

Sums of Possibly Associated Bernoulli Variables: The Conway–Maxwell–Binomial Distribution

Joseph B. Kadane

Abstract. The study of sums of possibly associated Bernoulli random variables has been hampered by an asymmetry between positive correlation and negative correlation. The Conway–Maxwell–Binomial (CMB) distribution gracefully models both positive and negative association. This distribution has sufficient statistics and a family of proper conjugate distributions. The relationship of this distribution to the exchangeable special case is explored, and two applications are discussed.

Keywords: correlated Bernoulli random variables, arithmetic binomial, multiplicative binomial, correlated binomial, beta-binomial.

1 Sums of possibly associated Bernoulli variables

There often are reasons to suggest that Bernoulli random variables, while identically distributed, may not be independent. For example, suppose pots are planted with six seeds each, where each pot has seeds from a unique plant, but different pot’s seeds came from different plants. Suppose that success of a seedling is well-defined. If genetic similarity is the dominant source of non-independence, it is reasonable to suppose positive association between seeds in the same pot. However, if competition for nutrients and sunlight predominates, association could be negative. Hence, it makes sense to find a functional form that gracefully allows for either positive or negative association.

“Association” here means something more general than correlation. Correlation is a particular measure of association, familiar because of its connection with the normal distribution, and its simple relationship to certain expectations. However, there is no particular reason why correlation should be used in non-normal situations if it has undesirable properties.

The desire for a functional form that allows for both positive and negative association in a symmetric way runs into the following familiar fact, which is well-known, but for completeness is proved in Appendix A:

Proposition 1. *Suppose X_1, \dots, X_m have (possibly different) means and variances and common pairwise correlations ρ . Then $\rho \geq -1/(m - 1)$.*

There are (at least) three different possible strategies for dealing with the asymmetry between positive and negative correlation revealed by the proposition:

Carnegie Mellon University, kadane@stat.cmu.edu

- (a) abandon correlation as a measure of association
- (b) abandon exchangeability of the Bernoulli random variables (exchangeability of order two implies identical one-and two-dimensional distributions)
- (c) model the sum directly, without fully specifying the distribution of the underlying Bernoulli random variables.

Some light on strategies (b) and (c) is shed by the following proposition, also proved in Appendix A.

Proposition 2. *Suppose X_1, \dots, X_m take values on $\{0, 1\}$. Let $P\{W = k\} = p_k \geq 0$, where $\sum_{k=0}^m p_k = 1$. Then there exists a unique distribution on X_1, \dots, X_m such that X_1, \dots, X_m are exchangeable of order m , and $\sum_{i=1}^m X_i$ has the same distribution as does W .*

Proposition 2 is reassuring with respect to strategy (c), since the set of distributions on the X 's corresponding to an arbitrary distribution on their sum is non-empty. However, it also shows that one can assume m -exchangeability among the X 's without restricting the distribution of their sum, so strategy (b) is superfluous. (This fact is also a consequence of Galambos' (1978) Theorem 3.2.1.)

The distribution studied in this paper pursues strategies (a) and (c) simultaneously.

The remainder of this paper is organized as follows: Section 2 reviews the literature on two-parameter extensions of the binomial distribution. Section 3 introduces the Conway–Maxwell–Binomial distribution and displays some of its mathematical properties. Section 4 gives sufficient statistics and discusses a conjugate prior family. Section 5 displays some examples, and gives expressions for its generating functions. The exchangeable case is examined in Section 6, and some applications are shown in Section 7.

2 Two-parameter generalizations of the binomial distribution

2.1 The beta-binomial distribution

Skellam (1948) proposed to endow the parameter p of the binomial distribution with a beta distribution with hyperparameters α and β . Then the probability mass function of the resulting distribution is

$$Pr\{W = k\} = \binom{m}{k} B(\alpha + k, \beta + m - k) / B(\alpha, \beta), \quad k = 0, 1, \dots, m, \quad (1)$$

where $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$. This distribution is widely used; see, for example, Chatfield and Goodhart (1970); Griffiths (1973) and Williams (1975).

At first it was believed that the beta-binomial could model positive association ($\alpha > 0, \beta > 0$), but not negative association. However, Prentice (1986) shows, to the contrary, that an extension of the beta-binomial can model negative association as well.

A family of distributions is marginally compatible if the marginal distributions are in the same family. De Finetti’s Theorem (1980) shows that mixtures of binomial distributions are marginally compatible. So the beta-binomial as a beta mixture of binomials is marginally compatible.

The beta-binomial distribution is not in the exponential family, which means it does not have sufficient statistics (except the whole sample). This implies that it is less computationally convenient than members of the exponential family, but still feasible using methods of George et al. (1993).

2.2 The multiplicative binomial model

This model is the first two terms in a log-linear model (Bishop et al., 1975). In a convenient parameterization, the probability mass function is

$$P\{W = k\} = \binom{m}{k} p^k q^{m-k} \theta^{k(m-k)} / f(\theta, p) \tag{2}$$

where $f(\theta, p) = \sum_{j=0}^m \binom{m}{j} p^j q^{m-j} \theta^{j(m-j)}$.

This model smoothly accommodates both positive ($\theta < 1$) and negative ($\theta > 1$) association. As discussed in Altham (1978), the multiplicative binomial model is not marginally compatible. This behavior is criticized in the context of a particular application, and a remedy proposed, by Verducci et al. (1988).

