Academic Advances of the CTO, Vol. 1, Issue 2, 2017
Original Research Article

Vertical Dependency in Sequences of Categorical Random Variables

Rachel Traylor*, Ph.D. and Jason Hathcock

Abstract
This paper develops a more general theory of sequences of dependent categorical random variables, extending the works of Korzeniowski (2013) and Traylor (2017) that studied first-kind dependency in sequences of Bernoulli and categorical random variables, respectively. A more natural form of dependency, sequential dependency, is defined and shown to retain the property of identically distributed but dependent elements in the sequence. The cross-covariance of sequentially dependent categorical random variables is proven to decrease exponentially in the dependency coefficient δ as the distance between the variables in the sequence increases. We then generalize the notion of *vertical dependency* to describe the relationship between a categorical random variable in a sequence and its predecessors, and define a class of generating functions for such dependency structures. The main result of the paper is that any sequence of dependent categorical random variables generated from a function in the class $\mathscr{C}_\delta$ that is *dependency continuous* yields identically distributed but dependent random variables. Finally, a graphical interpretation is given and several examples from the generalized vertical dependency class are illustrated.

Keywords
categorical variables — correlation — dependency — probability theory

Contents

*Office of the CTO, Dell EMC

Introduction

Many statistical tools and distributions rely on the independence of a sequence or set of random variables. The sum of independent Bernoulli random variables yields a binomial random variable. More generally, the sum of independent categorical random variables yields a multinomial random variable. Independent Bernoulli trials also form the basis of the geometric and negative binomial distributions, though the focus is on the number of failures before the first (or rth success). [2] In data science, linear regression relies on independent and identically distributed (i.i.d.) error terms, just to name a few examples.

The necessity of independence filters throughout statistics and data science, although real data rarely is actually independent. Transformations of data to reduce multicollinearity (such as principal component analysis) are commonly used before applying predictive models that require assumptions of independence. This paper aims to continue building a formal foundation of dependency among sequences of random variables in order to extend these into generalized distributions that do not rely on mutual independence in order to better model the complex nature of real data. We build on the works of Korzeniowski [1] and Traylor [3] who both studied first-kind (FK) dependence for Bernoulli and categorical random variables, respectively, in order to define a general class of functions that generate dependent sequences of categorical random variables.

Section 1 gives a brief review of the original work by Korzeniowski [1] and Traylor [3]. In section 2, a new dependency structure, *sequential dependency* is introduced, and the cross-covariance matrix for two sequentially dependent categorical random variables is derived. Sequentially dependent categorical random variables are identically distributed but dependent. Section 3.1 generalized the notion of *vertical dependency structures* into a class that encapsulates both the first-kind (FK) dependence of Korzeniowski [1] and Traylor [3] and shows that all such sequences of dependent random variables are identically distributed. We also provide a graphical interpretation and illustrations of several examples of vertical dependency structures.

1. Background

We repeat a section from [3] in order to give a review of the original first-kind (FK) dependency created by Korzeniowski.

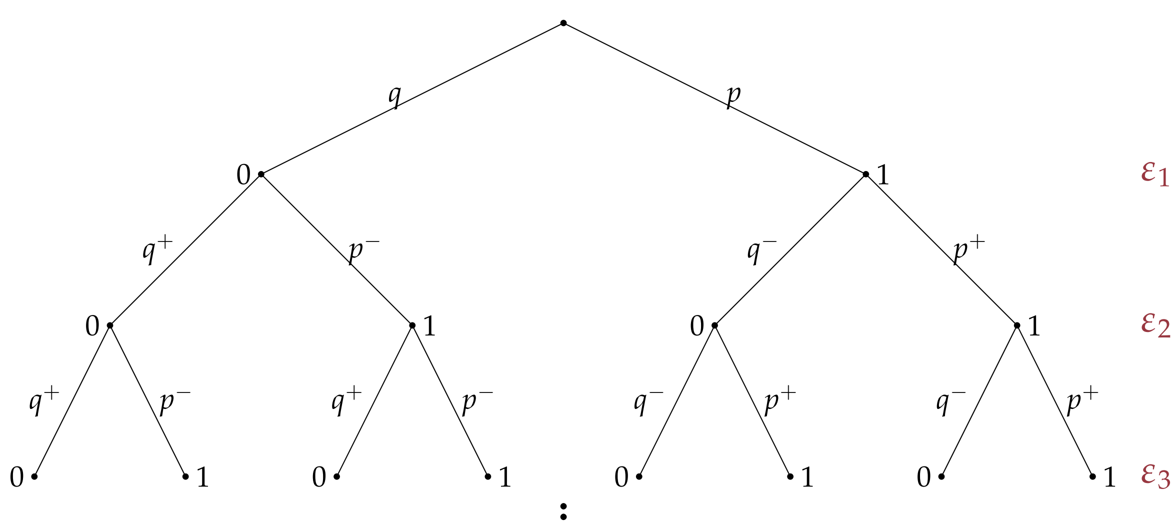

Figure 1. First Kind Dependence for Bernoulli Random Variables

Korzeniowski defined the notion of dependence in a way we will refer to here as *dependence of the first kind* (FK dependence). Suppose $(\varepsilon_1, \ldots, \varepsilon_N)$ is a sequence of Bernoulli random variables, and $P(\varepsilon_1 =$

$1) = p$. Then, for $\varepsilon_i, i \geq 2$, we weight the probability of each binary outcome toward the outcome of ε_1, adjusting the probabilities of the remaining outcomes accordingly.

