Binary Cumulants

David C. Torney

Theoretical Biology and Biophysics, Theoretical Division, Mail Stop K710,

Received May 27, 1998; accepted February 27, 2000

The binary cumulant is defined for joint probability distributions on binary sequences of finite length. The binary cumulant is bounded, in magnitude, by unity, and is shown to vanish if there exists any bipartition of the letter positions into statistically independent blocks. Probability distributions on binary $n$-sequences are shown to map injectively into their binary cumulants for all subsets of the set of letter positions. An inversion algorithm is established, recovering the joint distribution from its collection of binary cumulants. © 2000 Academic Press

1. INTRODUCTION

Injective mappings of the space of probability distributions are highly important—for both pure and applied mathematics—but they are also uncommon. An exhaustive, nondegenerate list of known injective mappings is fairly short. Distributions on binary sequences of finite length, for instance, map injectively into their moments and into their classical cumulants—evaluated for all subsets of the letter positions [1–4]. The respective codomains often constitute concise parameterizations of distributions. Consider, for instance, the importance of expectations, variances, and so on. The known injective mappings have all been extensively studied for over 100 years. Here, injectivity of the mapping into binary cumulants is established, albeit only for distributions on binary sequences of finite length.

All the sequences treated in this manuscript have letters from $\{0, 1\}$. It is easily seen, however, that, with a natural extension of the notion of parity,

$^1$ This manuscript is dedicated to the memory of Professor G.-C. Rota.
The properties of binary cumulants make them equally advantageous as classical cumulants for many applications—such as characterizing many samples of data sequences. Furthermore, their conceptual simplicity make them ideal candidates for solving fundamental mathematical problems. For instance, the concise and complete characterization of collections of numbers which may be the moments of probability distributions may be regarded as complete [5, 6]. This problem also has a close relationship with Hilbert’s 17th problem [7]. Next to nothing is known, however, about the analogous problem for classical cumulants: the concise and complete characterization of collections of numbers which may be the classical cumulants of probability distributions. This challenging problem was enthusiastically propounded by the late G.-C. Rota, who referred to it as the cumulant problem. It is reasonable to expect that it will be easier to resolve the analogous binary-cumulant problem and, also, that the methods of solution will be helpful for the resolution of the classical-cumulant problem.

In one regard, however, binary cumulants are more complex than classical cumulants. Analytical formulas, mentioned below, express probability distributions on binary sequences in terms of their classical cumulants [2]. On the other hand, an algorithm, such as the one given in the final section, is required for inversion: deriving a probability distribution from its binary cumulants.

2. DEFINITION

Consider a finite set $S = \{i_1, i_2, \ldots, i_{|S|}\} \subset \mathbb{N}$, where, without loss of generality, $i_j < i_{j+1}$ for $1 \leq j < |S|$. Let there be a given probability distribution $P_S(a_{i_1} a_{i_2} \cdots a_{i_{|S|}})$ for $(0, 1)$-binary $|S|$-sequences on the positions indexed by the elements of $S$. Then, the binary cumulant $\beta(S)$ has the following definition:

$$\beta(S) \overset{\text{def}}{=} 2^{|S|-1} \sum_{\xi} \left\{ \prod_{\xi} P_S(a_{i_1}^{(\xi)} a_{i_2}^{(\xi)} \cdots a_{i_{|S|}}^{(\xi)}) - \prod_{\xi} P_S(a_{i_1}^{(\xi)} a_{i_2}^{(\ xi)} \cdots a_{i_{|S|}}^{(\xi)}) \right\}. \quad (1)$$