Since it is a member of the exponential family, the multiplicative binomial model has sufficient statistics and is computationally convenient.

2.3 The additive binomial model

Altham (1978) and Kupper and Haseman (1978) proposed this model, based on Lancaster (1969) definition of additive interaction in contingency tables (Darroch, 1974). It also has antecedents in the work of Lazarsfeld and Bahadur (Bahadur, 1961). The random variables X are marginally $B(1, p)$, have correlation ρ and have zero higher-order additive interactions.

The probability mass function is

$$P\{W = k\} = \binom{m}{k} p^k (1-p)^{m-k} \left[1 + \frac{\rho\{(k - mp)^2 + k(2p - 1) + mp^2\}}{2p(1-p)} \right], \tag{3}$$

$$\frac{-2}{m(m-1)} \min\left(\frac{p}{1-p}, \frac{1-p}{p}\right) \leq \rho \leq \frac{2p(1-p)}{(m-1)p(1-p) + 0.25 - \gamma_0}$$

and

$$\gamma_0 = \min_{0 \leq k \leq m} [\{k - (m-p) - 1/2\}^2].^1$$

¹This corrects an error in the Kupper and Haseman (1978) paper (their equation (4)). Rudolfer (1990) correctly gives the probability mass function, but fails to report that Altham’s “additive binomial model” and Kupper and Haseman’s “correlated binomial model” are identical.

Although Kupper and Haseman refer to this model as the “correlated binomial model”, this paper refers to it as the “additive binomial model”, in order to reserve the name for another model, discussed next.

2.4 The correlated binomial model

Madsen (1993); Luceño (1995); Luceño and De Ceballos (1995) and Diniz et al. (2010) discuss the correlated binomial model. If $B(m, p)$ denotes a binomial distribution with probability p of success in each of m independent trials, the correlated binomial probability mass function can be expressed as follows:

$$\begin{aligned} W|Z = 0 &\sim B(m, p), \\ W|Z = 1 &\sim mB(1, p), \\ Z &\sim B(1, \rho). \end{aligned} \tag{4}$$

Application of Proposition 2 shows that there is a unique m -exchangeable distribution on the X 's that has the specified distribution for their sum, W . Simple calculations conditioning on Z show that the X 's have a $B(1, p)$ distribution, marginally, and the correlation of pairs of X 's is ρ .

This model requires $\rho \geq 0$, so it does not accommodate negative association. While it is not a member of the exponential family, Diniz et al. (2010) show that posterior computation can be accomplished using Tanner and Wong's (1987)'s “data augmentation” (which is an unfortunate name, “parameter augmentation” would be more apt).

The correlated binomial model satisfies marginal consistency.

Another possibility is to generate dependent Bernoulli random variables using discrete copulas. For an arbitrary exchangeable copula $C(u_1, \dots, u_m)$ and in particular any Archimedean copula, $X_i = I(U_i \leq p)$ gives suitably dependent Bernoulli variables with success probability p . This construction is used most prominently with Gaussian or multivariate- t copulas in quantitative risk management (see McNeil et al. (2005, Chapter 5). Genest and Nešlehová (2007) discuss difficulties in estimating such models, but Smith and Khaled (2012) demonstrate methods of overcoming those issues.

3 The Conway–Maxwell–Binomial distribution

The Conway–Maxwell–Binomial distribution (CMB) is a convenient two-parameter family that generalizes the binomial distribution and models both positive and negative association among the Bernoulli summands.

The probability mass function of a random variable W having the CMB distribution is given by

$$P\{W = k\} = \frac{p^k (1-p)^{m-k} \binom{m}{k}^\nu}{S(p, \nu)}, \quad k = 0, 1, \dots, m \tag{5}$$

where

$$S(p, \nu) = \sum_{k=0}^m p^k (1-p)^{m-k} \binom{m}{k}^\nu.$$

Here $0 \leq p \leq 1$ and $-\infty \leq \nu \leq \infty$ (see Shmueli et al. (2005, Eq. (13))). Of course, when $\nu = 1$, the binomial distribution results.

When $\nu > 1$, the center of the distribution is upweighted relative to the binomial distribution and the tails downweighted. In the limit as $\nu \rightarrow \infty$, W piles up at $m/2$ if m is even, and at $\lfloor m/2 \rfloor$ and $\lceil m/2 \rceil$ if m is odd. Thus the component Bernoulli random variables are negatively related. Conversely, when $\nu < 1$, the tails are upweighted relative to the binomial distribution, and the center downweighted. In the limit as $\nu \rightarrow -\infty$, (5) puts all its probability on $W = 0$ and $W = m$, which is the extreme case of positive dependence (all X 's have the same value). Thus, ν measures the extent of positive or negative association in the component Bernoulli random variables. Figure 1 (from Kadane and Naeshagen (2013)) illustrates these points.