Formally, let $0 \leq \delta \leq 1$, and $q = 1 - p$. Then define the following quantities

$$\begin{aligned} p^+ &:= P(\varepsilon_i = 1 | \varepsilon_1 = 1) = p + \delta q \qquad & p^- &:= P(\varepsilon_i = 0 | \varepsilon_1 = 1) = q - \delta q \\ q^+ &:= P(\varepsilon_i = 1 | \varepsilon_1 = 0) = p - \delta p \qquad & q^- &:= P(\varepsilon_i = 0 | \varepsilon_1 = 0) = q + \delta p \end{aligned} \tag{1}$$

Given the outcome i of ε_1, the probability of outcome i occurring in the subsequent Bernoulli variables $\varepsilon_2, \varepsilon_3, ..., \varepsilon_n$ is $p^+, i = 1$ or $q^+, i = 0$. The probability of the opposite outcome is then decreased to q^- and p^-, respectively.

Figure 1 illustrates the possible outcomes of a sequence of such dependent Bernoulli variables. Korzeniowski showed that, despite this conditional dependency, $P(\varepsilon_i = 1) = p \;\forall\, i$. That is, the sequence of Bernoulli variables is identically distributed, with correlation shown to be

$$Cor(\varepsilon_i, \varepsilon_j) = \begin{cases} \delta, & i = 1 \\ \delta^2, & i \neq j, \quad i, j \geq 2 \end{cases}$$

These identically distributed but correlated Bernoulli random variables yield a Generalized Binomial distribution with a similar form to the standard binomial distribution.

In [3], the concept of Bernoulli FK dependence was extended to categorical random variables. That is, given a sequence of categorical random variables with K categories, $P(\varepsilon_1 = i) = p_i, i = 1, ..., K$. $P(\varepsilon_j = i | \varepsilon_1 = i) = p_i^+ = p_i + \delta(1 - p_i)$, and $P(\varepsilon_j = k | \varepsilon_1 = i) = p_k^- = p_k - \delta p_k$, $i \neq k$, $k = 1, ..., K$. Traylor proved that FK dependent categorical random variables remained identically distributed, and showed that the cross-covariance matrix of categorical random variables has the same structure as the correlation between FK dependent Bernoulli random variables. In addition, the concept of a generalized binomial distribution was extended to a generalized multinomial distribution.

In the next section, we will explore a different type of dependency structure, *sequential dependency*.

2. Sequentially Dependent Categorical Random Variables

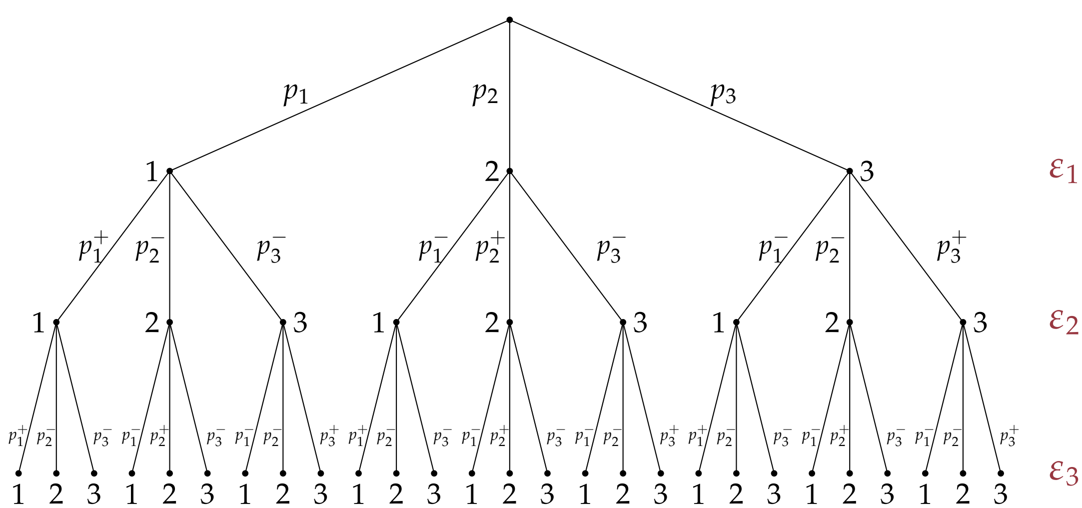

Figure 2. Probability Mass Flow of Sequentially Dependent Categorical Random Variables, $K = 3$.

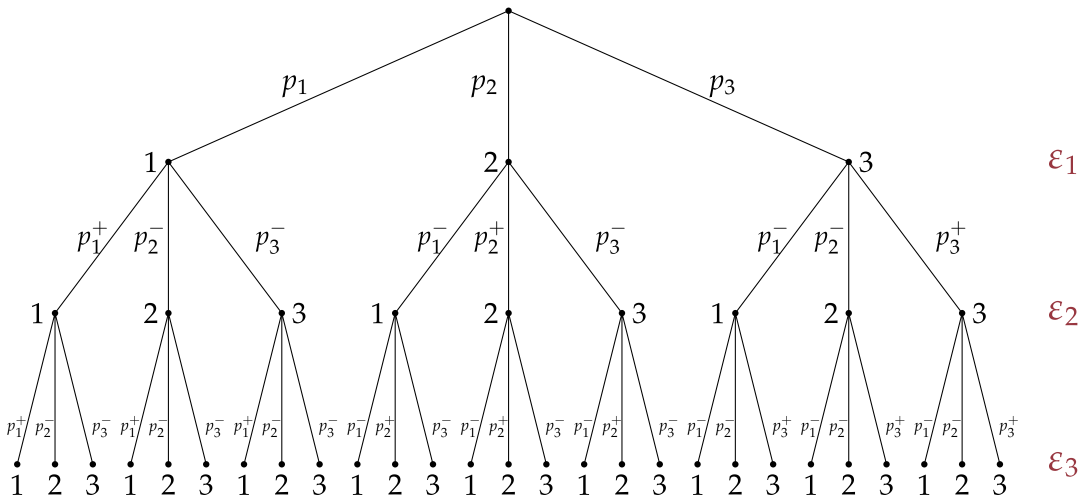

Figure 3. Probability Mass Flow of FK Dependent Categorical Random Variables, $K = 3$.

