ISSN 1066-5307, Mathematical Methods of Statistics, 2022, Vol. 31, No. 4, pp. 165–169. © Allerton Press, Inc., 2022.

What Intraclass Covariance Structures Can Symmetric Bernoulli Random Variables Have?

Iosif Pinelis*

Michigan Technological University, Houghton, Michigan, USA

Received June 16, 2022; revised September 24, 2022; accepted October 3, 2022

Abstract—The covariance matrix of random variables $X_1, \ldots, X_n$ is said to have an intraclass covariance structure if the variances of all the X_i's are the same and all the pairwise covariances of the X_i's are the same. We provide a possibly surprising characterization of such covariance matrices in the case when the X_i's are symmetric Bernoulli random variables.

DOI: 10.3103/S1066530722040020

Keywords: intraclass covariance structure, covariance matrix, symmetric Bernoulli random variables

For natural $n \geq 2$, let $\Sigma = [\Sigma_{i,j}]_{i,j\in[n]}$ be the covariance matrix of random variables (r.v.'s) $X_1, \ldots, X_n$ with finite second moments, so that $\Sigma_{i,j} = \mathsf{Cov}(X_i, X_j)$ for all i and j in the set $[n] := \{1, \ldots, n\}$. We are assuming that the matrix Σ is nonzero.

The covariance matrix Σ is said to have an intraclass covariance structure if (i) $\Sigma_{i,i} = \mathsf{Var}\, X_i = \mathsf{Cov}(X_i, X_i)$ is the same for all $i \in [n]$ and (ii) $\Sigma_{i,j} = \mathsf{Cov}(X_i, X_j)$ is the same for all distinct i and j in $[n]$. Let ICCS_n denote the set of all $n \times n$ covariance matrices that have an intraclass covariance structure.

In particular, if the r.v.'s $X_1, \ldots, X_n$ are exchangeabl—that is, if the joint distribution of the X_i's is invariant with respect to all permutations of the indices $1, \ldots, n$ (see e.g., [4] for much more on exchangeability of r.v.'s), then the covariance matrix Σ will be in the set ICCS_n. So, one may say that the covariance matrix Σ has an intraclass covariance structure if the r.v.'s $X_1, \ldots, X_n$ pertain to items that belong to one class and thus are exchangeable in a certain weak sense; this explains the use of the term "intraclass". The notion of an intraclass covariance structure was introduced by Fisher [3] and has been studied in many subsequent papers, including e.g., [7, 9, 10].

Obviously, the covariance matrix Σ is in the set ICCS_n if and only if

$$\Sigma = (a - b)I_n + b\,\mathbb{1}_n\mathbb{1}_n^\top \tag{1}$$

for some real numbers a and b, where I_n is the $n \times n$ identity matrix and $\mathbb{1}_n := [1, \ldots, 1]^\top$, the $n \times 1$ matrix of 1's.

Recall that a real $n \times n$ matrix is a covariance matrix if and only if it is positive semidefinite; cf. e.g., [2, Sect. III.6, Theorem 4]. Note that (i) $\mathbb{1}_n$ is an eigenvector of the matrix $\mathbb{1}_n\mathbb{1}_n^\top$ belonging to the eigenvalue n and (ii) any nonzero vector orthogonal to $\mathbb{1}_n$ is an eigenvector of the matrix $\mathbb{1}_n\mathbb{1}_n^\top$ belonging to the eigenvalue 0. So, the only eigenvalues of the matrix Σ of the form (1) are $a - b + bn$ and $a - b$.

It follows that the matrix Σ of the form (1) is in ICCS_n if and only if $-\frac{a}{n-1} \leq b \leq a$, that is, if and only if the pairwise correlation, $\rho = b/a$, between r.v.'s whose covariance matrix has an intraclass covariance structure is no less that $-1/(n-1)$:

$$\rho \geq \rho_{n,\min} := -\frac{1}{n-1}. \tag{2}$$

This is in contrast with the general lower bound -1 on the correlation between arbitrary r.v.'s. Let us refer to the values of ρ satisfying condition (2) as *good*.

*E-mail: ipinelis@mtu.edu

In the rest of this note, we shall consider the special case when the r.v.'s $X_1, \dots, X_n$ are symmetric Bernoulli, so that

$$\mathsf{P}(X_i = 1) = \tfrac{1}{2} = \mathsf{P}(X_i = 0) \tag{3}$$

for all $i \in [n]$. This important case has been extensively studied in computer science in general and in machine learning in particular (see e.g., [1, 5, 8, 11]), as well as in other applications of probability theory—though mainly when the X_i's are independent.

The question now is the following:

> For what values of pairwise correlation ρ do there exist symmetric Bernoulli r.v.'s $X_1, \dots, X_n$ whose covariance matrix Σ is in ICCS_n?