Here, $\xi$ ranges over the even-parity sequences on $S$ and $\xi$ ranges over the odd-parity sequences on $S$, with digits denoted by $a_{i_j}^{(\xi)}$ and $a_{i_j}^{(\xi)}$, respectively. As always, the parity of $a_{i_j}^{(\xi)} a_{i_2}^{(\xi)} \cdots a_{i_{|S|}}^{(\xi)}$ is congruent, modulo two, to $\sum_{j=1}^{|S|} a_{i_j}$, with $\forall \in \{\xi, \xi\}$. 
Define $\beta(\emptyset) = 1$. In addition, from (1),

$$\beta(\{i\}) = P_{\{i\}}(0) - P_{\{i\}}(1)$$

and

$$\beta(\{i, j\}) = 4(P_{\{i, j\}}(00)P_{\{i, j\}}(11) - P_{\{i, j\}}(01)P_{\{i, j\}}(10)),$$

where $i$ and $j$ are, of course, distinct. Mapping $\{0, 1\}$ to $\{-1, +1\}$, the foregoing binary cumulants are recognized as the classical cumulants of zero, one, and two specified positions for distributions on finite-length sequences with letters from $\{-1, +1\}$. However, no binary cumulant of a larger number of positions is equivalent to the corresponding classical cumulant.

3. ELEMENTALS

Let $U \overset{\text{def}}{=} \{1, 2, \ldots, n\}$, with $1 \leq n < \infty$, and let there be given a joint probability distribution $P(a_1a_2\cdots a_n)$ on binary $n$-sequences. Note that, when no ambiguity is engendered, $P(a_1a_2\cdots a_n)$ supplants $P_U(a_1a_2\cdots a_n)$.

In the remainder of this manuscript, $S = \{i_1, i_2, \ldots, i_{|S|}\} \subseteq U$ and $P_S(a_i, a_{i_2}\cdots a_{i_{|S|}})$ of (1) is taken to be the marginal probability of the given joint distribution [8, p. 166]. This, evidently, defines a $\beta(S)$ for all $S \subseteq U$.

Because the maximum of a product of non-negative real numbers with fixed sum is found by taking these numbers equal to one another [9, Theorem 9], the “normalization coefficient” $2^{|S|}2^{-|S|}$ ensures that $|\beta(S)| \leq 1$, for all $S \subseteq U$. In contrast, no tight upper bound is known for the magnitudes of all but a few of the classical cumulants, even in the special case of distributions on binary $n$-sequences [2].

Recalling the definition of statistical independence, a joint distribution on $n$-sequences decomposes into two independent blocks of letter positions if and only if the joint distribution equals the product of the marginal distributions on the positions from the two corresponding blocks [1, p. 36]. In this idiom, a telling property of binary cumulants is the following.

**Proposition 1.** $\beta(S)$ vanishes if there exists any bipartition of $S$ into two statistically independent blocks.

**Proof.** Let there be a probability distribution on binary $|S|$-sequences exhibiting two independent blocks, $L$ and $R$, of sizes $l$ and $|S| - l$, respectively, with, by definition, $1 \leq l \leq |S| - 1$, $S = L \cup R$, and $\emptyset = L \cap R$. From the definition of independence, the corresponding probability for any binary $|S|$-sequence equals the product of the marginal probability for the subsequence on $L$ with that for the subsequence on $R$. For the even-parity sequences, occurring in the left-hand product of (1), these subsequences
must both have even parity or odd parity. It is readily seen that the left-hand product equals the product of the marginal probabilities for all the $2^l$ possible binary $l$-sequences, on $L$, raised to the power $2^{|S|-l-1}$ (because each $l$-sequence occurs with $2^{|S|-l-1}$ $(|S|-l)$-sequences to generate the set of even-parity sequences) times the product of the marginal probabilities for all $2^{|S|-l}$ possible binary $(|S|-l)$-sequences, on $R$, raised to the power $2^{l-1}$. A similar analysis reveals an identical factorization for the right-hand product of (1). Therefore, the binary cumulant vanishes.

Note that classical cumulants vanish whenever there is a partition of the letter positions into any number of mutually independent blocks. Combining these blocks, as necessary, two independent blocks may always be obtained, demonstrating that binary cumulants are, with regard to statistical independence, perceptually indistinguishable from classical cumulants.