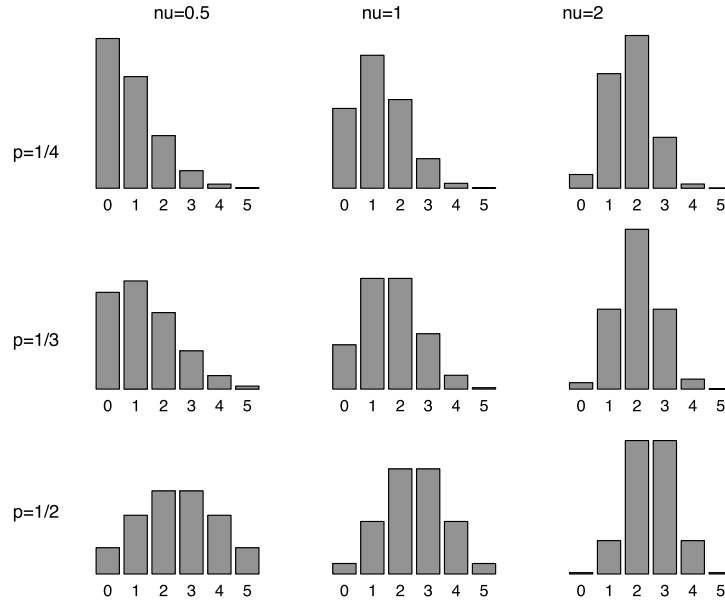


Figure 1: Examples of CMB when $m = 5$.

The name ‘‘Conway–Maxwell’’ comes from its relationship to the Conway and Maxwell (1962) generalization of the Poisson distribution, CMP (λ, ν) :

$$P\{W = x\} = \frac{\lambda^x}{(x!)^\nu M(\lambda, \nu)}, \quad x = 0, 1, \dots \tag{6}$$

where $M(\lambda, \nu) = \sum_{j=0}^\infty \lambda^j / (j!)^\nu$.

Shmueli et al. (2005) show that if $X \sim \text{CMP}(\lambda_1, \nu)$ and $Y \sim \text{CMP}(\lambda_2, \nu)$, X and Y independent, then

$$X \mid X + Y \sim \text{CMB} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}, \nu \right), \quad (7)$$

generalizing the familiar relationship between the Poisson and binomial distributions, when $\nu = 1$.

4 Sufficient statistics and a conjugate prior family

Imagine n samples from a CMB distribution, each with respect to a common m . Then the likelihood for p and ν is governed by the data k_1, \dots, k_n , and is given by

$$p(k_1, \dots, k_n \mid p, \nu) = \frac{\prod_{i=1}^n p^{k_i} (1-p)^{m-k_i} \binom{m}{k_i}^\nu}{[S(p, \nu)]^n}. \quad (8)$$

Provided n is considered known and fixed, the denominator is constant in the data, so it can be ignored. Then

$$\begin{aligned} p(k_1, \dots, k_n \mid p, \nu) &\propto (1-p)^{mn} \prod_{i=1}^n \left(\frac{p}{1-p} \right)^{k_i} \frac{m!^{\nu n}}{(k_i!(m-k_i!)^\nu} \\ &\propto \exp \left\{ \left(\sum_{i=1}^n k_i \right) (\log(p/1-p)) - \nu \sum_{i=1}^n \log[k_i!(m-k_i)!] \right\} \\ &= e^{S_1 \log(p/(1-p)) - \nu S_2} \end{aligned} \quad (9)$$

where $S_1 = \sum_{i=1}^n k_i$ and $S_2 = \sum_{i=1}^n \log[k_i!(m-k_i)!]$. Thus the CMB distribution is a member of the exponential family. Consequently, it has a conjugate prior family. To find a convenient form for this family, start over with the likelihood

$$p^k (1-p)^{m-k} \binom{m}{k}^\nu / \sum_{k=0}^m p^k (1-p)^{m-k} \binom{m}{k}^\nu. \quad (10)$$

We may take out the inessential factors of $(1-p)^m (m!)^\nu$, yielding

$$\left(\frac{p}{1-p} \right)^k \frac{1}{[k!(m-k)!]^\nu} / \sum_{k=0}^m \left(\frac{p}{1-p} \right)^k \frac{1}{[k!(m-k)!]^\nu}.$$

Let $\psi = \log(p/(1-p))$ and $t(k) = -\log(k!(m-k)!)$.

Then the probability mass function is

$$p(k \mid \psi, \nu) = \exp\{\psi k + \nu t(k) - M(\psi, \nu)\} \quad (11)$$

where $\exp(M(\psi, \nu)) = \sum_{k=0}^m \exp\{\psi k + \nu t(k)\}$.

Let $\theta = (\psi, \nu)$. Then (11) expresses the likelihood in the form of a natural exponential family.

I now explore the propriety of two different families of prior distributions, the first of which is the natural conjugate family associated with (11). Propriety of the prior distribution implies propriety of the posterior distribution, which is important for two reasons:

1. The usual argument for updating prior to posterior in conjugate families depends on the constant of proportionality being finite.
2. The proper behavior of numerical algorithms for computing posterior distributions, such as grid methods and Markov chain Monte Carlo (MCMC), also depend on the propriety of the posterior.

Since m is finite, $M(\theta) < \infty$ for all θ , so the natural parameter space for θ is \mathbb{R}^2 . The standard conjugate prior (Diaconis and Ylvisaker, 1979, Eq. (2.3)) is then

$$\pi((\psi, \nu)|a, b, c) \propto \exp\{\psi a + \nu b - cM(\psi, \nu)\}. \tag{12}$$

Theorem 1. *The distribution in (12) is proper if and only if*

$$(i) \ 0 < a/c < m$$

and

$$(ii) \ -\log(m!) < b/c < t(\lfloor a/c \rfloor) + (a/c - \lfloor a/c \rfloor)t(\lceil a/c \rceil) - t\lfloor a/c \rfloor. \tag{13}$$

The proof of Theorem 1 is given in Appendix B.