While FK dependence yielded some interesting results, a more realistic type of dependence is *sequential dependence*, where the outcome of a categorical random variable depends with coefficient δ on the outcome of the variable immediately preceeding it in the sequence. Put formally, if we let $\mathcal{F}_n = \{\varepsilon_1, ..., \varepsilon_{n-1}\}$, then $P(\varepsilon_n|\varepsilon_1, ..., \varepsilon_{n-1}) = P(\varepsilon_n|\varepsilon_{n-1}) \neq P(\varepsilon_n)$. That is, ε_n only has direct dependence on the previous variable ε_{n-1}. We keep the same weighting as for FK-dependence. That is,

$$P(\varepsilon_n = j|\varepsilon_{n-1} = j) = p_j^+ = p_j + \delta(1 - p_j), \qquad P(\varepsilon_n = j|\varepsilon_{n-1} = i) = p_j^- = p_j - \delta p_j; j = 1, ..., K, i \neq j \tag{2}$$

As a comparison, for FK dependence, $P(\varepsilon_n|\varepsilon_1, ..., \varepsilon_{n-1}) = P(\varepsilon_n|\varepsilon_1) \neq P(\varepsilon_n)$. That is, ε_n only has direct dependence on ε_1, and

$$P(\varepsilon_n = j|\varepsilon_1 = j) = p_j^+ = p_j + \delta(1 - p_j), \qquad P(\varepsilon_n = j|\varepsilon_1 = i) = p_j^- = p_j - \delta p_j; j = 1, ..., K, i \neq j \tag{3}$$

Let $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)$ be a sequence of categorical random variables of length n (either independent or dependent) where the number of categories for all ε_i is K. Denote Ω_n^K as the sample space of this random sequence. For example,

$$\Omega_3^3 = \{(1,1,1), (1,1,2), (1,1,3), (1,2,1), (1,2,2), ..., (3,3,1), (3,3,2), (3,3,3)\}$$

Dependency structures like FK-dependence and sequential dependence change the probability of a sequence ε of length n taking a particular $\omega = (\omega_1, ..., \omega_n) \in \Omega_n^K$. The probability of a particular $\omega \in \Omega_n^K$ is given by the dependency structure. For example, if the variables are independent, $P((1,2,1)) = p_1^2 p_2$. Under FK-dependence, $P((1,2,1)) = p_1 p_2^- p_1^+$, and under sequential dependence, $P((1,2,1)) = p_1 p_2^- p_1^-$. See Figures 2 and 3 for a comparison of the probability mass flows of sequential dependence and FK dependence. Sequentially dependent sequences of categorical random variables remain identically distributed but dependent, just like FK-dependent sequences. That is,

Lemma 1. *Let $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)$ be a sequentially dependent categorical sequence of length n with K categories. Then* $P(\varepsilon_j = i) = p_i; \quad i = 1, ..., K; \quad j = 1, ..., n; n \in \mathbb{N}$.

Proof. Fix $n \in \mathbb{N}$. Let $\Phi_n^K(i) = \{\omega \in \Omega_n^K : \omega_n = i\}, i = 1, ..., K$. Note that $|\Phi_n^K(i)| = K^{n-1}$. Then we may partition Ω_n^K using these $\Phi_n^K(i)$: $\Omega_n^K = \sqcup_{i=1}^K \Phi_n^K(i)$.[1]

[1]The union is disjoint

The event that $\varepsilon_n = i$ is given by $\bigcup_{\omega\in\Phi_n^K(i)}\omega$, and thus $P(\varepsilon_n = i) = P\left(\bigcup_{\omega\in\Phi_n^K(i)}\omega\right), i = 1,...,K$. Each sequence $\omega_j \in \Phi_n^K(i)$ has an associated probability $\pi_j = \prod_{l=1}^{n}\pi_{j_l}$, where π_{j_l} is the probability of the lth term in the sequence ω_{j_l}. Therefore,

$$P(\varepsilon_n = i) = \sum_{j=1}^{K^{n-1}}\pi_j = \sum_{j=1}^{K^{n-1}}\prod_{l=1}^{n}\pi_{j_l} \tag{4}$$

WLOG, assume $i = 1$. Then for $n = 2$, under sequential dependence,

$$P(\varepsilon_2 = 1) = p_1p_1^+ + p_2p_1^- + ... + p_Kp_1^- = p_1$$

by Lemma 1 of [3]. For $n = 3$,

$$\begin{aligned}P(\varepsilon_3 = 1) &= p_1p_1^+p_1^+ + p_1p_2^-p_1^- + p_1p_K^-p_1^- + ... + p_Kp_1^-p_1^+ + ... + p_Kp_K^+p_1^- \\ &= p_1^+\left(p_1p_1^+ + p_1^-\sum_{j\neq 1}p_j\right) + \sum_{i=2}^{K}p_1^-\left(p_ip_i^+ + p_i^-\sum_{j\neq i}p_j\right) \\ &= p_1^+p_1 + p_1^-\sum_{i=2}^{K}p_i \\ &= p_1\end{aligned}$$

It is clear that these hold true for $i = 2,...,K$. That is, $P(\varepsilon_2 = i) = p_i\ \forall\, i$ and $P(\varepsilon_3 = i) = p_i\ \forall\, i$. Now, assume that $P(\varepsilon_n = i) = p_i, i = 1,...,K$. Then, WLOG, we will show that $P(\varepsilon_{n+1} = 1) = p_1$. We have that $\Phi_{n+1}^K(1) = \Omega_n \times \{1\} = \left(\sqcup_{i=1}^K\Phi_n^K(i)\right)\times\{1\} = \sqcup_{i=1}^K\left(\Phi_n^K(i)\times\{1\}\right)$ Therefore,

$$\begin{aligned}P(\varepsilon_{n+1} = 1) &= P(\Phi_{n+1}^K(1)) \\ &= \sum_{i=1}^{K}P\left(\Phi_n^K(i)\times\{1\}\right) \\ &= p_1^+P(\Phi_n^K(i)) + p_1^-\sum_{i=2}^{K}P(\Phi_n^K(i)) \\ &= p_1^+p_1 + p_1^-\sum_{i=2}^{K}p_i \\ &= p_1\end{aligned}$$

