Information content of high-order associations of the human gut microbiota network

Weston D. Viles¹, Juliette C. Madan²,³,⁴, Hongzhe Li⁵, Jason C. Moore⁶, Margaret R. Karagas²,³, and Anne G. Hoen¹,²,³

Abstract

Background: The human gastrointestinal tract is an environment that hosts an ecosystem of microorganisms essential to human health. Vital biological processes emerge from fundamental inter- and intra-species molecular interactions that influence the assembly and composition of the gut microbiota ecology.

Results: Here we quantify the complexity of the ecological relationships of the infant gut microbiota ecosystem as a function of the information contained in the non-linear associations of a sequence of increasingly-specified maximum entropy representations of the system. Our paradigm frames the ecological state in terms of the presence or absence of an individual microbial taxonomic unit as a function of both the ecological states of its neighboring units and, departing from standard graphical model representations, the associations among the units within its neighborhood.

Conclusion: We characterize the order of the system based on the relative quantity of statistical information encoded by high-order statistical associations of the infant gut microbiota.

Background

The ecological relationships of the microbial ecosystem of the human gut are influenced by fundamental molecular interactions among microorganisms and underpin the formation of complex, robust communities that correspond to important biological functions for its host. [1, 2, 3, 4] The fundamental relationships among mutually-coupled microorganisms of the gut ecosystem possibly

¹Department of Biomedical Data Science, The Geisel School of Medicine at Dartmouth
²Department of Epidemiology, The Geisel School of Medicine at Dartmouth
³Children’s Environmental Health and Disease Prevention Research Center at Dartmouth
⁴Division of Neonatology, Department of Pediatrics, Children’s Hospital at Dartmouth
⁵Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania
⁶Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania
⁷Department of Mathematics and Statistics, University of Southern Maine
involve multiple taxonomically-distinct ecological units in commensal, symbiotic, and pathogenic relationships [5, 6]. Statistical analysis of the states of this system may reveal the ecological associations, e.g. combinations of ecological relationships among the incident ecological units, that manifest from these fundamental relationships. An ecological relationship is basically quantified by its order, e.g. the number of taxonomically-distinct units involved in the association, and the direction and magnitude of its corresponding statistical association measurement, e.g. correlation. Graphical models and network analysis methods are often-used tools for representing the observed second-order, e.g. pairwise, ecological relationships of the gut ecosystem [7, 8, 9] and, yet, do not sufficiently descriptive to express the stability-diversity paradox observed in nature [10, 11]. Associations have commonly been estimated with pairwise correlation [12, 13], despite its technical limitations [14, 15], or by partial correlation through generalized linear models [16], including logistic regression [17]. Network motifs and clusters are frequently identified subsequently as proxy higher-order associations [18, 19, 20] though, as we demonstrate in the following, this paradigm of communities as composites of pairwise relationships cannot fully capture the breadth of complex ecological associations in the system.

Early proponents [21, 22] for the existence of high-order, e.g. involving more than two units, interactions in ecological systems described the complex relationships in terms of one entity modifying the nature of an interaction in the system. Detecting the non-linear relationships that comprise a high-order association is a long-standing statistical problem [23]. More recently, the functionality of high-order interactions has been inferred to include the of promoting of stability and diversity in ecological communities [24, 25, 26]. Methodology for the integration of more than two operational taxonomic units (OTUs) in a regulatory triplet model [27] and estimation of dynamic networks that evolve according to a state space model [28] have been used in constructing microbial network models. Within the microbial communities of the human gut, higher-order interactions may occur on account of competing enzyme-antibiotic production [29] or environmentally adaptive trophic interactions [30], in which ancillary species influence pairwise interactions to support and regulate the diversity of multi-species communities [31, 32]. Such high-order associations have been shown to influence host fitness traits in Drosophila melanogaster fruit fly, the microbiome of which consists of few combinations of microorganisms [33] that are feasibly enumerated.

A common mathematical reduction of the state of an ecosystem is the binary vector of zero-one occurrence indicators of the taxonomically-distinct units of an ecosystem [9, 12]. The statistical properties of this state vector are derived from the probability distribution characterizing the likelihood of possible ecological states. We identify a class of low-rank approximations of this empirical probability distribution, commonly known as the maximum entropy distributions, on ecological subsystems, each consisting of a fixed number of taxonomically-distinct ecological units, and quantify the relative information gains associated with statistical representations of increasingly higher rank. As a corollary, we subsequently lower-bound the maximal order of association in the subsystems.
Based on a sample of 381 infants, to each corresponds a single observation, we demonstrate the broad existence of high-order ecological associations in sub-ecosystems of the human gut microbiota. Our conclusion supports modern ecological theory on the combinatorial nature of mutual influence among biological entities [11]. Consideration of higher-order statistical associations is pertinent to the accurate prediction/detection of abnormal states of the human microbiota.

The methodology we develop is a general procedure for estimating a low-rank approximation of the statistical distribution of states in a binary system and characterizing, in terms of order, the complexity of its associations. We understand complexity in the context a complex system and, in particular, the nonlinear associations that are a defining characteristic of such systems. We represent the observed states of the binary system (composed of zero-one ecological occurrence indicator variables) with a sequence of progressively more-specified maximum entropy models indexed by order. The increasingly-descriptive statistical models in this sequence potentially facilitate, in turn, an increase in the predictability of states as quantified through entropy-based measures. [34]

Specifically, we measure the relative statistical information attributed to the high-order statistical associations relative to those of lower-order in the sequence of estimated low-rank approximations of the statistical distribution of states. Our methodology implicitly quantifies the statistical information gained through representations of the high-order associations of the gut microbiota ecosystem and thereupon demonstrate that third-order ecological relationships are abundant and important for characterizing the statistical properties of the subsystems.

**Results**

To provide further context for our forthcoming results on the complexity of associations of the gut microbiota, we begin by providing insight on stereotypical non-linear statistical associations and the performance of our method in these situations. A non-linear association among three binary variables $x_1, x_2, x_3$ is exemplified through the standard exclusive or, e.g. “at least one but not both”, example [35] in which the variables $P(x_1 = 0, x_2 = 0) = p_{00}$ and $P(x_1 = x'_1, x_2 = x'_2) = (1 - p_{00})/3$ for $(x'_1, x'_2) \neq (0, 0)$ and $x_3 = x_1 + x_2 - 2x_1x_2$ so that $x_3 = 1$ when either $x_1 = 1$ or $x_2 = 1$ and $x_3 = 0$, otherwise. Note that the three variables are equi-probable with $P(x_j = 1) = 2(1 - p_{00})/3$ and equi-correlated with $Cov(x_i, x_j) = (4p_{00} - 1)(1 - p_{00})/9$, for $i, j \in \{1, 2, 3\}$ and $i \neq j$. If $p_{00} = 2/5$ then the three variables are positively correlated with $Cor(x_i, x_j) = 1/6$ and, yet, $P(x_1 = 1, x_2 = 1, x_3 = 1) = 0$. The mathematical relationship between $x_3$ and $x_1, x_2$ conceptualizes a non-linear ecological association in which a triad of taxonomically-distinct units do not occur collectively but are, otherwise, supportive of pairwise co-occurrences.