**Theorem 1.** Joint probability distributions on binary $n$-sequences, $P(a_1 a_2 \cdots a_n)$, map injectively into their collections of binary cumulants $\beta(S)$, where $S$ ranges over all subsets of $U$ and where the probabilities included in $\beta(S)$ are the marginal probabilities $P_S(\cdot \cdot \cdot)$.

**Proof.** Injectivity of the mapping holds if and only if a probability distribution on binary $n$-sequences is uniquely determined by the binary cumulants for all subsets of the letter positions, or indices. Injectivity holds for $n = 1$ because, from (1), $\beta(\{1\}) = P(0) - P(1)$, and, hence, $P(0) = (1 + \beta(\{1\}))/2$ and $P(1) = (1 - \beta(\{1\}))/2$. Assume, therefore, that the theorem is valid for joint probability distributions on binary $(n-1)$-sequences and consider a distribution on binary $n$-sequences. Under the induction hypothesis, all the marginal probabilities involving $n-1$ or fewer indices are uniquely determined by the corresponding binary cumulants, ranging over the subsets of the respective subset of indices. The following lemma isolates “fully joint” properties of distributions from “proper marginal” properties, which is integral to the proof of the theorem.

**Lemma 1.** The probability of an arbitrary binary $n$-sequence may be expressed in the form $A + \Omega$, when the sequence is of even parity, and in the form $A - \Omega$, when the sequence is of odd parity. Here, $A$ is a given summation, (3), of signed, proper marginal probabilities defined below, and $\Omega$ denotes the probability for the $n$-sequence consisting of all zeroes; $1 \leq n < \infty$.

**Proof.** Let $S = \{i_1, i_2, \ldots, i_{|S|}\} \subseteq U$. Also, let $M(S)$ denote the marginal probability of having zeroes on the positions indexed by the elements of $S$. Hence, $M(\varnothing) = 1$. Let $W(S)$ denote the probability $P$ of the $n$-sequence consisting of zeros at positions in $S$ and ones at positions in $U \setminus S$. The principle of inclusion and exclusion [10] yields the following formula for
\[ W(S) = \sum_{S \subseteq T \subseteq U} M(T) (-1)^{|T|-|S|}. \]  
\[ (2) \]

Taking \( \Omega = M(U) (= W(U)) \) it follows that
\[ A = \sum_{S \subseteq T \subseteq U} M(T) (-1)^{|T|-|S|}. \]  
\[ (3) \]

Noting the sign of \( \Omega \), the proof of the lemma is complete.  

Note that only the marginal probabilities for proper subsequences of all zeroes enter into (3) and, therefore, that \( A = 0 \) for \( W(U) (= \Omega) \).

According to the induction hypothesis, the marginal probabilities for proper subsequences are determined by the binary cumulants for proper subsets of the \( n \) indices. Thus, joint distributions on binary \( n \)-sequences, with given \( \beta(S) \) for all proper subsets \( S \subset U \), may differ only in the value taken by \( \Omega = M(U) \) (cf. Lemma 1). The proof of the theorem will, therefore, be completed by establishing a one-to-one correspondence between values of \( \Omega \) and values of \( \beta(U) \), assuming fixed proper marginal probabilities. Lemma 1 and (1) motivate the definition of the following function of the real variable \( \Omega \).

\[ b(\Omega) \overset{\text{def}}{=} 2^{(|U|-1)2^{(|U|-1)}} \left\{ \prod_{\varepsilon} (A_{\varepsilon} + \Omega) - \prod_{\varepsilon} (A_{\varepsilon} - \Omega) \right\}. \]

Here, \( \varepsilon \) ranges over the even-parity \( n \)-sequences and \( \varepsilon' \) ranges over the odd-parity \( n \)-sequences. Also, \( A_{\varepsilon} \) and \( A_{\varepsilon'} \) denote the constants derived from (3), for sequences \( \varepsilon \) and \( \varepsilon' \), respectively. Thus, \( b(\Omega) = \beta(U) \), given \( \Omega \) and all the proper marginal probabilities.

