ds \right)^2. \end{aligned} \quad (\text{A.2})$$

Thus, the conditional correlation, given $P > p$, $0 < p < 1$, between X_i and X_j , is

$$\begin{aligned} \rho_{i,j}(p) &= \frac{\left(\frac{1}{\bar{F}_P(p)} \int_p^1 a_2(s) f_P(s) ds - \left(\frac{1}{\bar{F}_P(p)} \int_p^1 sf_P(s) ds \right)^2 \right)}{\left(\frac{1}{\bar{F}_P(p)} \int_p^1 sf_P(s) ds - \left(\frac{1}{\bar{F}_P(p)} \int_p^1 sf_P(s) ds \right)^2 \right)} \end{aligned}$$

$$= \frac{\bar{F}_P(p) \int_p^1 a_2(s) f_P(s) ds - \left(\int_p^1 sf_P(s) ds \right)^2}{\bar{F}_P(p) \int_p^1 sf_P(s) ds - \left(\int_p^1 sf_P(s) ds \right)^2}. \quad (\text{A.3})$$

Denote $g(p) = (1 - p)f_P(p)/\bar{F}_P(p)$ and $r(p) = \int_p^1 sf_P(s) ds / \bar{F}_P(p)$. By assumption, $\lim_{p \rightarrow 1} g(p) = L < \infty$. Note that, by L'Hospital's rule,

$$\begin{aligned} \lim_{p \rightarrow 1} r(p) &= \lim_{p \rightarrow 1} \frac{\int_p^1 sf_P(s) ds}{\bar{F}_P(p)} = \lim_{p \rightarrow 1} \frac{-pf_P(p)}{-f_P(p)} = 1, \\ \lim_{p \rightarrow 1} \frac{a_2(p) - p^2}{1 - p} &= 2 - \lim_{p \rightarrow 1} a_2'(p) \equiv 2 - a_2'(1^-). \end{aligned} \quad (\text{A.4})$$

Finally, to identify the limit of $\rho_{i,j}(p)$ as $p \rightarrow 1$, we apply L'Hospital's rule two more times:

$$\begin{aligned} \rho_{i,j} &= \lim_{p \rightarrow 1} \rho_{i,j}(p) \\ &= \lim_{p \rightarrow 1} \frac{\bar{F}_P(p) \int_p^1 a_2(s) f_P(s) ds - \left(\int_p^1 sf_P(s) ds \right)^2}{\bar{F}_P(p) \int_p^1 sf_P(s) ds - \left(\int_p^1 sf_P(s) ds \right)^2} \\ &= \lim_{p \rightarrow 1} \frac{-\int_p^1 a_2(s) f_P(s) ds - a_2(p) \bar{F}_P(p) + 2p \int_p^1 sf_P(s) ds}{-\int_p^1 sf_P(s) ds - p \bar{F}_P(p) + 2p \int_p^1 sf_P(s) ds} \\ &= \lim_{p \rightarrow 1} \left(\left(2(a_2(p) - p^2) f_P(p) - a_2'(p) \bar{F}_P(p) \right. \right. \\ &\quad \left. \left. + 2 \int_p^1 sf_P(s) ds \right) \times \left(2(p - p^2) f_P(p) - \bar{F}_P(p) + 2 \int_p^1 sf_P(s) ds \right)^{-1} \right) \\ &\equiv \lim_{p \rightarrow 1} \frac{2((a_2(p) - p^2)/(1 - p)) g(p) - a_2'(p) + 2r(p)}{2pg(p) - 1 + 2r(p)} \\ &= \frac{2(2 - a_2'(1^-))L - a_2'(1^-) + 2}{2L - 1 + 2} \\ &= \frac{2(2L + 1) - (2L + 1)a_2'(1^-)}{2L + 1} = 2 - a_2'(1^-). \end{aligned} \quad (\text{A.5})$$