A linear classifier, e.g. logistic regression, for the occurrence variable $x_3$ based on the main variables $x_1$ and $x_2$ (no interaction term) would be expected...
to classify incorrectly \( x_3 \) with rate at least \( \mathbb{P}(x_1 = 1, x_2 = 1) = 1/5 \), for example, when trained on a random sample of sub-system states. To that end, the importance of the inclusion of third-order association among the three occurrence variables \( x_1, x_2, x_3 \) in a statistical representation of the occurrence states is depicted in Fig. 1b. For the sake of illustration, we include the \textit{max} model in which \( x_3 = \max\{x_1, x_2\} \) and the \textit{relaxed xor} model in which \( x_3 \), conditional on \( x_1, x_2 \), has the Bernoulli \( (1 - x_1 x_2^2) \) distribution.

Each path interpolates the points \((d, I_d)\), where \( I_d \) is the quantity of statistical information attributed to associations higher than \( d \)-th order as measured by the Kullback-Leibler divergence from the true probability distribution \( z \) of the three variables to the \( d \)-th order approximating maximum entropy distribution \( z_d \), for \( d \in \{1, 2, 3\} \). For all data sets, in general, all such paths are non-increasing functions of order \( d \in \{1, 2, \ldots, p\} \). The difference \( R_d \propto I_d - 1 - I_d \) is reflected in the negative magnitude of the jumps.

With respect to the xor model, the second-order maximum entropy distribution in the form of Eq. (2) is a modest improvement—accounting for an \( \approx 6\% \) reduction in the relative information from its first-order counterpart to the true probability distribution of the xor model. This carries the interpretation that the second-order probability model is expected to be equally predictive of occurrence states as the first-order probability model. However, with the third order association encoded, the third-order probability model approximates the true probability distribution of states with arbitrary accuracy. The second- and third-order associations present in the third-order model encode all of the statistical information attributable to the associations of this sub-system. While this is trivially the case since we encoded associations of all possible orders in this third-order probability model, it is straight-forward that the argument would remain the same, for example, in the context of a fourth and fully-independent occurrence variable \( x_4 \) and the maximal order of four would need not be encoded to represent all statistical information attributable to associations in the expanded subsystem.

The fully independent model encodes no associations and is provided as a baseline for comparison. The \textit{max} model is constructed to, like the xor model, define \( x_3 \) deterministically as a function of \( x_1, x_2 \) but to alleviate the necessity to encode a third-order association to represent the full statistical information of all orders. The \textit{relaxed or} assigns zero probability to the event \( x_j = 1 \) for \( j = 1, 2, 3 \) and, otherwise, is a Bernoulli(1/2) random variable negatively associated pairwise with \( x_1, x_2 \). A visualization of the existence, sign, and magnitude of the associations of the three different models in each of the approximating maximum entropy distributions of orders \( d \in \{1, 2, 3\} \) is seen in Fig. 1d. Select mathematical details are provided in the Supplementary Information.

To estimate the statistical information quantities of interest in sub-systems of \( p \) amplicon sequence variants (ASVs) ecological occurrence variables, for \( p \in \{5, 8, 10\} \), we obtain a random sample of 5000 subsets of \( p \) ASVs from the set of all such subsets and record their respective ASV occurrence profiles, e.g. observed ecological states, over all 381 infant samples, see Experimental Procedures. For each subset of \( p \) ASVs, we estimate the \( d^{th} \)-order maximum
Fig. 1. Illustration of the information content of third-order association. a. For three occurrence variables $x_1, x_2, x_3$ with $P(x_1 = 0, x_2 = 0) = 2/5$ and $P(x_1 = x_1', x_2 = x_2') = 1/5$ for $(x_1', x_2') \neq (0, 0)$ we define the models (red) “exclusive or” (xor) $x_3 = x_1 + x_2 - 2x_1x_2$, (green) max $x_3 = \max\{x_1, x_2\}$, (blue) relaxed xor $x_3 = 0$ if $x_1 = 1$ and $x_2 = 1$ and otherwise $x_3 \sim \text{Bernoulli}(1/2)$, and (black) $x_1, x_2, x_3$ mutually independent b. For each $d = 1, 2, 3$, node points on the paths represent the statistical information attributable to orders $> d^{th}$-order as a proportion of information attributable to all orders of association. c. Proportion of statistical information attributable to $d^{th}$-order associations $R_d$, for $d = 2, 3$. d. Visualization of associations represented in the maximum entropy distribution approximation to the distribution of states in the sub-system of three occurrence variables from each model. A relatively strong association is shaded darker than weaker associations. Positive associations are shaded blue whereas negative associations are shaded red.
fitness model $z_d = g_d(\beta_d \lambda)$ and compute $R_d = (I_{d-1} - I_d)/(I_1 - I_p)$, for $d \in \{1, 2, ..., 5\}$. We consider at most 5th-order representations due to the sample size and computational limitations related to the numerical estimation of $2^p - 1 = \sum_{j=1}^{p} (^p_j)$ parameters in the optimization problem in Eq. (4).

On sub-systems of $p = 5$ ASV occurrence variables, we estimate a median 0.7390 proportion of statistical information content attributed to second-order associations $R_2$, with an interquartile range of (0.636,0.871) over the 5000 subsets. Fig. 2a illustrates this prominent role for second-order associations in the prediction of occurrence states across subsets. The 0.355% of subsets which have statistical information allocated entirely in the second-order associations induce the bimodal distribution illustrated in Fig. 2b. The remaining statistical information is primarily attributable to third-order associations—the median 0.1616 proportion of statistical information attributed to third-order associations $R_3$. The median value 0.9040 of $R_2 + R_3$ over all 5000 subsets of $p = 5$ ASV occurrence variables—with an interquartile range of (0.868,0.991)–indicates that the vast majority of statistical information pertinent to prediction of ecological states is attributable to second- and third-order associations encoded in the third-order maximum entropy models. The median 0.777 value of $R_*$ statistic, with a corresponding inter-quartile range (0.713,0.819), implies that our model estimation procedure encoded through associations of all order at least 71.3% of the mutual information among $p = 5$ ASVs occurrence profiles in 75% of 5000 samples. The complementary quantity is the residual information that is attributable to a finite sample size and regularization, see Experimental Procedures and the Supplementary Information.

The third-order maximum entropy models for $p = 8$ and $p = 10$ occurrence variables are parameterized by 92 and 175 quantities, respectively. (By comparison, 31 for the fully-specified model for $p = 5$ variables) In lieu of $I_p$ for both $p = 8, 10$, we bound

$$\frac{I_1 - I_2}{I_1} \leq R_2 = \frac{I_1 - I_2}{I_1 - I_p} \leq \frac{I_1 - I_2}{I_1 - I_3}$$

$$\frac{I_2 - I_3}{I_1} \leq R_3 = \frac{I_2 - I_3}{I_1 - I_p} \leq \frac{I_2 - I_3}{I_1 - I_3},$$

since 0 $\leq I_p \leq I_3$, for $p = 8, 10$. Adding these inequality systems results in

$$1 - \frac{I_3}{I_1} \leq R_2 + R_3 \leq 1.$$

In Fig. 2c, we note that the median upper bound for $R_2$ for each of the $p \in \{5, 8, 10\}$ considered is apparently decreasing with $p$ and indicates a diminished prediction efficacy of the second-order maximum entropy models and necessarily greater quantities of statistical information attributed to higher associations in the probability models on increasingly larger sub-ecosystems.