Let us refer to such values of ρ as *symmetric-binary-good*. Clearly, any symmetric-binary-good value of ρ must be good. One then may wonder whether every good value of ρ is symmetric-binary-good.

The answer to this question may seem surprising:

- if n is even, then yes, every good value of ρ is symmetric-binary-good;
- if n is odd, then "nearly every" good value of ρ is symmetric-binary-good.

For symmetric Bernoulli r.v.'s $X_1, \dots, X_n$ whose covariance matrix Σ is in ICCS_n, it is a bit more convenient to deal with the probability

$$p := \mathsf{P}(X_1 = X_2)$$

than with the correlation ρ. It is easy to see that the values of ρ and p are in the simple bijective correspondence

$$(-1, 1) \ni 2p - 1 = \rho \longleftrightarrow p = \frac{1+\rho}{2} \in (0, 1), \tag{4}$$

so that $\mathsf{P}(X_i = X_j) = p$ for all distinct i and j in $[n]$.

Let us refer to the values of p corresponding to the good values of ρ as *good* values of p, and let us similarly define the *symmetric-binary-good* values of p. So, in view of (2) and (4), a value $p \in (0, 1)$ is good if and only if

$$p \geq p_n := \frac{n-2}{2(n-1)}. \tag{5}$$

Thus, we have to determine the symmetric-binary-good values of p.

Suppose for a moment that $p \in (0, 1)$ is symmetric-binary-good. Then there exist symmetric Bernoulli r.v.'s $X_1, \dots, X_n$ such that $\mathsf{P}(X_i = X_j) = p$ for all distinct i and j in $[n]$. Letting g stand for the joint probability mass function of the r.v.'s $X_1, \dots, X_n$, we note that g is a nonnegative function such that

(i) $\sum_{x \in \{0,1\}^n} g(x) = 1$,

(ii) $\sum_{x \in \{0,1\}^n} 1(x_i = 0) g(x) = \frac{1}{2}$ for all $i \in [n]$,

(iii) $\sum_{x \in \{0,1\}^n} 1(x_i = x_j) g(x) = p$ for all distinct i and j in $[n]$;

of course, here x_i denotes the ith coordinate of the vector $x = (x_1, \dots, x_n) \in \{0, 1\}^n$. By symmetry, conditions (i)–(iii) will hold with $\tilde{g}(x) := \frac{1}{n!} \sum_{\pi \in \Pi_n} g(\pi(x))$ in place of $g(x)$, where Π_n is the set of all permutations of the set $[n]$. Note that $\tilde{g}(x) = f(\sum_1^n x_i)$ for some nonnegative function $f\colon \{0, \dots, n\} \to \mathbb{R}$ and all $x \in \{0, 1\}^n$. So, conditions (i)–(iii) can be rewritten as

(I) $\sum_{k=0}^n \binom{n}{k} f(k) = 1$,

(II) $\sum_{k=0}^n \binom{n-1}{k} f(k) = \frac{1}{2}$ for all i,

(III) $\sum_{k=0}^n a_{n,k} f(k) = p$,

where

$$a_{n,k} = \binom{n-2}{k} + \binom{n-2}{k-2};$$

of course, $\binom{n-1}{n} = 0$, $\binom{n-2}{k} = 0$ if $k \geq n-1$ and $\binom{n-2}{k-2} = 0$ if $k \leq 1$.

Thus, for any given $n \geq 2$ and $p \in (0,1)$, we want to see whether there is a nonnegative function $f\colon \{0, \dots, n\} \to \mathbb{R}$ such that conditions (I)–(III) hold.

Towards this goal, consider the problem of finding the extrema of $\sum_{k=0}^{n} a_{n,k} f(k)$ over all $f \in F_n$, where F_n is the set of all nonnegative function $f\colon \{0, \dots, n\} \to \mathbb{R}$ satisfying condition (I). In view of the symmetries $\binom{n}{k} = \binom{n}{n-k}$ and $a_{n,k} = a_{n,n-k}$, without loss of generality the functions f are symmetric in the same sense: $f(k) = f(n-k)$ for all $k \in \{0, \dots, n\}$—otherwise, replacing $f(k)$ by $\frac{1}{2}\,(f(k) + f(n-k))$, we will have the sums in (I) and (III) unchanged. Next, consider the ratios

$$r_k := r_{n,k} := \frac{a_{n,k}}{\binom{n}{k}} = \frac{(n-k)(n-k-1) + k(k-1)}{n(n-1)}.$$

Note that $r_{k+1} \leq r_k$ if $0 \leq k \leq \frac{n-1}{2}$ and $r_{k+1} \geq r_k$ if $\frac{n-1}{2} \leq k \leq n-1$. Also, $r_k = r_{n-k}$. So, the smallest among the r_k's is/are the one/ones with index/indices k closest to $\frac{n}{2}$.