Remark. If it were possible that $L = \infty$, then the previous proof is easily modified to cover that case. In the third-to-last line of the final derivation, one simply divides the numerator

and denominator by $g(p)$ before taking the limit, which then becomes

$$\frac{\lim_{p \rightarrow 1} ((a_2(p) - p^2) / (1 - p)) - 0}{\lim_{p \rightarrow 1} (2p) - 0} = 2 - \lim_{p \rightarrow 1} a'_2(p). \quad (\text{A.6})$$

B. Proof of Theorem 4

As remarked after the statement of the theorem, the proof is based on parts of the proof of de Finetti's theorem, as given in Feller [20], but for the case of a finite sequence. As it is well known that de Finetti's theorem does not hold in general for finite sequences (see, e.g., Example 4 at the end of Section 4 of Feller [20]), we will take care to elucidate exactly the ingredients of the proof which are valid in our case.

To begin, we recall some identities for numerical sequences of finite length. For any numerical sequence $a = (a_j : 0 \leq j \leq n)$, denote the difference operator by Δ :

$$(\Delta a)_j := a_{j+1} - a_j, \quad 0 \leq j \leq n-1. \quad (\text{B.1})$$

Then, the m th-order difference operator, denoted by Δ^m , is defined recursively by

$$\Delta^{m+1} = \Delta \circ \Delta^m, \quad \Delta^1 \equiv \Delta. \quad (\text{B.2})$$

It is easily shown by induction, using the elementary identity, $\binom{m}{i-1} + \binom{m}{i} = \binom{m+1}{i}$, that for $m \leq n$ and $j \leq n-m$,

$$(\Delta^m a)_j = \sum_{i=0}^m (-1)^{m-i} \binom{m}{i} a_{j+i}. \quad (\text{B.3})$$

(This is equation (1.7) in Chapter VII of Feller [20].) Next, we cite the identity (1.9) from Chapter VII of Feller [20]:

$$a_0 = \sum_{i=0}^n \binom{n}{i} (-1)^{n-i} \Delta^{n-i} a_i. \quad (\text{B.4})$$

Finally, if for some probability measure ν , $a_j = \int_0^1 r^j \nu(dr)$, for all j , then it is easily established by induction that

$$(\Delta^m a)_j = (-1)^m \int_0^1 r^j (1-r)^m \nu(dr). \quad (\text{B.5})$$

(This is equation (3.3) in Chapter VII of Feller [20].) Putting (B.3) and (B.5) together yields

$$\sum_{i=0}^m (-1)^i \binom{m}{i} a_{k+i} = \int_0^1 r^k (1-r)^m \nu(dr) \geq 0. \quad (\text{B.6})$$

We now proceed with the proof of Theorem 4, starting with the verification of (11) and (12). Temporarily overriding the definition (12), denote

$$g_n(k; p) := \Pr[(X_1, \dots, X_k, X_{k+1}, \dots, X_n) = (1, \dots, 1, 0, \dots, 0) \mid P = p]. \quad (\text{B.7})$$

Due to the exchangeability assumption, the conditional joint probability that the random vector $(X_1, \dots, X_k, X_{k+1}, \dots, X_n)$ takes on any permutation of the vector $(1, \dots, 1, 0, \dots, 0)$ is $g_n(k; p)$. Therefore,

$$f_n(k; p) = \Pr[S_n = k \mid P = p] = \binom{n}{k} g_n(k; p), \quad k = 0, \dots, n, \quad (\text{B.8})$$

and, thus, for (11) and (12), we must show that

$$g_n(k; p) = \sum_{i=0}^{n-k} \binom{n-k}{i} (-1)^i a_{k+i}(p). \quad (\text{B.9})$$

As shown in (4.6) in Feller [20] (identifying our $g_n(k; p)$ and $a_k(p)$ with his $p_{k,n}$ and c_k , resp.),

$$g_n(k; p) = (-1)^{n-k} \Delta^{n-k} a_k(p) = \sum_{i=0}^{n-k} \binom{n-k}{i} (-1)^i a_{k+i}(p), \quad (\text{B.10})$$

where, for the second equality, we have used the result (B.3) with $m = n-k$.