---

1We compute $R_* = 1 - I_p/I_1$ for $d = 5$, see the Supplementary Information.
Fig. 2. Illustration of non-linear associations in a sub-system of three occurrence variables. a. Over a random sample of 5000 subsets of \( p = 5 \) ASVs the proportion of statistical information attributed to each order \( R_d \), for \( d \in \{2,3,4,5\} \), arranged according to the sorted values of \( R_2 \) with quartiles indicated. The sample median \( q_{0.5} = 0.7390 \) of the \( R_2 \) statistic indicates that the estimated second-order associations account for at most 75% of statistical information attributable to all orders of association in no more than half of the sampled five-vertex subsets. c. Upper bound for \( R_2 \), the proportion of statistical information attributed to second-order associations, for subsets of \( p \in \{5, 8, 10\} \) ASVs over a random sample of 5000 subsets each. d. Lower bound for \( R_2 + R_3 \), the proportion of statistical information attributed to third-order associations, respectively over the samples.
in the second- and third-order maximum entropy models are 0.683, 0.679, and 0.697, for \( p = 5, 8, 10 \), respectively. Fig. 2 illustrates empirically that the lower bound statistic for \( p = 5, 8, 10 \) are distributed similarly, each with an interquartile range of approximately \((0.607,0.754)\). This common statistical behavior of the estimated lower bound for the quantity of statistical information pertinent to prediction of occurrence states in small sub-ecosystem attributable to the associations encoded in the second- and/or third-order maximum entropy models predominates the total quantity of statistical information attributed to all order of association. In other words, second- and third-order associations prevail as the foundation of an accurate representation of the collective statistical behavior of occurrence variables in small sub-ecosystems of the human gut.

**Discussion**

The states of a microbial sub-ecosystem of the human gut elicit a statistical description which encodes high-order associations. In our manner of estimating and allocating the statistical information attributed to the associations encoded in a sequence of maximum entropy models of increasing specification in our observations of the infant gut microbiota ecosystem, we identified a prominent role for the third-order maximum entropy model in the prediction of ecological states in a statistical representation sub-ecosystems. Third-order associations predominate the estimated high-order associations and, in concert with the pairwise association they may encompass, are typically ascribed at least two-thirds of the information attributable to associations of all orders of ecological occurrence variables. Our statistical analysis provide convincing evidence that small sub-systems of \( 5 \leq p \leq 10 \) ecological occurrence variables are commonly at least third-order systems.

Our methodology for attributing the components of statistical information to specific orders of association is widely applicable within complex systems analysis. In particular, our estimation and evaluation procedure gathers evidence from the data to quantify the predictability of states as a function of model complexity which, in the present context, is indexed by the maximum order of a association in a maximum entropy distribution approximating the empirical distribution of the observed states.

Our subsequent analysis uncovered the existence of and quantified the extent to which associations among microbial occurrence variables are non-linear. The accurate prediction of outcomes from clinical interventions or perturbations of the gut ecosystem will require encoding of these complex relationships of the system. We have established the breadth of high-order associations that modify the lower-order association they encompass in the gut microbiota environment and, as a result, established a baseline for the level of difficulty in prediction of gut microbiota states.

The statistical behavior of ecological co-occurrences among the microorganisms of the infant gut microbiota is vital information for accurate prediction of possible states of the ecosystem. Based on 381 samples collected from participat-
ing infants, we demonstrated that small (up to 10 ASVs) subsystems the infant gut ecosystem predominantly require a probability model which encodes the statistical properties of third-order, e.g. involving three units, co-occurrences to represent the statistical information attributable to ecological associations and relevant to prediction of ecological states. We described our numerical procedure for the estimation of the proportion of statistical information ascribed to a range of orders of association in a binary representation of the co-occurrences of units in an ecosystem.

**Conclusion**

The statistical behavior of ecological co-occurrences among the microorganisms of the infant gut microbiota is vital information for accurate prediction of possible states of the ecosystem. Based on 381 samples collected from participating infants, we demonstrated that small subsystems the infant gut microbial ecosystem predominantly require a probability model which encodes the statistical properties of third-order co-occurrences of taxa in order to represent the statistical information attributable to ecological associations and relevant to prediction of ecological states. We described our numerical procedure for the estimation of the proportion of statistical information ascribed to a range of orders of co-occurrences of taxa in an ecosystem and propose that our findings imply an important role for complex interactions of microbes in the infant gut.

**Abbreviations**

ASV: amplicon sequence variant  
OTU: operational taxonomic unit  
rRNA: ribosomal ribonucleic acid

**Methods**

We describe the sample collection and the numerical methods of our procedure in the following.

**Sample collection, sequencing, and processing**

Mothers aged 18 to 45 years participating in the New Hampshire Birth Cohort Study at Dartmouth provided infant stool samples that were collected at regularly scheduled maternal six-week postpartum follow-up visits (mean: 46 days, median: 44 days, range: 14-153 days, s.d.: 13.7 days). Institutional review board approval was obtained at Dartmouth with yearly renewal. Subjects provided written informed consent to participate on behalf of themselves and their infants. Stool was aliquoted in sterile tubes and frozen at $-80^\circ$C within 24 hours of receipt. Samples were thawed and DNA was extracted using the Zymo
DNA extraction kit (Zymo Research). The quantity and purity of the DNA were determined by OD260/280 nanodrop measurement. Illumina tag sequencing of the 16S rRNA gene v4-v5 hypervariable region and initial quality control was performed at the Marine Biological Laboratory in Woods Hole, Massachusetts. Quality control procedures eliminated sequences containing more than one ambiguous nucleotide, removed sequences with a length outside of the expected distribution, and eliminated chimeric reads using the UCHIME algorithm [36] de novo and with reference within the USEARCH program [37]. Sequences were processed using the DADA2 sequence processing pipeline (v.1.6.0) [38] to infer the amplicon sequence variants (ASVs) present and their relative abundances across samples. Sequencing and sequence read processing were done using established methods that have been previously described [39, 40, 41]. We identified 2663 ASVs from 381 infant stool samples.

The ASV abundances recorded in the 381 \times 2663 ASV table were transformed to the 381 \times 2663 binary matrix with the \((i,j)\) element equal to one if the abundance of the \(j^{th}\) ASV in the \(i^{th}\) sample exceeds the detection limit, i.e. the corresponding entry in the ASV table is positive and otherwise zero, for \(i \in \{1,2,\ldots,381\}\) and \(j \in \{1,2,\ldots,2663\}\). We compute the univariate information entropy \(H(\hat{p}_j) = -\hat{p}_j \log(\hat{p}_j) - (1 - \hat{p}_j) \log(1 - \hat{p}_j)\) of occurrences for the \(j^{th}\) taxonomic unit, where \(\hat{p}_j\) is the occurrence rate (sample proportion) of the \(j^{th}\) taxon occurrence variable for \(j \in \{1,2,\ldots,2663\}\), and eliminate those with low entropy via the elbow method [42]. This amounts to retaining 295 ASVs in the 381 \times 295 binary matrix \(D\) with occurrence rates in the interval \([0.0210, 0.9291]\) and focuses our exploration for high-order associations in high-entropy components of the ecosystem. The discarded low- and high-rate ASVs are predominated (63.88%) by ASVs with a single occurrence in the data set and are easily predictable in their own rights.

**Low-rank Approximation**

The occurrence of 295 taxonomically-distinct ecological units in an observation of the gut ecosystem is a binary (one if present, zero if absent) state vector \(x \in \{0,1\}^{295}\) in a sample space of \(2^{295} > 10^{88}\) states. We concentrate on sub-ecosystems consisting of a fixed number of taxonomically-distinct units \(p\), for \(p \in \{5,8,10\}\), for which the quantities of statistical information encoded by associations can be feasibly computed and recorded over a multitude of instances. Statistical properties of the \(p\)-length binary random vector \(x \in \{0,1\}^p\) are functions of its probability distribution \(P : \{0,1\}^p \mapsto [0,1]\). Let \(x_k \in \{0,1\}^p\) be the binary representation of the number \(k\), for \(k \in \{0,1,2,\ldots,2^p - 1\}\), and define the \(2^p \times p\) matrix \(X\) to have \(k^{th}\) row \(X_{k} = x_k\). The sample space of all \(p\)-length binary vectors \(x \in \{0,1\}^p\) is the union of the row vectors of \(X\).