A similar argument follows for $i = 2,....,K$ to conclude that for any n, $P(\varepsilon_n = i) = p_i$ under sequential dependence. □

2.1 Cross-Covariance Matrix

The $K \times K$ cross-covariance matrix $\Lambda^{m,n}$ of ε_m and ε_n in a sequentially dependent categorical sequence has entries $\Lambda_{i,j}^{m,n}$ that give the cross-covariance $Cov([\varepsilon_m = i], [\varepsilon_n = j])$, where $[\cdot]$ denotes an Iverson bracket. In the FK-dependent case, the entries of $\Lambda^{m,n}$ are given by [3]

$$\Lambda_{ij}^{1,n} = \begin{cases}\delta p_i(1-p_i), & i = j\\ -\delta p_ip_j, & i\neq j\end{cases}, n \geq 2, \text{ and } \Lambda_{ij}^{m,n} = \begin{cases}\delta^2 p_i(1-p_i), & i = j\\ -\delta^2 p_ip_j, & i\neq j\end{cases}, n > m, m \neq 1.$$

Thus, the cross covariance between any two ε_m and ε_n in the FK-dependent case is never smaller than δ^2 times the independent cross-covariance. In the sequentially dependent case, the cross-covariances of ε_m and ε_n decrease in powers of δ as the distance between the two variables in the sequence increases.

Theorem 1 (Cross-Covariance of Dependent Categorical Random Variables)**.** *Denote $\Lambda^{m,n}$ as the $K \times K$ cross-covariance matrix of ε_m and ε_n in a sequentially dependent sequence of categorical random variables of length N, $m \leq n$, and $n \leq N$, defined as $\Lambda^{m,n} = E[(\varepsilon_m - E[\varepsilon_m])(\varepsilon_n - E[\varepsilon_n])]$. Then the entries of the matrix are given by* $\Lambda^{m,n}_{ij} = \begin{cases} \delta^{n-m} p_i(1-p_i), & i = j \\ -\delta^{n-m} p_i p_j, & i \neq j \end{cases}$

The pairwise covariance between two Bernoulli variables in a sequentially dependent sequence is given in the following corollary

Corollary 1. *Denote $P(\varepsilon_i = 1) = p; i = 1,...,n$, and let $q = 1-p$. Under sequential dependence, $Cov(\varepsilon_m, \varepsilon_n) = pq\delta^{n-m}$.*

We move the proof of Theorem 1 to Section 5. We give some examples to illustrate.

Example 1 (Bernoulli Random Variables)**.** *If we want to find the covariance between ε_2 and ε_3, then we note that the set $S = \{\omega \in \Omega_3^2 : \omega_2 = 1, \omega_3 = 1\}$ is given by $S = \{(1,1,1),(0,1,1)\}$. $P(\varepsilon_2 = 1, \varepsilon_3 = 1) = P(S)$. Thus,*

$$\begin{aligned}
Cov(\varepsilon_2, \varepsilon_3) &= P(\varepsilon_2 = 1, \varepsilon_3 = 1) - P(\varepsilon_2 = 1)P(\varepsilon_3 = 1) \\
&= pp^+p^+ + qp^-p^+ - p^2 \\
&= p^+(pp^+ + qp^-) - p^2 \\
&= p(p^+ - p) \\
&= pq\delta
\end{aligned}$$

Similarly,

$$\begin{aligned}
Cov(\varepsilon_1, \varepsilon_3) &= P(\varepsilon_1 = 1, \varepsilon_3 = 1) - P(\varepsilon_1 = 1)P(\varepsilon_3 = 1) \\
&= (pp^+p^+ + pq^-p^-) - p^2 \\
&= p((p + \delta q)^2 + pq(1-\delta)^2) - p^2 \\
&= p(p^2 + pq + \delta^2 q^2 + \delta^2 pq) - p^2 \\
&= p(p(p+q) + \delta^2 q(q+p)) - p^2 \\
&= p(p + \delta^2 q) - p^2 \\
&= pq\delta^2
\end{aligned}$$

Example 2 (Categorical Random Variables)**.** *Suppose we have a sequence of categorical random variables, where $K = 3$. Then $[\varepsilon_m = i]$ is the Bernoulli random variable with the binary outcome of 1 if $\varepsilon_m = i$ and 0 if not. Thus, $Cov([\varepsilon_m = i], [\varepsilon_n = j]) = P(\varepsilon_m = i \vee \varepsilon_n = j) - P(\varepsilon_m = i)P(\varepsilon_n = j)$. We have shown that every ε_n in the sequence is identically distributed, so $P(\varepsilon_m = i) = p_i$ and $P(\varepsilon_n = j) = p_j$.*

$$\begin{aligned}
Cov([\varepsilon_2 = 1], [\varepsilon_3 = 1]) &= (p_1 p_1^+ p_1^+ + p_2 p_1^- p_1^+ + p_3 p_1^- p1^+) - p_1^2 \\
&= p_1^+(p_1 p_1^+ + p_2 p_1^- + p_3 p_1^-) - p_1^2 \\
&= p_1^+ p_1 - p_1^2 \\
&= \delta p_1(1 - p_1)
\end{aligned}$$

$$\begin{aligned}
Cov([\varepsilon_2 = 1], [\varepsilon_3 = 2]) &= (p_1 p_1^+ p_2^- + p_2 p_1^- p_2^- + p_3 p_1^- p_2^-) - p_1 p_2 \\
&= p_2^-(p_1 p_1^+ + p_2 p_1^- + p_3 p_1^-) - p_1 p_2 \\
&= p_1 p_2^- - p_1 p_2 \\
&= -\delta p_1 p_2
\end{aligned}$$

We may obtain the other entries of the matrix in a similar fashion. So, the cross-covariance matrix for a ε_2 and ε_3 with $K = 3$ categories is given by

$$\Lambda^{2,3} = \delta \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & -p_1p_3 \\ -p_1p_2 & p_2(1-p_2) & -p_2p_3 \\ -p_1p_3 & -p_2p_3 & p_3(1-p_3) \end{bmatrix}$$

Note that if ε_2 and ε_3 are independent, then the cross-covariance matrix is all zeros, because $\delta = 0$.