More specifically, if $n = 2m-1$ is odd, then $r_k \geq r_m = r_{m-1}$ for all $k \in \{1, \dots, n-1\}$. Letting then

$$f_{\min}^{\text{odd}}(m-1) := \frac{1/2}{\binom{n}{m-1}} = \frac{1/2}{\binom{n}{m}}, \quad f_{\min}^{\text{odd}}(m) := \frac{1/2}{\binom{n}{m}} = \frac{1/2}{\binom{n}{m-1}},$$
$$f_{\min}^{\text{odd}}(k) := 0 \quad \text{for all} \quad k \in \{0, \dots, n\} \setminus \{m-1, m\},$$

we see that $f_{\min}^{\text{odd}}$ is a symmetric function in F_n and

$$(r_k - r_m)(f_{\min}^{\text{odd}}(k) - f(k)) \leq 0$$

for all $k \in \{0, \dots, n\}$ and all symmetric functions $f \in F_n$, which implies

$$\begin{aligned}
\sum_{k=0}^{n} a_{n,k} f_{\min}^{\text{odd}}(k) - \sum_{k=0}^{n} a_{n,k} f(k) &= \sum_{k=0}^{n} a_{n,k} (f_{\min}^{\text{odd}}(k) - f(k)) \\
&= \sum_{k=0}^{n} \binom{n}{k} r_k (f_{\min}^{\text{odd}}(k) - f(k)) \\
&= \sum_{k=0}^{n} \binom{n}{k} (r_k - r_m)(f_{\min}^{\text{odd}}(k) - f(k)) \leq 0.
\end{aligned}$$

It follows that $f_{\min}^{\text{odd}}$ is a minimizer of $\sum_{k=0}^{n} a_{n,k} f(k)$ over all $f \in F_n$, that is, over all nonnegative f satisfying condition (I). Moreover, condition (II) is satisfied with $f_{\min}^{\text{odd}}$ in place of f.

We conclude that, in the case when $n = 2m-1$ is odd, $f_{\min}^{\text{odd}}$ is a minimizer of $\sum_{k=0}^{n} a_{n,k} f(k)$ over all nonnegative f satisfying *both* conditions (I) and (II). The corresponding minimum value of $\sum_{k=0}^{n} a_{n,k} f(k)$ is

$$p_{n,\min}^{\text{odd}} := \sum_{k=0}^{n} a_{n,k} f_{\min}^{\text{odd}}(k) = \frac{m-1}{2m-1} = \frac{n-1}{2n}.$$

Similarly, in the case when $n = 2m$ is even, a minimizer of $\sum_{k=0}^{n} a_{n,k} f(k)$ over all nonnegative f satisfying both conditions (I) and (II) is given by

$$f_{\min}^{\text{even}}(m) := \frac{1}{\binom{n}{m}} \quad \text{and} \quad f_{\min}^{\text{even}}(k) := 0 \ \text{ for all } \ k \in \{0, \dots, n\} \setminus \{m\},$$

and the corresponding minimum value of $\sum_{k=0}^n a_{n,k} f(k)$ is

$$p_{n,\min}^{\text{even}} := \sum_{k=0}^n a_{n,k} f_{\min}^{\text{even}}(k) = \frac{m-1}{2m-1} = \frac{n-2}{2(n-1)}.$$

The above minimization can of course be recognized as something similar to, or even a special case of, the Neyman–Pearson lemma [6, part III].

The just considered cases of odd and even n can be summarized as follows. For

$$m_n := \lceil n/2 \rceil,$$

let $f_{\min}$ be the symmetric function in F_n such that $\sum_{k\in\{m_n, n-m_n\}} f_{\min}(k) = 1$, so that $f(k) = 0$ for $k \in \{0,\dots,n\} \setminus \{m_n, n-m_n\}$. Then $f_{\min}$ is a minimizer of $\sum_{k=0}^n a_{n,k} f(k)$ over all nonnegative f satisfying conditions (I) and (II). The corresponding minimum value of $\sum_{k=0}^n a_{n,k} f(k)$ is

$$p_{n,\min} := \sum_{k=0}^n a_{n,k} f_{\min}(k) = \frac{m_n - 1}{2m_n - 1}.$$

The extremal joint distribution of the binary r.v.'s $X_1,\dots,X_n$ corresponding to the minimizer $f_{\min}$ can be described as follows: the random set $I := \{i \in [n] \colon X_i = 1\}$ is uniformly distributed on the set $S_n := \binom{[n]}{m_n} \cup \binom{[n]}{n-m_n}$, where $\binom{[n]}{k}$ denotes the set of all subsets of cardinality k of the set $[n]$; of course, $S_n := \binom{[n]}{n/2}$ if n is even.