Define \(z \in \mathbb{R}^{2^p}\) with \(\|z\|_1 = 1\) as the probability vector representing the likelihood of states \(x \in \{0,1\}^p\) to have components \(z_k = P(x = x_k)\)
for $k = 0, 1, \ldots, 2^p - 1$. Let $u(x) = \{ i : x_i = 1 \}$ be the indexes of components in $x$ equal to 1 and define the sequence of indicator functions

$$T_k(x) = 1\{u(x) \subseteq u(x_k)\}.$$  

The moments $m_k$, for $k \in \{1, 2, \ldots, 2^p - 1\}$, of $z$ are enumerated as

$$m_k = \mathbb{E}_z [T_k(x)] = \sum_{j=0}^{2^p-1} z_j T_k(x_j)$$

By the fact that $z$ is a probability vector, it is straightforward that $m_0 = 1$. We construct the $2^p \times (2^p - 1)$ zero-indexed matrix $T$ with elements

$$T_{jk} = T_k(x_j)$$

and note that the moments $m = (m_1, \ldots, m_{2^p-1})' \in [0, 1]^{2^p-1}$ of $z$ satisfy

$$m = T'z.$$  

This illustrates the bijection between the state probability vector $z$ and the moment sequence $m$.

Provided that $z$ is strictly positive, the statistical distribution of states $x \in \{0, 1\}^p$ may be represented as the Gibbs distribution

$$z_k = P(x = x_k|\gamma) = \exp\left\{ \sum_{j=1}^{2^p-1} \gamma_j T_j(x_k) - \log(Z(\gamma)) \right\}, \quad (1)$$

for some $\gamma \in \mathbb{R}^{2^p-1}$, where $Z(\gamma)$ is the partition function

$$Z(\gamma) = \sum_{k=0}^{2^p-1} \exp\left\{ \sum_{j=1}^{2^p-1} \gamma_j T_j(x_k) \right\},$$

e.g. normalizing constant of the probability distribution. With $\exp\{\cdot\}$ as the element-wise exponential function, we have

$$z = \exp\{T\gamma - \log(Z(\gamma))1\}. \quad (2)$$

Define $g : \mathbb{R}^{2^p-1} \mapsto \mathbb{R}^{2^p}$ as $g(\gamma) = \exp\{T\gamma - \log(Z(\gamma))1\}$ and note that the columns $T$ are the basis vectors of the nonlinear transformation $g(\cdot)$ from $\mathbb{R}^{2^p-1}$ to $[0, 1]^{2^p}$.

A low-rank approximation of the probability vector $z$ will exploit any redundancy in the moment sequence $m$. We classify moments of $z$ according to the number $|u(x)|$ of active states, e.g. number of ones in $x$. To that end, define the increasing sequence of index sets

$$I_d = \{ k : |u(x_k)| \leq d \},$$

11
for $d \in \{1, 2, \ldots, p\}$ and note that $|I_d| = t_d$, where $t_d = \sum_{i=1}^{d} \binom{p}{i}$. Correspondingly, let $T_d$ be the $2^p \times t_d$ matrix constituted by the columns of $T$ corresponding to the indexes in $I_d$, for $d \in \{1, 2, \ldots, p\}$. Finally, we define the increasing subsets $U_d \subseteq [0, 1]^{2^p}$ according to

$$U_d = \{ y \in \mathbb{R}^{2^p} : \exists \beta \in \mathbb{R}^{t_d} \text{ s.t. } y = g_d(\beta) \},$$

where $g_d(\beta) = \exp \{ T_d \beta - \log(Z(\beta)) \mathbf{1} \}$, for $d \in \{1, 2, \ldots, p\}$. That is, $U_d$ is the image of all $\beta \in \mathbb{R}^{t_d}$ under the non-linear transformation $g_d(\beta)$.

**Definition 1.** The binary system with states $x \in \{0, 1\}^p$ and state probability vector $z \in [0, 1]^{2^p}$ is a $d$th-order system if

$$z \in U_d \quad \text{and} \quad z \notin U_{d-1},$$

for some $d \in \{2, 3, \ldots, p\}$.

**Model Identification**

Let $\beta \in \mathbb{R}^{t_d}$ and define $z_d = g_d(\beta)$ for some $d \in \{2, 3, \ldots, p\}$. The cross entropy $H(z, z_d)$ from $z_d$ to $z$ is

$$H(z, z_d) = -\sum_{k=0}^{2^p-1} z_k \log(z_{dk})$$

$$= -z' [T_d \beta - \log(Z(\beta)) \mathbf{1}]$$

$$= -m_d \cdot \beta + \log(Z(\beta)),$$

where $m_d = T_d' z$. If $z_d = z$ then there is no statistical information lost in representing $z$ with $z_d = g_d(\beta)$. In this case $H(z, z_d) = H(z)$ is the entropy of $z$ and the predictability of states encoded in $z$ has been captured in the $d$th-order statistical representation $z_d = g_d(\beta)$. The gradient

$$\frac{d}{d\beta} H(z, z_d) = \frac{d}{d\beta} H(z, g_d(\beta)) = -T_d' z + \frac{d}{d\beta} \log(Z(\beta)).$$

Since $\frac{d}{d\beta} Z(\beta) = \mathbb{E}_{z_d} [T_j(x)]$, the gradient takes the form

$$\frac{d}{d\beta} H(z, z_d) = T_d'(g_d(\beta) - z).$$

The system of equations

$$0 = T_d'(g_d(\beta) - z)$$

are precisely those which identify the $d$th-order maximum entropy distribution approximating the probability vector $z$. That is, all moments of at most $d$th-order are conserved in the solution probability vector $\hat{z}_d$. The remaining moments are left unconstrained.
The \(d^{th}\)-order low-rank approximation \(\tilde{z}_d = g_d(\hat{\beta})\) of \(z\), for solution vector \(\hat{\beta} \in \mathbb{R}^{t_d}\) and based on the column vectors of \(T_d\) and subject to the constraint \(|\tilde{z}_d|_1 = 1\), satisfies the system in Eq. (3). Since the Hessian matrix of \(H(z, g_d(\beta))\) is positive definite for \(\beta \in \mathbb{R}^{t_d}\), the solution \(\hat{\beta}\) to the system in Eq. (3) obtains the minimal cross entropy to \(z\) from any probability distribution \(z_d \in U_d\) and may be computed with a gradient descent procedure, see the Supplementary Information.

If \(z \in U_d\) then \(z_d = z\) and the order of the system with state probability distribution vector \(z\) is at most \(d^{th}\)-order. More precisely, the Kullback-Leibler divergence

\[
\text{KL}(z|\tilde{z}_d) = - \sum_{k=0}^{2^p-1} z_k \log(\tilde{z}_k) - \log(z_k),
\]

the relative entropy from the \(d^{th}\)-order probability distribution \(\tilde{z}_d\) approximating the probability vector \(z\), equals zero precisely when \(\tilde{z}_d = z\). In reference to Definition 1, if \(\text{KL}(z|\tilde{z}_d) = 0\) for some \(d \in \{1, 2, \ldots, p\}\) then \(z \in U_d\) and, more generally, \(\text{KL}(z|\tilde{z}_d') = 0\) for \(d' \in \{d, d+1, \ldots, p\}\).