3. Dependency Generators

Both FK-dependency and sequential dependency structures are particular examples of a class of *vertical dependency structures*. We denote the dependency of subsequent categorical random variables on a previous variable in the sequence as a *vertical dependency*. In this section, we define a class of functions that generate vertical dependency structures with the property that all the variables in the sequence are identically distributed but dependent.

3.1 Vertical Dependency Generators Produce Identically Distributed Sequences

Define the set of functions $\mathcal{C}_\delta = \{\alpha : \mathbb{N}_{\geq 2} \to \mathbb{N} : \alpha(n) \geq 1 \wedge \alpha(n) < n \; \forall \, n\}$. The mapping defines a direct dependency of n on $\alpha(n)$, denoted $n \overset{\delta}{\rightsquigarrow} \alpha(n)$. We have already defined the notion of direct dependency, but we formalize it here.

Definition 1 (Direct Dependency)**.** *Let ε_m, ε_n be two categorical random variables in a sequence, where $m < n$. We say ε_n has a **direct dependency** on ε_m, denoted $\varepsilon_n \overset{\delta}{\rightsquigarrow} \varepsilon_m$, if $P(\varepsilon_n | \varepsilon_{n-1}, ..., \varepsilon_1) = P(\varepsilon_n | \varepsilon_m)$.*

Example 3. *For FK-dependence, $\alpha(n) \equiv 1$. That is, for any n, $\varepsilon_n \overset{\delta}{\rightsquigarrow} \varepsilon_1$*

Example 4. *The function $\alpha(n) = n - 1$ generates the sequential dependency structure of Section 2.*

We now define the notion of dependency continuity.

Definition 2 (Dependency Continuity)**.** *A function $\alpha : \mathbb{N}_{\geq 2} \to \mathbb{N}$ is dependency continuous if $\forall \, n \exists j \in \{1, ..., n-1\}$ such that $\alpha(n) = j$*

We require that the functions in $\mathcal{C}_\delta$ be dependency continuous. Thus, the formal definition of the class of dependency generators $\mathcal{C}_\delta$ is

Definition 3 (Dependency Generator)**.** *We define the set of functions $\mathcal{C}_\delta = \{\alpha : \mathbb{N}_{\geq 2} \to \mathbb{N}\}$ such that*

- $\alpha(n) < n$
- $\forall \, n \exists \, j \in \{1, ..., n-1\} : \alpha(n) = j$

*and refer to this class as **dependency generators**.*

Each function in $\mathcal{C}_\delta$ generates a unique dependency structure for sequences of dependent categorical random variables where the individual elements of the sequence remain identically distributed.

Theorem 2. *Let $\alpha \in \mathcal{C}_\delta$. Then for any $n \in \mathbb{N}, n \geq 2$, the dependent categorical random sequence generated by α has identically distributed elements.*

Proof. Let α be a dependency continuous function in $\mathcal{C}_\delta$. Then $\alpha(2) = 1 \Rightarrow \varepsilon_2 \overset{\delta}{\rightsquigarrow} \varepsilon_1$. Thus, ε_2 and ε_1 are a sequentially dependent sequence of length 2 and identically distributed. Now, $\alpha(3) \in \{1,2\}$, and thus either $\varepsilon_3 \overset{\delta}{\rightsquigarrow} \varepsilon_1$, in which case ε_3 and ε_1 are identically distributed, and thus $\varepsilon_1, \varepsilon_2$, and ε_3 are identically distributed, or $\varepsilon_3 \overset{\delta}{\rightsquigarrow} \varepsilon_2$. But since $\varepsilon_2 \overset{\delta}{\rightsquigarrow} \varepsilon_1$, it follows that $\varepsilon_3 \overset{\delta}{\rightsquigarrow} \varepsilon_2 \overset{\delta}{\rightsquigarrow} \varepsilon_1$, which is a sequentially dependent sequence of length 3, and thus all elements are identically distributed. Now, for and $n > 3$, $\alpha(n) = j \in \{1, ..., n-1\}$. If $j = 1$ or 2, then we are done. Suppose then that $j > 2$. Then by dependency continuity, $\exists l \in \{1, ..., j-1\}$ such that $\alpha(j) = l$. If $l = 1$ or 2, then ε_n is in a sequentially dependent subsequence of length 4 or 5, respectively, and is thus identically distributed within the subsequence. If $l \geq 3$, then $\exists m \in \{1, ..., l-1\}$ such that $\alpha(l) = m$. We may proceed in this fashion until we reach a q such that $\alpha(q) = 1$ or 2, which is forced by dependency continuity. Thus, for any n, ε_n is appended to a sequentially dependent subsequence that is identically distributed among its members and with ε_1. Thus all subsequences are identically distributed with ε_1 and hence each other as well. Thus, the entire sequence is identically distributed with $P(\varepsilon_n = i) = p_i, i = 1, ..., K \; \forall \, n$. □

3.2 Graphical Interpretation and Illustrations

We may visualize the dependency structures generated by $\alpha \in \mathcal{C}_\delta$ via directed *dependency graphs*. Each ε_n represents a vertex in the graph, and a directed edge connects n to $\alpha(n)$ to represent the direct dependency generated by α. This section illustrates some examples and gives a graphical interpretation of the result in Theorem 2.