When propriety holds, updating is accomplished by

$$a' = a + k, b' = b + \log(k!(m - k)!) \quad \text{and} \quad c' = c + 1. \tag{14}$$

Suppose one wanted to center the prior distribution on the symmetric binomial distribution $B(m, 1/2)$, which corresponds to $\psi = 0$ and $\nu = 1$. One way to find approximate values for a, b , and c that do this is to observe that the mode of π satisfies

$$\nabla M(\psi, \nu) = (a/c, b/c). \tag{15}$$

Recalling that the gradient of M gives the expected values of the sufficient statistics k and $t(k)$, we have

$$a/c = m/2 \tag{16}$$

and

$$\begin{aligned} b/c &= \frac{-\sum_{k=0}^m \log[k!(m - k)!] \binom{m}{k}}{\sum_{k=0}^m \binom{m}{k}} \\ &= \frac{-\sum_{k=0}^m \log[k!(m - k)!] \binom{m}{k}}{2^m}. \end{aligned} \tag{17}$$

Furthermore, the referee suggests the following result, whose proof is in Appendix B.

Theorem 2. *The natural conjugate prior is symmetric around 0 if and only if*

$$a/c = m/2.$$

Another way is to notice that one can multiply the prior in (12) by any probability density $g(\psi, \nu)$ that leads to a proper distribution, and the result will be closed under sampling, with the same updating formulas. The following theorem gives a class of g 's leading to a proper such prior:

Theorem 3. Let $g(\psi, \nu)$ be a probability distribution defined on \mathbb{R}^2 with a finite moment generating function J . Then the prior proportional to

$$\exp(\psi a + \nu b - cM(\psi, \nu)g(\psi, \nu)) \quad (18)$$

is proper for all a, b , and c .

The proof of Theorem 3 is in Appendix B.

With the family (18), one can choose any distribution $g(\psi, \nu)$ whose mean vector is $(0, 1)$, with $a = b = c = 0$, and in particular

$$g(\psi, \nu) = \phi(\psi)\phi(\nu - 1), \quad (19)$$

where ϕ is the standard normal probability density function, to center at $B(m, 1/2)$.

5 Understanding the CMB distribution

One way to understand a distribution is to look at some representative examples of it. Figure 1 offers a matrix of such examples, for different values of p and ν .

Another way to understand a distribution is by way of its generating functions. These are derived next. Reconsider

$$\begin{aligned} S(p, \nu) &= \sum_{k=0}^m \binom{m}{k}^\nu p^k (1-p)^{m-k} \\ &= (1-p)^m \sum_{k=0}^m \binom{m}{k}^\nu \left(\frac{p}{1-p}\right)^k \\ &= (1-p)^m T\left(\frac{p}{1-p}, \nu\right), \end{aligned} \quad (20)$$

where $T(x, \nu) = \sum_{k=0}^m x^k \binom{m}{k}^\nu$.

Then the probability generating function of the CMB distribution can be expressed as

$$\begin{aligned} E(t^x) &= \sum_{k=0}^m t^k p^k (1-p)^{m-k} \binom{m}{k}^\nu / S(p, \nu) \\ &= (1-p)^m \sum_{k=0}^m \left(\frac{tp}{1-p}\right)^k \binom{m}{k}^\nu / S(p, \nu) \\ &= T(tp/(1-p), \nu) / T(p/(1-p), \nu). \end{aligned} \quad (21)$$

Similarly, the moment generating function and the characteristic function are, respectively,

$$E(e^{tx}) = T(e^t p / (1-p), \nu) / T(p / (1-p), \nu) \quad (22)$$

and

$$E(e^{itx}) = T(e^{it} p / (1-p), \nu) / T(p / (1-p), \nu). \quad (23)$$

6 Exchangeability of order 2

The CMB distribution is a distribution on the sum of m (possibly dependent) Bernoulli components without specifying anything else about the joint distribution of those components. This section explores the consequences of assuming in addition that those components are exchangeable of order 2.

To establish notation, let

$$p_{1,\dots,i_m} = P\{X_1 = i_1, X_2 = i_2, \dots, X_m = i_m\}, \tag{24}$$

where each $i_j \in \{0, 1\}$. Let π be a permutation of (i_1, \dots, i_m) . Then the random variables X are called m -exchangeable just in case

$$p_{1,\dots,i_m} = p_{\pi(i_1,i_2,\dots,i_m)} \tag{25}$$

for all permutations π .

Let $s(\ell, m)$ be the set of sequences (i_1, \dots, i_m) with exactly ℓ 1's, i.e., satisfying $\sum_{j=1}^m i_j = \ell$. There are $\binom{m}{\ell}$ such sequences in $s(\ell, m)$. The following theorem is given in the literature (see Diaconis (1977, Theorem 1) and the references cited there):

Theorem 4. *The set \mathcal{E}_m of m -exchangeable sequences is a convex set whose extreme points are e_0, \dots, e_m , where e_ℓ is the measure that puts probability $1/\binom{m}{\ell}$ on each element of $s(\ell, m)$ and 0 otherwise. Each point $x \in \mathcal{E}_m$ has a unique representation as a mixture of the $m + 1$ extreme points.*

Viewed in this light, the m -exchangeable set of CMB distributions specifies a particular two parameter family, with parameters p and ν , of weights on the extreme points e_0, \dots, e_m .