### Approximating the Probability Distribution of the Data

In practice, samples \(x^{(1)}, x^{(2)}, \ldots, x^{(n)}\) are observed instances of the infant gut ecosystem and \(z\) is directly estimated with the empirical probability distribution \(\tilde{z}\) of relative frequencies, typically, many of which are zero. On account of these zero components of \(\tilde{z}\), the probability vector cannot be expressed in the form of Eq. (2) and, consequently, \(\text{KL}(\tilde{z}|g_d(\beta)) > 0\) for all \(\beta \in \mathbb{R}^{t_d}\) and \(d \in \{1, 2, \ldots, p\}\). Moreover, let \(\tilde{m}_d = T_d'\tilde{z}\) be the vector of sample moments up to \(d^{th}\)-order and note that if there exists a zero component in \(\tilde{m}_d\) then the system in Eq. (3) does not have a solution \(\hat{\beta} \in \mathbb{R}^{t_d}\). These observations indicate that, in order to estimate the \(d^{th}\)-order solution \(\hat{\beta} \in \mathbb{R}^{t_d}\) with a method modified from that of the preceding section, it is necessary that the first \(t_d\) elements of \(\tilde{z}\) are strictly positive. More generally, we define the regularized \(d^{th}\)-order cross entropy optimization problem:

\[
\text{minimize} \quad H(\tilde{z}, g_d(\beta)) + \lambda \beta' W \beta \quad \text{for} \quad \beta \in \mathbb{R}^{t_d},
\]

for some hyperparameter \(\lambda > 0\), where \(W = \text{diag}(w)\) and \(w \in \mathbb{R}^{t_d}\) with element \(w_j = \frac{1}{(d - 1)(p_j)}\) for \(j \in \{p + 1, \ldots, t_d\}\) and \(w_j = 0\) for \(j \in \{1, \ldots, p\}\) so that \(|w|_1 = 1\) and that the weights \(w\) are such that the regularization \(\lambda \beta' W \beta\) is order-wise equally applied to the value of the objective function, except to the first-order which is not regularized.

This regularized \(d^{th}\)-order objective function in Eq. (4) has corresponding gradient

\[
\frac{d}{d\beta} \left[H(\tilde{z}, g_d(\beta)) + \lambda \beta' W \beta\right] = T_d' [g_d(\beta) - \tilde{z}] + 2\lambda W \beta,
\]
For some \( d \in \{2, 3, \ldots, p\} \) and \( \lambda > 0 \), the solution
\[
\hat{\beta}_{d\lambda} = \arg \min_{\beta \in \mathbb{R}^{d}} \ H(\tilde{z}, g_{d}(\beta)) + \lambda \beta' \mathbf{W} \beta
\] (5)
to the regularized optimization Eq. (4) is readily obtained via gradient descent
and the regularization parameter \( \lambda > 0 \) is selected via leave-one-out cross-
validation, see the Supplementary Information. We denote \( \hat{z}_{d} = g_{d}(\hat{\beta}_{d\lambda}) \), e.g.
without the \( \lambda \) subscript, as the estimated probability distribution that results
from an automatically selected \( \lambda \) via the cross validation subroutine.

### Information Content of Higher-order Associations

We seek to quantify the statistical information attributable to \( d^{th} \)-order associations in the statistical distribution \( \hat{z}_{d} = g_{d}(\beta_{d\lambda}) \), for \( d \in \{2, 3, \ldots, p\} \).
Understanding entropy as a measure of the lack of predictability of states based
a statistical representation of their respective likelihoods, the Kullback-Leibler
divergence from the estimated \( d^{th} \)-other probability distribution \( \hat{z}_{d} \) to the
empirical probability distribution of the data \( \tilde{z} \) is the information discrepancy
\( I_{d} = \text{KL}(\tilde{z} || \hat{z}_{d}) \). For values of \( d \in \{1, 2, \ldots, p\} \), we define the sequence
\[
I_{d} = -\sum_{j=0}^{2^{p}-1} \tilde{z}_{j} [\log(\hat{z}_{dj}) - \log(\tilde{z}_{j})]
= H(\tilde{z}, \hat{z}_{d}) - H(\hat{z}),
\]
where \( H(\hat{z}) = -\sum_{k=0}^{2^{p}-1} \hat{z}_{k} \log(\hat{z}_{k}) \) is the entropy of \( \hat{z} \). Note that
\[
I_{1} = [I_{1} - I_{p}] + I_{p}
\]
decomposes into (i) \( I_{1} - I_{p} \), a global measure of statistical dependence
known as the mutual information from the first-order \( \hat{z}_{1} \) distribution to the
fully-specified \( \hat{z}_{p} \) probability distribution, and (ii) \( I_{p} \) is the lack of fit in approximating \( \hat{z} \) with \( \hat{z}_{p} \) on account of the inexpressibility of \( \hat{z} \) in the form of Eq. (2). The information gained from approximating \( \hat{z} \) with the estimated \( d^{th} \)-order
distribution \( \hat{z}_{d} \) relative to approximating \( \hat{z} \) with the \( (d - 1)^{th} \)-order distribution
\( \hat{z}_{d-1} \) is
\[
I_{d-1} - I_{d} = -\sum_{j=0}^{2^{p}-1} \tilde{z}_{j} [\log(\hat{z}_{(d-1)j}) - \hat{z}_{dj}],
\]
for \( d \in \{2, 3, \ldots, p\} \), and is interpreted as the quantity of statistical information encoded by \( d^{th} \) order statistical associations detected in the system with empirical probability distribution \( \hat{z} \). This information quantity is a proportion
\[
R_{d} = \left( \frac{I_{d-1} - I_{d}}{I_{1} - I_{p}} \right) \in [0, 1],
\] (6)
of the total quantity of detected statistical information contained in associations of all orders, for \( d \in \{2, 3, \ldots, p\} \). Clearly, \( \sum_{d=2}^{p} R_d = 1 \). Finally, let

\[
R_* = \frac{I_1 - I_p}{I_1} = 1 - \frac{I_p}{I_1}
\]

be the measure of the quality-of-fit in the approximating \( \hat{z} \) with the distributions \( \hat{z}_d \) of the form in Eq. (2), for any \( d \in \{2, 3, \ldots, p\} \). This quantity may be intuited as an information-analogue to the coefficient of determination in linear models and its value is attributed to (i) the inexpressibility of empirical probability distribution in the form of Eq. (2) and (ii) the regularization involving \( \lambda > 0 \). The contributions of each type to the \( R_* \) statistic may be diminished in the large sample limit. For example, provided that the unknown, true probability distribution of the states in the system is stationary and strictly positive, e.g. is representable in the form of Eq. (2), a sufficiently-large sample size is to include observations on all possible states so that \( R_* = 1 \) with high probability. On the other hand, the leave-one-out cross validation procedure have, in the large sample limit, more similarly distributed training data sets and a correspondingly smaller \( \lambda \) value on problems of fixed size.

Acknowledgments

The authors are grateful to the children and families that made this study possible and to the staff of the New Hampshire Birth Cohort Study.

References

[1] Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-Bacterial Mutualism in the Human Intestine. Science. 2005;307(5717):1915–1920.

[2] Bäumler AJ, Sperandio V. Interactions between the microbiota and pathogenic bacteria in the gut. Nature. 2016;535:85–93.

[3] Trosvik P, Rudi K, Straetkvern KO, Jakobsen KS, Naes T, Stenseth NC. Web of ecological interactions in an experimental gut microbiota. Environmental Microbiology. 2010;10:2677–2687.