3.2.1 First-Kind Dependence

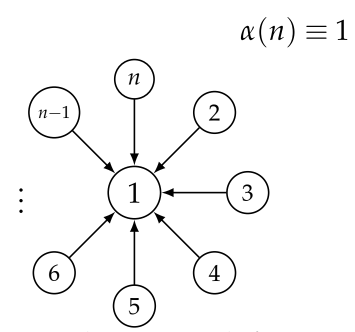

Figure 4. Dependency Graph for FK Dependence

For FK-dependency, $\alpha(n) \equiv 1$ generates the graph in Figure 4. Each ε_i depends directly on ε_1, and thus we see no connections between any other vertices i, j where $j \neq 1$. There are $n-1$ separate subsequences of length 2 in this graph.

3.2.2 Sequential Dependency

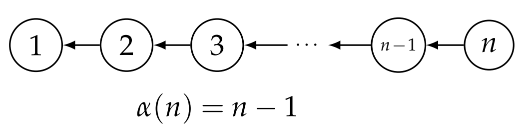

Figure 5. Dependency Graph for Sequential Dependence

$\alpha(n) = n-1$ generates the sequential dependency structure we studied in Section 2. We can see that a path exists from any n back to 1. This is a visual way to see the result of Theorem 2, in that if a path exists from any n back to 1, then the variables in that path must be identically distributed. Here, there is only one sequence and no subsequences.

3.2.3 A Monotonic Example

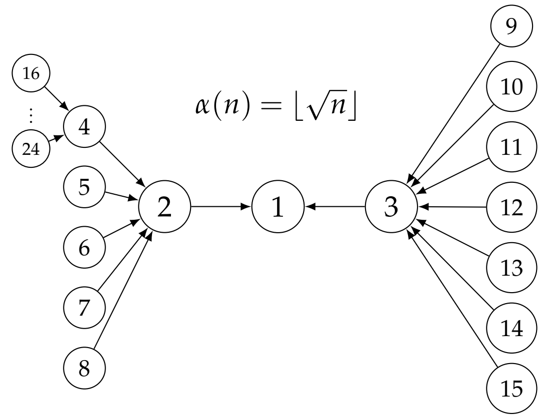

Figure 6. Dependency Graph for $\alpha(n) = \lfloor\sqrt{n}\rfloor$

Figure 6 gives an example of a monotonic function in $\mathcal{C}_\delta$ and the dependency structure it generates. Again, note that any n has a path to 1, and the number of subsequences is between 1 and $n-1$.

3.2.4 A Nonmonotonic Example

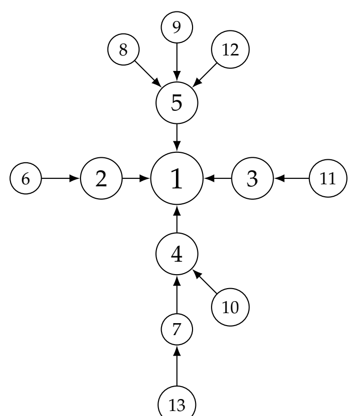

Figure 7. Dependency Graph for $\alpha(n) = \left\lfloor \frac{\sqrt{n}}{2}(\sin(n)) + \frac{n}{2} \right\rfloor$

Figure 7 illustrates a more contrived example where the function is nonmonotonic. It is neither increasing nor decreasing. The previous examples have all been nondecreasing.

3.2.5 A Prime Example

Let $\alpha \in \mathcal{C}_\delta$ be defined in the following way. Let p_m be the mth prime number, and let $\{kp_m\}$ be the set of all postive integer multiples of p_m. Then the set $\mathscr{P}_m = \{kp_m\} \setminus \cup_{i=1}^{m-1}\{kp_i$ gives a disjoint partitioning of $\mathbb{N}$. That is, $\mathbb{N} = \sqcup_m \mathscr{P}_m$, and thus for any $n \in \mathbb{N}$, $n \in \mathscr{P}_m$ for exactly one m. Now let $\alpha(n) = m$. Thus, the function is well-defined. We may write $\alpha : \mathbb{N}_{\geq 2} \to \mathbb{N}$ as $\alpha[\mathscr{P}_m] = m$. As an illustration,

$$\begin{aligned}
\alpha[\{2k\}] &= 1 \\
\alpha[\{3k\} \setminus \{2k\}] &= 2 \\
\alpha[\{5k\} \setminus (\{2k\} \cup \{3k\})] &= 3 \\
&\vdots \\
\alpha[\{kp_m\} \setminus (\cup_{i=1}^{m-1}\{kp_i\})] &= m
\end{aligned}$$

4. Conclusion

This paper has extended the concept of dependent random sequences first put forth in the works of Korzeniowski [1] and Traylor [3] and developed a generalized class of vertical dependency structures. The sequential dependency structure was studied extensively, and a formula for the cross-covariance obtained. The class of dependency generators was defined in Section 3.1 and shown to always produce a unique dependency structure for any $\alpha \in \mathscr{C}_\delta$ in which the a sequence of categorical random variables under that α is identically distributed but dependent. We provided a graphical interpretation of this class, and illustrated with several key examples.