Because m -exchangeability applies to every permutation of length m , it implies m' exchangeability for each $m' < m$. Hence as m increases, m -exchangeability becomes increasingly restrictive. In the limit at $m = \infty$, de Finetti's Theorem shows that sums of exchangeable random variables are mixtures of Binomial random variables. Because the marginal distribution of each component is Bernoulli, interest centers on the joint distribution of pairs of such variables. By exchangeability of order 2, every pair has the same distribution as every other pair, so concentrating on (X_1, X_2) suffices. Exchangeability of order 2 implies that $P\{X_1 = 0, X_2 = 1\} = P\{X_1 = 1, X_2 = 0\}$, so there are really three probabilities to consider jointly, $p_{00} = P\{X_1 = 0, X_2 = 0\}$, $p_{01} = p_{10} = P\{X_1 = 0, X_2 = 1\}$, and $p_{11} = P\{X_1 = 1, X_2 = 1\}$. Diaconis (1977, p. 274) introduces a convenient way of graphing these quantities. The graph is reminiscent of barycentric coordinates, only here the constraint is slightly different:

$$p_{00} + 2p_{01} + p_{11} = 1; \quad p_{ij} \geq 0. \tag{26}$$

Figures 2 and 3 display the possible values of the exchangeable CMB distribution for specified values of m and ν , as p varies from 0 to 1.

In Figure 2, which is computed at $m = 3$, the curve for $\nu = 4$ is the highest, showing, as expected, more weight on $p_{01} = p_{10}$. The curve $\nu = 1$ is in the middle; this

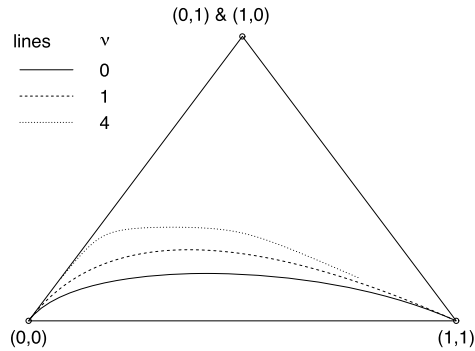


Figure 2: Possible values for $P\{X_1 = i, X_2 = j\}$ when $m = 3$.

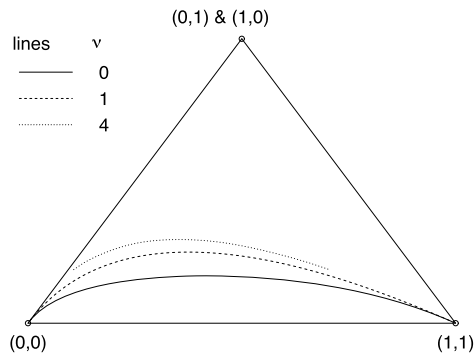


Figure 3: Possible values for $P\{X_1 = i, X_2 = j\}$ when $m = 5$.

one corresponds to independence, and is known to be $p(1-p)$. The curve for $\nu = 0$ is lowest. As $\nu \rightarrow -\infty$, this curve descends to the p_{00} to p_{11} line, indicating that all the probability is at the extremes.

Figure 3 shows the same curve, when $m = 5$. The main difference is that the $\nu = 4$ curve is flatter. Indeed, as $m \rightarrow \infty$, this curve will collapse to the $\nu = 1$ curve.

7 Applications

7.1 An agricultural experiment

Diniz et al. (2010) use a correlated binomial model to analyze data from an experiment on soybean seeds.

The data come from having planted six seedlings in each of 20 pots, and using the judgment of an expert as to which seedlings were successful. The goal was to examine the extent to which competition among the seedlings affected the outcomes. The raw

data given by Diniz et al. (2010) is reported in Table 1. They use an MCMC with data augmentation to fit the correlated binomial model to this data set.

# of “good” plants	# of pots observed
0	0
1	2
2	2
3	5
4	5
5	3
6	3

Table 1: Observed frequency of “good” plants from Diniz et al. (2010).

To employ the CMB model, I choose to use the prior specified by (19), with $a = b = c = 0$. This prior is centered on a Binomial model with $p = 1/2$ (which implies $\psi = 0$), which seems reasonable.

The contours of the resulting posterior distribution are shown in Figure 4. The maximum posterior point is $\hat{\psi} = 0.30$ and $\hat{\nu} = 0.52$, with inverse Hessian

$$\Sigma = \begin{pmatrix} 0.028 & 0.019 \\ 0.019 & 0.065 \end{pmatrix}.$$

In view of the elliptical shape of the contours in Figure 4, it is reasonable to approximate the posterior with a normal distribution with mean $(\hat{\psi}, \hat{\nu})$ and covariance Σ , as would be suggested by the asymptotic distribution of posterior distributions from conditionally independent models.

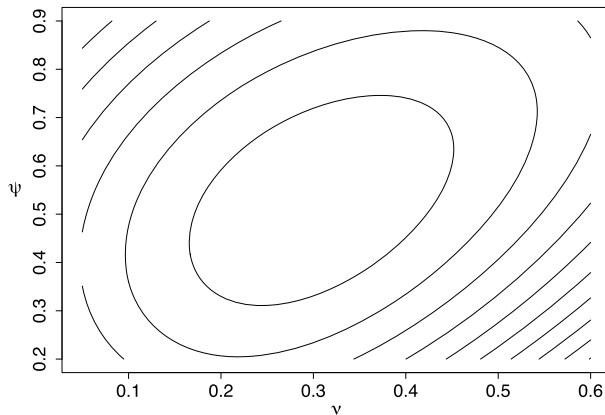


Figure 4: Contour plot of the CMB posterior distribution.