[4] Kährström CT, Pariente N, Weiss U. Intestinal Microbiota in health and disease. Nature. 2016;535:47–47.

[5] Haque SZ, Haque M. The ecological community of commensal, symbiotic, and pathogenic gastrointestinal microorganisms - an appraisal. Clinical and experimental gastroenterology. 2017;10:91–103.

[6] Shoaei S, Karlsson F, Mardinoglu A, Nookaew I, Bordel S, Nielsen J. Understanding the interactions between bacteria in the human gut through metabolic modeling. Scientific reports. 2013;3:2532–2532.
[7] Freilich S, Kreimer A, Meilijson I, Gophna U, Sharan R, Ruppin E. The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic Acids Research. 2010;38:3857–3868.

[8] Layeghifard M, Hwang DM, Guttman DS. Disentangling Interactions in the Microbiome: a Network Perspective. Trends in Microbiology. 2017;25:217–228.

[9] Ramette A. Multivariate analyses in microbial ecology. FEMS Microbiology Ecology. 2007;62(2):142–160.

[10] Ives AR, Carpenter SR. Stability and diversity of ecosystems. Science. 2007;317:58–62.

[11] Bairey E, Kelsic ED, Kishony R. High-order species interactions shape ecosystem diversity. Nature Communications. 2016;7:12285.

[12] Mandakovic D, Rojas C, Maldonado J, Latorre M, Travisany D, Delage E, et al. Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. Scientific Reports. 2018;8:5875.

[13] Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. Journal of Clinical Microbiology. 2005;43:5721–5732.

[14] Aitchison J. A new approach to null correlations of proportions. Journal of the International Association for Mathematical Geology. 1981;13:175–189.

[15] Poudel R, Jumpponen A, Schlatter DC, Paulitz TC, Gardener BBM, Kinkel LL, et al. Microbiome Networks: A Systems Framework for Identifying Candidate Microbial Assemblages for Disease Management. Phytopathology. 2016;106:1083–1096.

[16] Faust K, Raes J. Microbial interactions: from networks to models. Nature Reviews Microbiology. 2012;10:538–550.

[17] van den Bergh MR, Biesbrock G, Rossen JWA, de Steenhuijsen Piters WAA, Bosch AATM, van Gils EJM, et al. Associations between Pathogens in the Upper Respiratory Tract of Young Children: interplay between Viruses and Bacteria. PLoS ONE. 2012;7:e4771.

[18] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002;298(5594):824–827.

[19] Alon U. Network motifs: theory and experimental approaches. Nature Reviews Genetics. 2007;8:450–461.
[20] Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial Co-occurrence Relationships in the Human Microbiome. PLoS Computational Biology. 2012;8:e1002606.

[21] Abrams PA. Arguments in Favor of Higher Order Interactions. The American Naturalist. 1983;121(6):887–891.

[22] Billick I, Case TJ. Higher Order Interactions in Ecological Communities: What Are They and How Can They be Detected? Ecology. 1994;75(6):1530–1543.

[23] Case TJ, Bender EA. Testing for higher order interactions. American Naturalist. 1981;118:920–929.

[24] Grilli J, Barabás G, Michalska-Smith MJ, Allesina S. Higher-order interactions stabilize dynamics in competitive network models. Nature. 2017;548:210–213.

[25] Levine J, Bascompte J, Adler P, Allesina S. Beyond pairwise coexistence: biodiversity maintenance in complex ecological communities. Nature. 2017;546:56–64.

[26] Friedman J, Higgins LM, Gore J.;

[27] Tsai KN, Lin SH, Liu WC, Wang D. Inferring microbial interaction network from microbiome data using RMN algorithm. BMC Systems Biology. 2015; p. 54.

[28] Chen I, Kelkar YD, Gu Y, Zhou J, Qiu X, Wu H. High-dimensional linear state space models for dynamic microbial interaction networks. PLoS One. 2017;12:e0187822.

[29] Kelsic ED, Zhao J, Vetsigian K, Kishony R. Counteraction of antibiotic production and degradation stabilizes microbial communities. Nature. 2015;521:516–519.

[30] Beckerman AP, Uriarte M, Schmitz OJ. Experimental evidence for a behavior-mediated trophic cascade in a terrestrial food chain. Proceedings of the National Academy of Sciences of the USA. 1997;94:10735–10738.

[31] Poisot T, Stouffer DB, Gravel D. Beyond species: why ecological interactions vary through space and time. Oikos. 2015;124:243–251.

[32] Wootton JT. The nature and consequences of indirect effects in ecological communities. Annual Review Ecology and Systematics. 1994;25:443–466.

[33] Gould AL, Zhang V, Lamberti L, Jones EW, Obadia B, Korasidis N, et al. Microbiome interactions shape host fitness. Proceedings of the National Academy of Sciences. 2018;115(51):E11951–E11960.
Gradient Descent

The gradient descent update

$$\beta^{(1)} = \beta^{(0)} - T_d \left[ g_d(\beta^{(0)}) - z \right]$$  (7)
(19) is our preferred method for numerical optimization in this context. Strict positivity of \( z \) is sufficient for the above iteration to converge to the optimum \( \hat{\beta}_d \in \mathbb{R}^{t_d} \) such that \( H(z, \hat{z}_d) < H(z, z_d) \) for any \( z_d \in [0,1]^{2^p} \), where \( \hat{z}_d = g_d(\beta_d) \).

For general \( \beta \in \mathbb{R}^{t_d} \) and \( z_p = g_d(\beta) \), the gradient
\[
\frac{d}{d\beta} g_d(\beta) = \frac{d}{d\beta} \exp\{T_d \beta - \log(Z(\beta))1\} = [T_d' - T_d' \cdot z_p \cdot 1'] \text{diag}(z_p) = T_d' [I_{2^p} - z_p \cdot 1'] \text{diag}(z_p),
\]
which implies that
\[
\frac{d^2}{d\beta d\beta'} H(z, g_d(\beta)) = \frac{d}{d\beta} [T_d' g_d(\beta) - z] = T_d' [I_{2^p} - z_p \cdot 1'] \text{diag}(z_d) T_d = T_d' [\text{diag}(z_d) - \text{diag}(z_p) J_{2^p} \text{diag}(z_p)] T_d = T_d' R T_d,
\]
where \( R = \text{diag}(z_d) - \text{diag}(z_p) J_{2^p} \text{diag}(z_p) \). Note that the \( k^{th} \) diagonal element \( R_{kk} = z_k(1 - z_k) \) and the sum of the off diagonal elements of the \( k^{th} \) row of \( R \) is
\[
R_{k} = -z_{dk} \sum_{j=0, j\neq k}^{2^p-1} z_{dj} = z_{dk}^2 - z_{dk} \sum_{j=0}^{2^p-1} z_{dj} = z_{dk}(1 - z_{dk})
\]
which demonstrates that \( R \) is weakly diagonally dominant with non-negative diagonal elements and, thus, positive semi-definite. In particular, since \( R1 = 0 \) then \( v = 1 \) is the eigenvector of \( R \) with eigenvalue \( \lambda = 0 \). Since no column of \( T_d \) is proportional to \( 1 \) and the columns of \( T_d \) are linearly independent then it follows that \( T_d' RT_d \) is positive definite and \( H(z, g_d(\beta)) \) is a convex function of \( \beta \).