5. Appendix

5.1 Proof of Theorem 1

The proof can be divided into two statements:

$$\begin{cases} \text{Statement 1}: & Cov([\epsilon_m = i],[\epsilon_n = i]) = p_i(1-p_i)\delta^{n-m} \\ \text{Statement 2}: & Cov([\epsilon_m = i],[\epsilon_n = j]) = -p_i p_j \delta^{n-m}; \quad m > 1, n > m \end{cases}$$

We prove Statement 1 first for $m = 1$. It suffices to show that $P(\epsilon_1 = i, \epsilon_n = i) = p_i(p_i + \delta^{n-1}(1-p_i))$. This will be shown via induction, though the base case is 6. $n = 2,..,5$ are calculated directly. Fix the number of categories K, and denote the sample space of a sequentially dependent categorical sequence of length n with K categories by Ω_n. Fix $i \in \{1,...,K\}$ Denote the following sets:

$$\Theta_n^{(i)} = \{\omega \in \Omega_n : \omega_1 = \omega_n = i\}$$
$$\Phi_n^{(i)} = \{\omega \in \Omega_n : \omega_n = i\}$$

For $n = 2$, $P(\epsilon_1 = i, \epsilon_2 = i) = p_i p_i^+ = p_i(p_i + \delta(1-p_i))$, $i = 1,...,K$. For $n = 3$, $\Theta_3^i = \{i\} \times \Phi_2^{(i)}$. Then

$$P(\epsilon_1 = i, \epsilon_3 = i) = P(\Theta_3^i) = p_i(p_i^+ p_i^+) + p_i\left(\sum_{j\neq i} p_j^-\right) p_i^- = p_i(p_i + (1-p_i)\delta^2)$$

For $n = 4$,

$$\begin{aligned} \Theta_4^{(i)} &= \{i\} \times \Phi_3^{(i)} \\ &= \{i\} \times \left(\bigcup_{j=1}^{K}\{j\}\right) \times \Phi_2^{(i)} \\ &= \left[\{i\} \times \Theta_3^{(i)}\right] \cup \left[\bigcup_{j\neq i}\{i\} \times \{j\} \times \Phi_2^{(i)}\right] \end{aligned}$$

Let $\epsilon_{(-k)}$ be a sequence with the first k terms missing.

$$\begin{aligned} P(\Theta_4^{(i)}) &= P(\epsilon_{(-1)} \in \Theta_3^{(i)} | \epsilon_1 = i)P(\epsilon_1 = i) + \sum_{j\neq i} P(\epsilon_{(-2)} \in \Phi_2^{(i)} | \epsilon_1 = i, \epsilon_2 = j)P(\epsilon_2 = j | \epsilon_1 = i)P(\epsilon_1 = i) \\ &= p_i p_i^+ (p_i + \delta^2(1-p_i) + \sum_{j\neq i}) p_i p_j^- p_i^- (1+\delta) \\ &= p_i p_i^+ (p_i + \delta^2(1-p_i) + p_i(1-p_i)(1-\delta-\delta^2+\delta^3) \\ &= p_i(p_i + (1-p_i)\delta^3) \end{aligned}$$

The calculations for $n=5$ proceed in a similar fashion. $P(\Theta_5^{(i)}) = p_i(p_i+(1-p_i)\delta^4)$. Then for $n=6$, which is our true base case for the inductive hypothesis,

$$\Theta_6^{(i)} = \{i\}\times\Theta_5 \cup \left(\bigcup_{j\neq i}\{i\}\times\{j\}\times\Phi_4^{(i)}\right)$$

We already know that $P(\varepsilon_{(-1)}\in\Theta_5|\varepsilon_1=i) = p_i p_i^+(p_i+(1-p_i)\delta^4)$. Now, fix j. Then we may partition $\{i\}\times\{j\}\times\Phi_4^{(i)}$.

$$\{i\}\times\{j\}\times\Phi_4^{(i)} = \left[\{i\}\times\{j\}\times\Theta_4^{(i)}\right]\cup\left[\{i\}\times\{j\}\times\{j\}\times\Phi_3^{i}\right]\cup\left(\bigcup_{k\neq i,j}\left[\{i\}\times\{j\}\times\{k\}\times\Phi_3^{(i)}\right]\right) \tag{5}$$

Then we calculate the probability of each partition separately.

$$P\left(\{i\}\times\{j\}\times\Theta_4^{(i)}\right) = p_i p_j^- p_i^-(p_i+(1-p_i)\delta^3) \tag{6}$$

$$\begin{aligned} P\left(\{i\}\times\{j\}\times\{j\}\times\Phi_3^{i}\right) &= p_i p_j^- p_j^+ P(\varepsilon_{(-3)}\in\Phi_3^{(i)}|\varepsilon_1=i,\varepsilon_2=\varepsilon_3=j) \\ &= p_i(p_i p_j(1-\delta)(1-\delta^3)(p_j+(1-p_j)\delta)) \end{aligned} \tag{7}$$

and for fixed $k\neq i,j$,

$$\begin{aligned} P\left(\{i\}\times\{j\}\times\{k\}\times\Phi_3^{(i)}\right) &= p_i p_j^- p_k^- P(\varepsilon_{(-3)}\in\Phi_3)|\varepsilon_1=i,\varepsilon_2=j,\varepsilon_3=k) \\ &= p_i\left(\prod_{l=1}^{K} p_l\right)(1-\delta)^2(1-\delta^3) \end{aligned} \tag{8}$$

Then $P\left(\bigcup_{k\neq i,j}\left[\{i\}\times\{j\}\times\{k\}\times\Phi_3^{(i)}\right]\right)$ is obtained by summing the probabilities in (6) - (8) over all $j\neq i$. That is,

$$\begin{aligned} P\left(\bigcup_{k\neq i,j}\left[\{i\}\times\{j\}\times\{k\}\times\Phi_3^{(i)}\right]\right) &= \sum_{\substack{j=1\\ j\neq i}}^{K}\Big(p_i p_j^- p_i^-(p_i+(1-p_i)\delta^3) + p_i(p_i p_j(1-\delta)(1-\delta^3)(p_j+(1-p_j)\delta)) \\ &\quad + \sum_{l\neq i,j} p_i\left(\prod_{l=1}^{K}p_l\right)(1-\delta)^2(1-\delta^3)\Big) \\ &= p_i(1-p_i)(1-\delta-\delta^4+\delta^5) \end{aligned}$$