Extending Table 1, Table 2 below reports the estimated fitted values of each model.

The CMB and CB fits are from the maximum posterior point; the others are from maximum likelihood estimates. Table 2 shows that at least for this data set and this

measure of fit, the CMB has a smaller sum of squared errors than the other models. The CMB and binomial numbers are those of Diniz et al. (2010).

It is notable that the CMB estimate of ν is less than 1, indicating positive association in the soybean seeds. This suggests that competition for nutrients is not the dominant phenomenon in this data set. Further investigation and experimentation might then be warranted to discover the reasons for this positive association.

# of “good” plants	observed data	B	CB	BB	MB	AB	CMB
0	0	0.06	1.19	0.37	0.37	0.19	0.37
1	2	0.61	0.79	1.31	1.31	1.17	1.25
2	2	2.46	2.73	2.71	2.69	2.89	2.72
3	5	5.28	5.03	4.10	4.10	4.10	4.27
4	5	6.37	5.21	4.83	4.90	4.80	4.97
5	3	4.10	2.87	4.30	4.37	4.68	4.18
6	3	1.09	2.17	2.39	2.26	2.17	2.23
Sum of squared errors		8.96	4.16	4.01	4.32	5.86	3.73

Key: B = Binomial
 CB = Correlated Binomial
 BB = Beta-Binomial
 MB = Multiplicative Binomial
 AB = Arithmetic Binomial
 CMB = Conway–Maxwell–Binomial

Table 2: Fits of various models to the soybean data.

7.2 Killings in medieval Norway

In Norway just after the Viking Period, the law distinguished a killing from a murder. In both, there was somebody dead. However, in the former, the killer went to the King’s representative within 24 hours and confessed. (Absent such prompt confession, it would be a murder, punishable by execution or banishment). The King’s representative would write a letter to the killer stating that the killer was under the protection of the King. An investigation would ensue, resulting in a second letter to the killer, specifying how much was owed to the King, and how much to the family of the deceased. There would then be receipts to the killer for the payments (two more letters), and a final letter from the King’s representative to the killer saying that it was all over. Thus the killer would have received five letters.

Several hundred of these letters have survived in the intervening centuries, and a complete list of those found is available. Additionally, there are mentions of killings in other documents such as private letters, Bishop’s records, etc. A simple representation of the data is a 6×2 matrix, where the first dimension records the number of letters to the killer that survive, and the second is whether or not the killing is mentioned in other sources. Of course, there is the $(0, 0)$ cell of killings for which no letters survive and for

which there are no other mentions. To estimate this cell, and hence the total number of killings, Kadane and Naeshagen (2013, 2014) resort to a dual-systems estimate.

Since there’s no obvious reason why the survival of letters in the killer’s archive should be related to whether the killing is mentioned in the other sources, an independence assumption between the two dimensions seems reasonable. To model the number of letters from a given killing that might survive, a first thought might be a binomial model. However, since all five letters went to the killer, and were likely stored together, at least at first, it is reasonable to suppose that the event of the survival of a given letter to the killer would be positively associated with the event of the survival of the other letters to the same killer. Thus one would expect $\nu \leq 1$ in the CMB, and Kadane and Naeshagen imposed a prior on ν putting zero probability on the space $\nu \geq 1$. As it happened, the data favors $\nu > 1$, so the posterior piled up at $\nu = 1$, the binomial model.

Nonetheless, this was a successful application of the CMB, in that it allowed for (and rejected) what appeared to be the biggest reasonable threat to the binomial model.

8 Conclusion

This paper explores the properties of the Conway–Maxwell–Binomial (CMB) distribution. Table 3 gives a summary of the properties of this distribution in comparison to other two-parameter generalizations of the binomial distribution.

	BB	MB	AB	CB	CMB
1. Positive and Negative Association	Y	Y	Y	N	Y
2. Exponential Family	N	Y	N	N	Y
3. Marginally Compatible	Y	N	Y	Y	N

Table 3: Summary of properties of one-parameter generalization of binomial distribution.

It is not surprising that the distributions (CMB and MB) that are members of the exponential family are exactly those that are not marginally compatible. To be marginally compatible for all sample sizes requires (by de Finetti’s Theorem) that the distribution be a mixture of binomials, which (except for the trivial case) prevents it from being in the exponential family.

The Conway–Maxwell distribution is an alternative that deserves to be in a statistician’s toolbox.

Appendix A. Proof of Propositions 1 and 2

Suppose X_1, X_2, \dots, X_m have the same means and variances, and identical correlations ρ . Then $\rho \geq -1/(m - 1)$.

Proof of Proposition 1. Let $Y_i = (X_i - E(X_i))/\sigma(X_i)$, $i = 1, \dots, m$. Then Y_1, Y_2, \dots, Y_m satisfy $E(Y_i) = 0$ and $\text{Var}(Y_i) = 1$. Because correlations are unaffected by location and scale changes, they still have common covariance ρ . Now

$$\begin{aligned} 0 &\leq \text{Var}\left(\sum_{i=1}^m Y_i\right) = E\left(\sum_{i=1}^m Y_i\right)^2 - \left(E\left(\sum_{i=1}^m Y_i\right)\right)^2 \\ &= E\left(\sum_{i=1}^m Y_i\right)^2 = E\sum_{i=1}^m Y_i^2 + \sum_{\substack{i=1 \\ j=1 \\ i \neq j}}^m \sum_{i=1}^m EY_i Y_j \\ &= m + m(m-1)\rho \end{aligned}$$

from which the desired result follows immediately. \square

Remark. If the correlation between X_i and X_j is $\rho_{i,j}$ (not necessarily equal), the same proof shows that the average of the $\rho_{i,j}$'s is bounded below by $-1/(m-1)$.