For the converse part of the theorem, we first show that $\sum_{k=0}^n f_n(k; p) = 1$, regardless of how the sequence $(a_k(p) : 0 \leq k \leq n)$ is chosen, as long as $a_0(p) = 1$. By (B.3), we recognize that

$$g_n(i; p) = (-1)^{n-i} \Delta^{n-i} a_i(p), \quad 1 \leq i \leq n. \quad (\text{B.11})$$

Therefore, by the identity (B.4), with $a_i \equiv a_i(p)$,

$$1 = a_0(p) = \sum_{i=0}^n \binom{n}{i} (-1)^{n-i} \Delta^{n-i} a_i(p) = \sum_{i=0}^n \binom{n}{i} g_n(i; p) = \sum_{i=0}^n f_n(i; p). \quad (\text{B.12})$$

Therefore, to show that f_n is a valid pmf, we must show that each term $f_n(k; p)$ or equivalently $g_n(k; p)$ is nonnegative. To achieve that, we simply apply (B.6), with $m = n-k$, to $a_k \equiv a_k(p)$, in case $a_k(p) = \int_0^1 r^k \nu(dr)$, $k = 0, 1, \dots$, as in the hypothesis of the converse part of the theorem. As a byproduct, we obtain the conditional binomial representation

$$g_n(k; p) = \int_0^1 r^k (1-r)^{n-k} \nu(dr; p). \quad (\text{B.13})$$

Finally, with the conditional pmf as described at (13), $(X_i)_{i=1}^n$ is clearly exchangeable. The proof of the converse will be complete as soon as it is shown that $\Pr[(X_1, \dots, X_k) = (1, \dots, 1) \mid P = p] = a_k(p)$, for all $k = 1, \dots, n$. For $k = n$, it follows by definition that $g_n(n; p) = a_n(p)$. For fixed $k < n$, denote

$$\mathcal{X}_j := \left\{ x \in \{0, 1\}^{n-k} : \sum_{i=1}^{n-k} x_i = j \right\}, \quad 0 \leq j \leq n-k; \quad (\text{B.14})$$

that is, each x in $\mathcal{X}_j$ has precisely j components equal to 1 and $n - k - j$ components equal to 0. Then,

$$\begin{aligned}
 & \Pr[(X_1, \dots, X_k) = (1, \dots, 1) \mid P = p] \\
 &= \sum_{j=0}^{n-k} \sum_{x \in \mathcal{X}_j} \Pr[(X_1, \dots, X_n) = (1, \dots, 1, x) \mid P = p] \\
 &= \sum_{j=0}^{n-k} \binom{n-k}{j} g_n(k+j; p) \\
 &= \sum_{j=0}^{n-k} \binom{n-k}{j} \\
 &\quad \times \int_0^1 r^{k+j} (1-r)^{n-k-j} \nu(dr; p), \quad \text{by (B.13),} \\
 &= \int_0^1 r^k \sum_{j=0}^{n-k} \binom{n-k}{j} r^j (1-r)^{n-k-j} \nu(dr; p) \\
 &= \int_0^1 r^k \nu(dr; p) = a_k(p),
 \end{aligned} \tag{B.15}$$

where the second line follows from the exchangeability of $(X_i)_{i=1}^n$; the second-to-last equality follows from the binomial expansion of $1 = [r + (1-r)]^{n-k}$.

Disclaimer

The views expressed in this paper are those of the authors and might not represent the views of the authors' affiliated institutions.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors are grateful to Bloomberg for granting permission to use the delinquency data in the examples of Section 5. Taehan Bae is supported by the Discovery Grant Program of the Natural Science and Engineering Council of Canada.