Therefore,

$$\begin{aligned} P\left(\Theta_6^{(i)}\right) &= p_i p_i^+(p_i+(1-p_i)\delta^4) + p_i(1-p_i)(1-\delta-\delta^4+\delta^5) \\ &= p_i(p_i+(1-p_i)\delta^5) \end{aligned}$$

For the inductive hypothesis, assume for $6\leq m\leq n$,

$$P(\varepsilon_{(-1)}\in\Theta_{n-1}|\varepsilon_1=i) = p_i^+(p_i+(1-p_i)\delta^{n-2}) \tag{9}$$

$$P(\varepsilon_{(-1)} \in \Theta_{n-1} | \varepsilon_1 \neq i) = p_i^-(p_i + (1-p_i)\delta^{n-2}) \tag{10}$$

$$P(\varepsilon_{(-3)} \in \Phi_{n-3}^{(i)} | \varepsilon_1 = i, \varepsilon_2 = \varepsilon_3 = j) = p_j(1-\delta^{n-3}) \tag{11}$$

and for $k \neq i, j$,

$$P(\varepsilon_{(-3)} \in \Phi_{n-3}^{(i)} | \varepsilon_1 = i, \varepsilon_2 = j \varepsilon_3 = k) = (1-\delta^{n-3}) \prod_{l \neq j,k} p_l \tag{12}$$

Then $P(\varepsilon_1 = i, \varepsilon_{n+1} = i) = P(\Theta_{n+1}^{(i)})$. Now,

$$\begin{aligned}\Theta_{n+1}^{(i)} &= \{i\} \times \Phi_n^{(i)} \\ &= (\{i\} \times \Theta_n) \cup \left(\bigcup_{j \neq i} \{i\} \times \{j\} \times \Phi_{n-1} \right)\end{aligned}$$

Then we can break $\{i\} \times \Theta_n$ down further and express it as

$$\{i\} \times \Theta_n = (\{i\} \times \{i\} \times \Theta_{n-1}) \cup \left(\bigcup_{k \neq i} \{i\} \times \{k\} \times \Phi_{n-1} \right) \tag{13}$$

We can then see that $P(\{i\} \times \{i\} \times \Theta_{n-1}) = p_i p_i^+(p_i + (1-p_i)\delta^{n-2})$ by the inductive hypothesis. Then, we break down (for fixed k) $\{i\} \times \{k\} \times \Phi_{n-1}$ into

$$\{i\} \times \{k\} \times \Phi_{n-1} = (\{i\} \times \{k\} \times \Theta_{n-1}) \cup \left(\bigcup_{l \neq i} \{i\} \times \{k\} \times \{l\} \times \Phi_{n-2} \right) \tag{14}$$

By the inductive hypothesis, we may calculate (in a similar manner as for the base case $n = 6$) that

$$P(\bigcup_{k \neq i} \{i\} \times \{k\} \times \Phi_{n-1}) = p_i(1-p_i)(1-\delta-\delta^{n-2}+\delta^{n-1})$$

and therefore

$$P(\{i\} \times \Theta_n) = p_i p_i^+(p_i + (1-p_i)\delta^{n-2}) + p_i(1-p_i)(1-\delta-\delta^{n-2}+\delta^{n-1}) = p_i(p_i + (1-p_i)\delta^{n-1})$$

In a similar manner, we can break down $\{i\} \times \{j\} \times \Phi_{n-1}$ for fixed j, and, using the inductive hypotheses (and some very tedious arithmetic), calculate that

$$P\left(\bigcup_{k \neq i} \{i\} \times \{k\} \times \Phi_{n-1} \right) = p_i(1-p_i)(1-\delta-\delta^{n-1}+\delta^n)$$

Combining yields $P(\Theta_{n+1}^{(i)}) = p_i(p_i + (1-p_i)\delta^n)$ and the statement is proven.

To prove the Statement 1 for $m > 1$, we abuse notation and redefine the set $\Theta_{n-m}^{(i)} = \{\omega \in \Omega_n : \omega_m = i, \omega_n = i\}$. Then it also suffices to show that $P(\Theta_{n-m}^{(i)}) = p_i(1-p_i)\delta^{n-m}$.

$$\Theta_{n-m}^{(i)} = \Omega_m \times \Theta_{n-m}$$

where Ω_m consists of sequences (or subsequences) of length m and ending at index $m-1$. Then

$$\Omega_m \times \Theta_{n-m}^{(i)} = \bigcup_{j=1}^{K} \Phi_m^{(j)} \times \Theta_{n-m}^{(i)} \tag{15}$$

Now, since the sequential variables are identically distributed, $P(\Phi_m^j) = P(\varepsilon_{m-1} = j) = p_j$. Thus,

$$\begin{aligned} P(\Theta_{n-m}^{(i)}) &= p_i(p_i^+(p_i + (1-p_i)\delta^{n-m})) + \sum_{j \neq i} p_j p_i^-(p_i + (1-p_i)\delta^{n-m})) \\ &= p_i(p_i + (1-p_i)\delta^{n-m}) \end{aligned}$$

The proof of Statement 2 follows in a similar fashion.

Acknowledgments

The authors are grateful to the DPD CTO of Dell EMC and Stephen Manley for continued funding and support.

References

[1] Andrzej Korzeniowski. On correlated random graphs. *Journal of Probability and Statistical Science*, pages 43–58, 2013.

[2] Joseph McKean Robert Hogg and Allen Craig. *Introduction to Mathematical Statistics*. Prentice Hall, 6 edition.

[3] Rachel Traylor. A generalized multinomial distribution from dependent categorical random variables. *Academic Advances of the CTO*, 2017.