Next, letting

$$f_{\max}(0) := \tfrac{1}{2}, \quad f_{\max}(n) := \tfrac{1}{2}, \quad f_{\max}(k) := 0 \text{ for all } k \in \{1,\dots,n-1\},$$

we see that the nonnegative function $f_{\max}$ satisfies conditions (I) and (II), and also $\sum_{k=0}^n a_{n,k} f_{\max}(k) = 1$. On the other hand, for any nonnegative function f satisfying conditions (I) and (II), the sum $\sum_{k=0}^n a_{n,k} f(k)$ is a probability and hence does not exceed 1. We conclude that $f_{\max}$ is a maximizer of $\sum_{k=0}^n a_{n,k} f(k)$ over all nonnegative f satisfying conditions (I) and (II). The corresponding maximum value of $\sum_{k=0}^n a_{n,k} f(k)$ is

$$p_{n,\max} := \sum_{k=0}^n a_{n,k} f_{\max}(k) = 1.$$

The extremal joint distribution of the binary r.v.'s $X_1,\dots,X_n$ corresponding to the maximizer $f_{\max}$ can be described as follows: the random set $I = \{i \in [n] \colon X_i = 1\}$ is uniformly distributed on the set $\{\emptyset, [n]\}$; that is, $\mathsf{P}(I = \emptyset) = \frac{1}{2} = \mathsf{P}(I = [n])$.

Now note that the set of all values of $\sum_{k=0}^n a_{n,k} f(k)$, where $f\colon \{0,\dots,n\} \to \mathbb{R}$ is a nonnegative function such that conditions (I) and (II) hold, is convex and therefore coincides with the interval $[p_{n,\min}, p_{n,\max}] = [p_{n,\min}, 1]$.

Thus, a value $p \in (0,1)$ is symmetric-binary-good if and only if

$$p \ge p_{n,\min} = \frac{m_n - 1}{2m_n - 1} = \begin{cases} \dfrac{n-2}{2(n-1)} = p_n & \text{if } n \text{ is even} \\ \dfrac{n-1}{2n} = p_{n+1} > p_n & \text{if } n \text{ is odd,} \end{cases}$$

where p_n is as in (5).

Because p_{n+1} is close to p_n for large n and in view of the correspondence (4) between ρ and p, we have now confirmed that

- if n is even then every good value of ρ is symmetric-binary-good;
- if n is odd then, for large n, nearly every good value of ρ is symmetric-binary-good.

One may also note here that for large n the lower bound $\rho_{n,\min}$ (defined in (2)) is close to (but less than) 0, whereas the lower bound $p_{n,\min}$ is close to (but less than) $\frac{1}{2}$.

REFERENCES

1. P. Baldi and R. Vershynin, "A theory of capacity and sparse neural encoding," Neural Networks **143**, 12–27 (2021).
2. W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. II, 2nd ed. (John Wiley and Sons, Inc., New York-London-Sydney, 1971).
3. F. A. Fisher. *Statistical Methods for Research Workers. Pure and Applied Mathematics* (New York, Oliver and Boyd, 1932).
4. O. Kallenberg, *Probabilistic Symmetries and Invariance Principles. Probability and its Applications* (New York, Springer, New York, 2005).
5. N. Natarajan, I. S. Dhillon, P. K. Ravikumar, and A. Tewari, Learning with noisy labels, Ed. by C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Advances in Neural Information Processing Systems, Vol. 26 (Curran Associates, Inc., 2013).
6. J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," Mathematics, Mathematical Proceedings of the Cambridge Philosophical Society **231**, 289–337 (1933).
7. S. J. Press, "Structured multivariate Behrens–Fisher problems," Sankhya: The Indian Journal of Statistics, Series A (1961–2002) **29** (1), 41–48 (1967).
8. K. Senel and E. G. Larsson, Joint user activity and non-coherent data detection in mmtcenabled massive mimo using machine learning algorithms, in WSA 2018; 22nd International ITG Workshop on Smart Antennas (2018), pp. 1–6.
9. M. S. Srivastava and M. Singull, "Testing sphericity and intraclass covariance structures under a growth curve model in high dimension," Communications in Statistics—Simulation and Computation **46** (7), 5740–5751 (2017).
10. J. E. Walsh, "Concerning the Effect of Intraclass Correlation on Certain Significance Tests," The Annals of Mathematical Statistics **18** (1), 88–96 (1947).
11. F. Zhang, W. Wang, J. Hou, J. Wang, and J. Huang, "Tensor restricted isometry property analysis for a large class of random measurement ensembles," Sci. China Inf. Sci. **64** (1), 119101 (2021).