Proof of Proposition 2. For each k , there are $\binom{m}{k}$ different arrangements of k 1's and $m-k$ 0's. Let each of them have probability $p_k/\binom{m}{k}$. Then $P\{\sum_{i=1}^m X_i = k\} = p_k$ and the X 's are exchangeable of order m .

To show uniqueness, if the sum of the probabilities of the sequences with exactly k 1's is not p_k , the sum condition is violated. If their probabilities are not equal, exchangeability of order m is violated. \square

Appendix B: Proofs of Theorems 1, 2, and 3

Proof of Theorem 1. Let μ be a measure on \mathbb{R}^2 that puts mass $1/(m+1)$ on each of the points $(k, t(k))$. Then (12) is the natural exponential family thus generated. According to Theorem 1 of Diaconis and Ylvisaker (1979), (12) is proper if and only if $(a, b)/c$ lies in the interior H of the convex hull \bar{H} of the points $(k, t(k))$, $k = 0, \dots, m$.

At integer points

$$k = 0, \dots, m, \quad -t(k) = \log(k!(m-k)!) = \log \Gamma(k+1) + \log \Gamma(m-k+1). \quad (\text{B.1})$$

Since the log-gamma function is strictly convex (i.e., has positive second derivative), so is $-t(k)$. Therefore, $t(k)$ is strictly concave. Consequently each of the points $(k, t(k))$, $k = 0, \dots, m$ is an extreme point of \bar{H} . Therefore, the boundary of \bar{H} consists of the line segments joining $t, t(k)$ and $(t+1, t(k+1))$, $k = 0, \dots, m-1$ and the segment joining $(0, t(0))$ and $(m, t(m))$.

Since $t(0) = t(m) = -\log(m!)$, the line segment joining $(0, t(0))$ and $(m, t(m))$ takes the constant value $-\log(m!)$ when $0 \leq t \leq m$. Furthermore, $m!e^{t(k)} = \binom{m}{k, m-k}$ so $t(k)$ is minimized at $t(0) = t(m)$. The line segment joining $(k, t(k))$ and $(k+1, t(k+1))$ is

$$t(k) + (x-k)[t(k+1) - t(k)] \text{ for } k \leq x \leq k+1.$$

Therefore, the elements of H , namely the elements of \bar{H} that are not boundary points, are those triples (a, b, c) satisfying

$$0 < a/c < m$$

and

$$-\log(m!) < b/c < t\lfloor a/c \rfloor + (a/c - \lfloor a/c \rfloor)(t\lceil a/c \rceil - t(\lfloor a/c \rfloor)) \quad \square$$

Proof of Theorem 2. The condition for symmetry around $\psi = 0$ is, for all ψ and some ν ,

$$a\psi - cM(\psi, \nu) = a(-\psi) - cM(-\psi, \nu). \quad (\text{B.2})$$

Now

$$M(-\psi, \nu) = \log \left(\sum_{k=0}^m e^{-\psi k + \nu t(k)} \right).$$

Let $j = m - k$. Then

$$M(-\psi, \nu) = \log \left(\sum_{j=0}^m e^{-\psi(m-j) + \nu t(m-j)} \right).$$

But $t(m - j) = t(j)$, so

$$M(-\psi, \nu) = -m\psi + \log \left(\sum_{j=0}^m e^{\psi j + \nu t(j)} \right) = -m\psi + M(\psi, \nu).$$

Therefore, (B.2) becomes

$$a\psi - cM(\psi, \nu) = a(-\psi) - c(-m\psi + M(\psi, \nu)),$$

or

$$(a + a - cm)\psi = 0.$$

Therefore,

$$a/c = m/2. \quad \square$$

Proof of Theorem 3. For large absolute values of ψ and ν , $M(\psi, \nu)$ is asymptotically linear. Consequently, for fixed values of a, b , and c , π is proper if g has a finite moment generating function. \square

References

- Altham, P. (1978). "Two generalizations of the binomial distribution." *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 27(2): 162–167. [MR0506839](#). doi: <http://dx.doi.org/10.2307/2346943>. 405