In the closed interval \( T \) of admissibility for \( \Omega \), \( (A_{\varepsilon} + \Omega) \) and \( (A_{\varepsilon} - \Omega) \) are non-negative, for all \( \varepsilon \) and \( \varepsilon' \), respectively. From the preceding equation,
\[ \frac{db(\Omega)}{d\Omega} = 2^{(|U|-1)2^{(|U|-1)}} \left\{ \prod_{\varepsilon} (A_{\varepsilon} + \Omega) \sum_{\varepsilon'} (A_{\varepsilon'} + \Omega)^{-1} + \prod_{\varepsilon} (A_{\varepsilon} - \Omega) \sum_{\varepsilon'} (A_{\varepsilon'} - \Omega)^{-1} \right\}. \]
\[ (4) \]

Here, as usual, \( \varepsilon \) and \( \varepsilon' \) range over the even-parity \( n \)-sequences and \( \varepsilon \) and \( \varepsilon' \) range over the odd-parity \( n \)-sequences. As remarked, for admissible \( \Omega \) all multiplicands and summands on the right-hand side of (4) are non-negative. Noting that there must be at least one nonvanishing probability everywhere in \( T \), the right-hand side of (4) must be positive for all \( \Omega \in T \). Therefore, \( b(\Omega) \) is strictly isotone with \( \Omega \) within \( T \). Because \( b(\Omega) \) is evidently continuous, the desired one-to-one correspondence between \( \beta(U) \)
and $\Omega$ is established. Thus, the map from probability distributions on binary $n$-sequences into the collection of binary cumulants for all subsets of the letter indices is injective. Therefore, by induction, the theorem is valid for $1 \leq n < \infty$.

An elementary corollary of (2) is that there is an injection of joint probability distributions on binary $n$-sequences into collections of the foregoing marginal probabilities $M(S)$, for all non-empty $S \subseteq U$.

4. INVERSION

“Inversion” denotes the determination of a joint probability distribution from an injective mapping of the space of distributions. For conventional mappings, inversion is accomplished analytically. For instance, encoding the symbols of the binary sequence into $\{-1, +1\}$, the moment representation is inverted by summing all “signed” moments, with the appropriate (sequence-dependent) signs [2]. Theorem 1 establishes the invertibility of binary cumulants. Furthermore, the methods and statements of the proof of Theorem 1 yield the following inversion algorithm.

Let there be given the collection of $\beta(S)$ for $S \subseteq U$. Recall that $M(\emptyset) = 1$. Next, determine the $n$ marginal probabilities, $M(\{i\})$—the marginal probability of a zero occurring at position $i$—with $1 \leq i \leq n$. From (1), $M(\{i\}) = (1 + \beta(\{i\}))/2$. There is also an explicit formula for $M(\{i, j\}) = \beta(\{i, j\})/4 + M(\{i\})M(\{j\})$. The monotonicity and continuity of $b(\Omega)$ may, alternatively, be employed, in conjunction with Newton’s method, to determine $M(\{i, j\})$ from the corresponding $\beta(\{i, j\}), M(\{i\}),$ and $M(\{j\})$, for all $1 \leq i < j \leq n$. Repeated application of this procedure will yield the $M(\{i, j, k\})$ from the corresponding $\beta(\{i, j, k\})$, for all $1 \leq i < j < k \leq n$, and so on. Thus, “levels” one through three of the lattice $B_n$ are assigned $\beta$s. At level $\ell$, the marginal probabilities, $M(S)$ for $S \subseteq U : |S| = \ell$, are found analogously, using the previously determined marginal probabilities (for smaller-size subsets). The algorithm terminates with the determination of the marginal probability for the entire set, $M(U)$. The latter equals an $\Omega$ which, from Lemma 1, will, with the previously determined marginal probabilities, yield the probabilities for all sequences.

ACKNOWLEDGMENTS

David J. Balding provided valuable comments on this manuscript. This work was performed with the support of the U.S.D.O.E. Human Genome Project and the Center for Nonlinear Studies of Los Alamos National Laboratory.
REFERENCES