**Cross Validation**

Viewing \( \tilde{z} \) as a probability vector over the integers \( \mathcal{S}_p = \{0,1,\ldots,2^p - 1\} \), define \( K = \{k \in \mathcal{S}_p : \tilde{z}_k > 0\} \) as the integers corresponding to the binary representations of the observed state vectors obtained in the sample. Let \( \tilde{z}^{(k)} \in [0,1]^{2^p} \) be the empirical probability distribution vector obtained after removing one observations corresponding to some \( k \in K \). The \( j^{th} \) element of \( \tilde{z}^{(k)} \) is
\[
\tilde{z}^{(k)}_j = \frac{n}{n-1} \left( \tilde{z}_j - \frac{1}{n} 1\{j = k\} \right),
\]
for \( j \in \{0, 1, \ldots, 2^p - 1\} \). The probability vector \( \hat{z}_d^{(k)} = g_d(\hat{\beta}_{d\lambda}^{(k)}) \) estimate based on \( \hat{z}^{(k)} \) is identified by

\[
\hat{\beta}_{d\lambda}^{(k)} = \arg \min_{\beta \in \mathbb{R}^d} H(\hat{z}^{(k)}, g_d(\beta)) + \frac{\lambda}{2} \| w \circ \beta \|^2
\]  

(8)

Let \( z_{d\lambda k}^{(k)} \) be the \( k^{th} \) element of \( \hat{z}_{d\lambda}^{(k)} \) and define

\[
\hat{\lambda}_d = \arg \min_{\lambda > 0} - \sum_{k \in \mathcal{K}} \hat{z}_k \log \left( \frac{\hat{z}_{d\lambda k}^{(k)}}{\sum_{j \in \mathcal{K}} \hat{z}_{d\lambda j}^{(j)}} \right)
\]  

(9)

as the \( \lambda > 0 \) value which minimizes the loss function which is a generalization of cross entropy between the empirical probability vector \( \hat{z} \) and the collection of leave-one-out probability vector estimates \( \hat{z}_{d\lambda}^{(k)} \), for \( k \in \mathcal{K} \). The normalization factor is included to standardize the loss function. The regularized parameter estimate \( \hat{\beta}_{d\lambda} \) identifies the regularized \( d^{th} \)-order maximum entropy distribution \( \hat{z}_{d\lambda} = g_d(\hat{\beta}_{d\lambda}) \). Selecting \( \hat{\lambda}_d \) is automatic and, accordingly, we suppress the \( \lambda \) notation in the \( d^{th} \)-order maximum entropy distribution \( \hat{z}_d = \hat{z}_{d\lambda} \).

**Examples of Three-Variable Non-linear Associations**

In the following, we suppose that the binary variables are jointly distributed as

\[
\mathbb{P}(x_1 = 0, x_2 = 1) = p_{00}
\]

\[
\mathbb{P}(x_1 = x_1', x_2 = x_2') = \frac{1}{3} (1 - p_{00})
\]

for \((x_1', x_2') \neq (0, 0)\). Derived from this distribution are the probabilities

\[
\mathbb{P}(x_1 = 1) = \frac{2}{3} (1 - p_{00})
\]

\[
\mathbb{P}(x_2 = 1) = \frac{2}{3} (1 - p_{00})
\]

\[
\mathbb{P}(x_1 x_2 = 1) = \frac{1}{3} (1 - p_{00}).
\]

Note that the covariance

\[
\text{Cov}(x_1, x_2) = \frac{1}{3} (1 - p_{00}) - \frac{4}{9} (1 - p_{00})^2
\]

\[
= \frac{1}{3} (1 - p_{00}) \left[ 1 - \frac{4}{3} (1 - p_{00}) \right]
\]

\[
= \frac{1}{9} (1 - p_{00}) [4p_{00} - 1].
\]
We set $p_{00} = 2/5$ in which case $\text{Cor}(x_1, x_2) > 0$. In the following we determine the coefficients of the second- and third-order maximum entropy distributions and subsequently compute the quantity of statistical information attributable to the corresponding associations.

**The xor Relationship**

We define $x_3 = x_1 + x_2 - 2x_1x_2$ so that $x_3 = 1$ if and only if exactly one of $x_1 = 1$ or $x_2 = 1$ is true. In this case, the joint distribution of $x_1, x_2, x_3$ is recorded as

| $x_1$ | $x_2$ | $x_3$ | $z$ |
|-------|-------|-------|-----|
| 0     | 0     | 0     | $p_{00}$ |
| 1     | 0     | 0     | 0   |
| 0     | 1     | 0     | 0   |
| 1     | 1     | 0     | $(1 - p_{00})/3$ |
| 0     | 0     | 1     | 0   |
| 1     | 0     | 1     | $(1 - p_{00})/3$ |
| 0     | 1     | 1     | $(1 - p_{00})/3$ |
| 1     | 1     | 1     | 0   |

Note that

$$P(x_3 = 1) = \frac{2}{3}(1 - p_{00})$$

$$P(x_1x_3 = 1) = P(x_2x_3 = 1) = \frac{1}{3}(1 - p_{00})$$

so that the collection of $x_1, x_2, x_3$ have equal individual success probabilities and are equi-correlated. This implies that the second-order maximum entropy model is of the form

$$P(\mathbf{x}) \propto \exp\{\alpha (x_1 + x_2 + x_2) + \beta (x_1x_2 + x_1x_3 + x_2x_3)\}$$

$$= \exp\{(2\alpha + \beta)(x_1 + x_2 - x_1x_2)\},$$

by applying $x_3 = x_1 + x_2 - 2x_1x_2$. Finally, note that $\text{max}\{x_1, x_2\} = x_1 + x_2 - x_1x_2$.

The matrix of basis vectors

$$T_2 = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}$$
leads to the moment equation

\[
\frac{1}{3} \left(1 - p_{00}\right) \begin{pmatrix} 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{Z(\alpha, \beta)} T_2' \begin{pmatrix} \exp\{2\alpha + \beta\} \\ \exp\{2\alpha + \beta\} \\ \exp\{2\alpha + \beta\} \\ \exp\{2\alpha + \beta\} \\ \exp\{2\alpha + \beta\} \\ \exp\{2\alpha + \beta\} \end{pmatrix}
\]

We exploited the symmetry of the problem to designate \(\alpha = \alpha_1 = \alpha_2 = \alpha_3\) and \(\beta = \beta_1 = \beta_2 = \beta_3\) as outlined above.

\[
\frac{1}{3} \left(1 - p_{00}\right) \begin{pmatrix} 2 \\ 2 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{Z(\alpha, \beta)} \begin{pmatrix} \exp\{\alpha\} + 2\exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta\} \\ \exp\{\alpha\} + 2\exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta\} \\ \exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta\} \\ \exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta\} \\ \exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta\} \end{pmatrix}
\]

Comparing the first and fourth equations

\[
\exp\{\alpha\} = \exp\{3\alpha + 3\beta\}
\]

\[
\alpha = 3\alpha + 3\beta
\]

\[
\alpha = -\frac{3}{2}\beta.
\]

The partition function

\[
Z(\alpha, \beta) = 1 + 3\exp\{\alpha\} + 3\exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta\}
\]

\[
= 1 + 4\exp\{-3\beta/2\} + 3\exp\{-2\beta\}.
\]

Solving the fourth equation

\[
\frac{1}{3} \left(1 - p_{00}\right) = \frac{\exp\{-3\beta/2\} + \exp\{-2\beta\}}{1 + 4\exp\{-3\beta/2\} + 3\exp\{-2\beta\}}.
\]

If \(p_{00} = 2/5\) then \(\beta \approx 0.605\) and, accordingly, \(\alpha \approx -0.908\).