- Bahadur, R. (1961). “A representation of the joint distribution of n dichotomous items.” In: Solomon, H. (ed.), *Studies in Item Analysis and Predictions*. Stanford, California: Stanford University Press. [MR0121893](#). 405
- Bishop, Y., Fienberg, S., and Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice*. M.I.T. Press. [MR0381130](#). 405
- Borges, P., Rodrigues, J., Balakrishnan, N., and Bazan, J. (2014). “A COM-Poisson type generalization of the binomial distribution and its properties and applications.” *Statistics and Probability Letters*, 87: 158–166. [MR3168950](#). doi: <http://dx.doi.org/10.1016/j.spl.2014.01.019>. 420
- Chatfield, C. and Goodhart, G. (1970). “The beta-binomial model for consumer purchasing behavior.” *Applied Statistics*, 27: 34–37. 404
- Conway, R. and Maxwell, W. (1962). “A queuing model with state dependent service rates.” *Journal of Industrial Engineering*, 12: 132–136. 407
- Darroch, J. (1974). “Multiplicative and additive interaction in contingency tables.” *Biometrika*, 61(2): 207–214. [MR0403087](#). 405
- de Finetti, B. (1980). “Foresight: Its logical laws, its subjective sources.” In: Kyburg, H. and Smokler, H. (eds.), *Studies in Subjective Probability*, 55–118. New York: Krieger Publishing. [MR0584256](#). 405
- Diaconis, P. (1977). “Finite forms of de Finetti’s Theorem on exchangeability.” *Synthese*, 36: 271–281. [MR0517222](#). 411
- Diaconis, P. and Ylvisaker, D. (1979). “Conjugate priors for exponential families.” *The Annals of Statistics*, 7(2): 269–281. [MR0520238](#). 409, 416
- Diniz, A., Tutia, M., and Laite, J. (2010). “Bayesian analysis of a correlated binomial model.” *Brazilian Journal of Probability and Statistics*, 24(1): 68–77. [MR2580989](#). doi: <http://dx.doi.org/10.1214/08-BJPS014>. 406, 412, 413, 414
- Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics*. New York: J. Wiley and Sons. [MR0489334](#). 404
- Genest, C. and Nešlehová, J. (2007). “A primer on copulas for discrete data.” *Astin Bulletin*, 37(2): 475–515. [MR2422797](#). doi: <http://dx.doi.org/10.2143/AST.37.2.2024077>. 406
- George, E., Makov, U., and Smith, A. (1993). “Conjugate likelihood distributions.” *Scandinavian Journal of Statistics*, 20(2): 147–156. [MR1229290](#). 405
- Griffiths, D. (1973). “Maximum likelihood estimation for the beta-binomial distribution of the total number of cases of a disease.” *Biometrics*, 29: 637–648. 404
- Kadane, J. and Naeshagen, F. (2013). “The number of killings in southern rural Norway, 1300–1569.” *The Annals of Applied Statistics*, 7(2): 846–859. [MR3113492](#). 407, 415
- (2014). “Homicide rates in rural southern Norway 1300–1569.” *Scandinavian Journal of History*, 39(3): 287–298. Published online 01 Apr 2014. doi: <http://dx.doi.org/10.1080/03468755.2014.882270>. 415

- Kupper, L. and Haseman, J. (1978). “The use of a correlated binomial model for the analysis of certain toxicological experiments.” *Biometrics*, 34: 69–76. 405
- Lancaster, H. (1969). *The Chi-Squared Distribution*. London: Wiley. MR0253452. 405
- Luceño, A. (1995). “A family of partially correlated Poisson models for overdispersion.” *Computational Statistics and Data Analysis*, 20: 511–520. 406
- Luceño, A. and De Ceballos, F. (1995). “Describing extra-binomial variation with partially correlated models.” *Communications in Statistics: Theory and Methods*, 24(6): 1637–1653. 406
- Madsen, R. (1993). “Generalized binomial distributions.” *Communications in Statistics: Theory and Methods*, 22(11): 3065–3086. MR1245533. doi: <http://dx.doi.org/10.1080/03610929308831203>. 406
- McNeil, A., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management*. Princeton University Press. MR2175089. 406
- Prentice, R. (1986). “Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors.” *Journal of American Statistical Association*, 81: 321–327. 404
- Rudolfer, S. (1990). “A Markov chain model of extrabinomial variation.” *Biometrika*, 77(2): 255–264. MR1064797. doi: <http://dx.doi.org/10.1093/biomet/77.2.255>. 405
- Shmueli, G., Minka, T., Kadane, J., Borle, S., and Boatwright, P. (2005). “A useful distribution for fitting discrete data: Revival of the Conway–Maxwell–Poisson distribution.” *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 54(1): 127–142. MR2134602. doi: <http://dx.doi.org/10.1111/j.1467-9876.2005.00474.x>. 407
- Skellam, J. (1948). “A probability distribution derived from the binomial distribution by regarding the probability of success as variable between sets of trials.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 10: 257–261. MR0028539. 404
- Smith, M. and Khaled, M. (2012). “Estimation of Copula Models using discrete margins via Bayesian Data Augmentation.” *Journal of American Statistical Association*, 107(497): 290–303. MR2949360. doi: <http://dx.doi.org/10.1080/01621459.2011.644501>. 406
- Tanner, M. and Wong, W. (1987). “The calculation of posterior distributions by data augmentation.” *Journal of the American Statistical Association*, 82: 528–540. MR0898357. 406
- Verducci, J., Mack, M., and DeGroot, M. (1988). “Estimating multiple rater agreement for a rare diagnosis.” *Journal of Multivariate Analysis*, 27: 512–535. MR0970970. doi: [http://dx.doi.org/10.1016/0047-259X\(88\)90145-5](http://dx.doi.org/10.1016/0047-259X(88)90145-5). 405
- Williams, D. (1975). “The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity.” *Biometrics*, 31: 949–952. 404

Acknowledgments

I thank a careful and persistent referee for saving me from publishing theorems (and proofs!) that were wrong. I thank Bronwyn Woods, Christian Robert, Kim Sellers, Galit Shmueli, Rebecca Steorts and Rhiannon Weaver for helpful comments. Additionally, I thank Josemar Rodrigues for bringing Borges et al. (2014) to my attention. This paper independently finds the sufficiency properties and generating functions of the COM-Binomial distribution.