References

- [1] C. Donnelly and P. Embrechts, "The devil is in the tails: actuarial mathematics and the subprime mortgage crisis," *ASTIN Bulletin*, vol. 40, no. 1, pp. 1–33, 2010.
- [2] Basel Committee on Banking Supervision, "International convergence of capital measurement and capital standards: a revised framework," 2006, <http://www.bis.org/publ/bcbs128a.pdf>.
- [3] D. Li, "On default correlation: a copula function approach," *Journal of Fixed Income*, vol. 9, pp. 43–54, 2000.
- [4] L. Andersen and J. Sidenius, "Extensions to the Gaussian copula: random recovery and random factor loadings," *Journal of Credit Risk*, vol. 1, pp. 29–70, 2005.
- [5] X. Bertschell, J. Gregory, and J. P. Laurent, "Beyond the gaussian copula: stochastic and local correlation," *Journal of Credit Risk*, vol. 3, pp. 31–62, 2007.
- [6] X. Bertschell, J. Gregory, and J.-P. Laurent, "A comparative analysis of CDO pricing models under the factor copula framework," *Journal of Derivatives*, vol. 16, no. 4, pp. 9–37, 2009.
- [7] P. J. Schönbucher and D. Schubert, *Copula Dependent Default Risk in Intensity Models*, 2001.
- [8] C. Bluhm, L. Overbeck, and C. Wagner, *Introduction to Credit Risk Modeling*, Chapman & Hall/CRC, London, UK, 2003.
- [9] A. Cousin and J.-P. Laurent, "Comparison results for exchangeable credit risk portfolios," *Insurance: Mathematics & Economics*, vol. 42, no. 3, pp. 1118–1127, 2008.
- [10] R. Frey and A. J. McNeil, "VaR and expected shortfall in portfolios of dependent credit risks: conceptual and practical insights," *Journal of Banking and Finance*, vol. 26, no. 7, pp. 1317–1334, 2002.
- [11] R. Frey and A. McNeil, "Dependent defaults in models of portfolio credit risk," *Journal of Risk*, vol. 6, pp. 59–92, 2003.
- [12] KMV-Corporation, "Modelling default risk," 1997, <http://www.moodysanalytics.com>.
- [13] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management*, Princeton University Press, Princeton, NJ, USA, 2005.
- [14] A. J. McNeil and J. P. Wendin, "Bayesian inference for generalized linear mixed models of portfolio credit risk," *Journal of Empirical Finance*, vol. 14, no. 2, pp. 131–149, 2007.
- [15] F. Moraux, "Sensitivity analysis of credit risk measures in the beta binomial framework," *The Journal of Fixed Income*, vol. 19, no. 3, pp. 66–76, 2010.
- [16] RiskMetrics-Group, "CreditMetrics—Technical Document," 1997, <http://www.defaultrisk.com>.
- [17] T. Bae and I. Iscoe, "Correlations under stress," *International Review of Applied Financial Issues and Economics*, vol. 2, pp. 248–271, 2010.
- [18] M. Kalkbrener and N. Packham, "Correlation under stress in normal variance mixture models," *Mathematical Finance*. To appear.
- [19] Sh. Sharakhmetov and R. Ibragimov, "A characterization of joint distribution of two-valued random variables and its applications," *Journal of Multivariate Analysis*, vol. 83, no. 2, pp. 389–408, 2002.
- [20] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2, John Wiley & Sons, New York, NY, USA, 2nd edition, 1970.
- [21] O. Vasicek, "Limiting loan loss probability distribution," 1991, <http://www.moodysanalytics.com>.
- [22] M. R. Hardy, "A regime-switching model of long-term stock returns," *North American Actuarial Journal*, vol. 5, no. 2, pp. 41–53, 2001.
- [23] C. Acerbi and D. Tasche, "Expected shortfall: a natural coherent alternative to value at risk," *Economic Notes*, vol. 31, no. 2, pp. 379–388, 2002.
- [24] T. Bae, I. Iscoe, and C. Kim, "Valuing retail credit tranches with structural, double mixture models," *Journal of Futures Markets*. To appear.