Note that the cross entropy from the maximum entropy model to the true distribution is

\[
H(\mathbf{z}, \mathbf{z}_2) = -\left[\frac{3}{5} (2\alpha + \beta) - \log(Z)\right]
\]

\[
= \frac{6}{5}\beta + \log(1 + 4\exp\{-3\beta/2\} + 3\exp\{-2\beta\})
\]

\[
\approx 1.9812
\]
Since the entropy

\[ H(z) = -\frac{2}{5} \log\left(\frac{2}{5}\right) - \frac{3}{5} \log\left(\frac{1}{5}\right) \]

\[ \approx 1.3321 \]

which implies that \( \text{KL}(z|z_2) \approx 0.6490 \) so that there is additional statistical information ascribed to associations of orders greater than second-order.

In terms of the third-order model, the infeasible system of moment equations becomes

\[
\frac{1}{3}(1 - p_{00}) \begin{pmatrix} 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{Z(\alpha, \beta, \gamma)} T_3' \begin{pmatrix} 1 \\ \exp\{\alpha\} \\ \exp\{\alpha\} \\ \exp\{2\alpha + \beta\} \\ \exp\{\alpha\} \\ \exp\{2\alpha + \beta\} \\ \exp\{2\alpha + \beta\} \end{pmatrix}
\]

\[
= \frac{1}{Z(\alpha, \beta, \gamma)} \begin{pmatrix} \exp\{\alpha\} + 2 \exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta + \gamma\} \\ \exp\{\alpha\} + 2 \exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta + \gamma\} \\ \exp\{\alpha\} + 2 \exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta + \gamma\} \\ \exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta + \gamma\} \\ \exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta + \gamma\} \\ \exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta + \gamma\} \end{pmatrix}
\]

Considering the first and fourth equations,

\[ \exp\{\alpha\} = \exp\{3\alpha + 3\beta + \gamma\} \]

\[ \alpha = 3\alpha + 3\beta + \gamma \]

\[ \alpha = -\frac{3}{2}\beta - \frac{1}{2}\gamma. \]

Note that

\[ \exp\{3\alpha + 3\beta + \gamma\} = \exp\{-(3\beta + \gamma)/2\} \]

which implies that the solution to the third-order moment equations does not exist in the real numbers. Never-the-less, in the limit \( 3\beta + \gamma \to \infty \). Additionally, since \( \alpha = -(3\beta + \gamma)/2 \) then \( \alpha \to -\infty \). The partition function

\[ Z(\alpha, \beta, \gamma) = 1 + 3 \exp\{\alpha\} + 3 \exp\{2\alpha + \beta\} + \exp\{3\alpha + 3\beta + \gamma\} \]

\[ = 1 + 3 \exp\{-(2\beta + \gamma)/2\} + 4 \exp\{-(3\beta + \gamma)/2\} \]

All of the equations are equivalent to

\[
\frac{1}{3}(1 - p_{00}) = \frac{\exp\{-(3\beta + \gamma)/2\} + \exp\{-(2\beta + \gamma)\}}{1 + 4 \exp\{-(3\beta + \gamma)/2\} + 3 \exp\{-(2\beta + \gamma)\}}\]

\[
\to \frac{\exp\{-(2\beta + \gamma)\}}{1 + 3 \exp\{-(2\beta + \gamma)\}}.
\]
as $3\beta + \gamma \to \infty$. With $p_{00} = 2/5$ then $2\beta + \gamma = \log(2)$. Since
\[
\alpha = -\frac{3}{2}\beta - \frac{1}{2}\gamma \\
= -\frac{1}{2}[3\beta + \gamma] \\
= -\frac{1}{2}\beta - \frac{1}{2}[2\beta + \gamma] \\
\to -\frac{1}{2}\beta - \frac{\log(2)}{2},
\]
which means that $\beta \sim -2\alpha \to \infty$. Finally,
\[
\log(2) = 2\beta + \gamma \\
\log(2) - 2\beta = \gamma
\]
which means $\gamma \to -\infty$. The cross entropy
\[
H(z, z_3) = -\left[\frac{3}{5}(2\alpha + \beta) - \log(1 + 3\exp\{-2\beta + \gamma\})\right] \\
= -\frac{3}{5}(-2\beta - \gamma) + \log\left(\frac{5}{2}\right) \\
= \frac{3}{5}\log(2) + \log\left(\frac{5}{2}\right) \\
\approx 1.3321 \\
= H(z).
\]
Certainly this equality achieved only in the limit and a numerical evaluation requires regularization.

**Regularization**

The gradient of
\[
F_\lambda(\beta) = H(z, g_d(\beta)) + \lambda \beta^T W \beta,
\]
as written in Equation [4] in the main article, is
\[
\frac{d}{d\beta} [F_\lambda(\beta)] = T_d' [g_d(\beta) - z] + 2\lambda W \beta,
\]
which leads to the Newton update
\[
\beta^{(1)} = \beta^{(0)} - \left(T_d' [g_d(\beta^{(0)}) - z] + 2\lambda W \beta^{(0)}\right)
\]
Take the last of the third-order moment equations above and note that its regularized counterpart is
\[
0 = \exp\{-3(3\beta + \gamma)/2\} + 2\lambda \gamma/3,
\]
where $\gamma$ is the final element of the $\beta$ vector and the $W_{7,7} = 1/3$, which illustrates the finiteness of the solution.
The *max*, *relaxed or*, and *independence* Models

The joint distribution of the three binary variables $x_1, x_2, x_3$ under the *max* and *relaxed or* models begin, as before, with the joint distribution of $x_1, x_2$ as

\[
P(x_1 = 0, x_2 = 1) = p_{00} \\
\mathbb{P}(x_1 = x_1', x_2 = x_2') = \frac{1}{3}(1 - p_{00})
\]

for $(x_1', x_2') \neq (0, 0)$.

In the *max* model we define $x_3 = \max\{x_1, x_2\}$ so that

\[
\begin{array}{ccc|c}
  x_1 & x_2 & x_3 & z_{\text{max}} \\
  0 & 0 & 0 & p_{00} \\
  1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  0 & 0 & 1 & 0 \\
  1 & 0 & 1 & (1 - p_{00})/3 \\
  0 & 1 & 1 & (1 - p_{00})/3 \\
  1 & 1 & 1 & (1 - p_{00})/3 \\
\end{array}
\]

where as the *relaxed or* we define $x_3 = 0$ if $x_1 x_2 = 1$ and, otherwise, $x_3 \sim \text{Bernoulli}(1/2)$ so that

\[
\begin{array}{ccc|c}
  x_1 & x_2 & x_3 & z_{\text{relxor}} \\
  0 & 0 & 0 & \frac{1}{2} p_{00} \\
  1 & 0 & 0 & \frac{1}{6}(1 - p_{00}) \\
  0 & 1 & 0 & \frac{1}{6}(1 - p_{00}) \\
  1 & 1 & 0 & \frac{1}{6}(1 - p_{00}) \\
  0 & 0 & 1 & \frac{1}{2} p_{00} \\
  1 & 0 & 1 & \frac{1}{6}(1 - p_{00}) \\
  0 & 1 & 1 & \frac{1}{6}(1 - p_{00}) \\
  1 & 1 & 1 & 0. \\
\end{array}
\]

Residual Information

A plot in the format of Figure 2a, we provide the quantiles of the $R_*$ quality of fit statistic in Figure 3.
Fig. 3. Supplemental. Quantiles of the $R_*$ statistic on the quality of model fit to the data.