A groupwise approach to the birthday paradox

Kaiji Motegi* and Soma Hayashi†
Kobe University

October 6, 2023

Abstract

A key quantity of the birthday paradox is the probability that all $K$ individuals have distinct birthdays across $N$ calendar days. The permutation approach computes this probability directly, using permutations of all individuals. The pairwise approach sequentially computes the conditional probability for each pair of individuals to have distinct birthdays given that all previous pairs do. We propose a groupwise approach in which all possible groups of size $J$ are handled sequentially. The groupwise approach contains the permutation and pairwise approaches as special cases. The conditional groupwise probability of no birthday collisions takes various shapes such as decreasing, increasing, and hump-shaped curves. We characterize these patterns with $(N, K, J)$, and derive a number of new insights on the birthday paradox.

MSC 2010 Classification Codes: 60C05, 65C50, 97K50.

Keywords: collision, conditional probability, exponentiation approximation, pairwise approach, permutation approach.

* Corresponding author. Graduate School of Economics, Kobe University. 2-1 Rokkodai-cho, Nada, Kobe, Hyogo 657-8501 Japan. E-mail: motegi@econ.kobe-u.ac.jp
†Faculty of Economics, Kobe University. 2-1 Rokkodai-cho, Nada, Kobe, Hyogo 657-8501 Japan. E-mail: 2362186e@stu.kobe-u.ac.jp
1 Introduction

In a class of randomly selected individuals, the probability that at least one pair shares the same birthday is higher than it sounds. Such a counter-intuitive phenomenon is called the birthday paradox, and its literature dates back to Mosteller (1962) at least. The birthday paradox itself is a curious problem, and is related to various fields including Markov chains (Kim, Montenegro, Peres, and Tetali, 2010) and cryptology (Bellare and Kohno, 2004, Suzuki, Tonien, Kurosawa, and Toyota, 2006). Indeed, an attempt to find a collision of a hash function, which is a major concern of cryptology, is called a birthday attack.

A basic set-up of the birthday paradox is as follows. Let \( \mathcal{K} = \{1, \ldots, K\} \) be the set of individuals, and let \( \mathcal{N} = \{1, \ldots, N\} \) be the set of calendar days, where \( 2 \leq K \leq N \). Let \( A_{k,n} \) be an event that the birthday of individual \( k \in \mathcal{K} \) is day \( n \in \mathcal{N} \). Assume hereafter that uniform probability is assigned to each profile of the \( K \) individuals’ birthdays:

\[
\Pr \left( \bigcap_{k=1}^{K} A_{k,n_k} \right) = \frac{1}{N^K}, \quad \forall\{n_1, \ldots, n_K\} \in \mathcal{N}^K. \tag{1}
\]

The probability of no birthday collisions is defined as follows.

\[
Q(N, K) = \Pr(\text{The birthdays of the } K \text{ individuals are all distinct}). \tag{2}
\]

A core statement of the birthday paradox is that \( Q(N, K) \) takes a smaller value than one would imagine. The number of permutations of choosing \( K \) distinct birthdays out of \( N \) days is given by \( \mathcal{N}P_K = \prod_{i=0}^{K-1} (N - i) = N!/(N - K)! \). Under Assumption (1), \( Q(N, K) \) can be characterized by

\[
Q^*(N, K) = \frac{\mathcal{N}P_K}{N^K}. \tag{3}
\]

Equation (3) is called the permutation approach, since it incorporates all possible permutations such that the \( K \) individuals have distinct birthdays.

---

All $K$ individuals have distinct birthdays if and only if all pairs of individuals do. Based on this insight, Motegi and Woo (2023) (“MW2023” hereafter) took the pairwise approach to characterize $Q(N, K)$ as the product of the conditional probabilities that each pair have distinct birthdays given that all previous pairs do. A conventional simplification called the exponentiation approximation imposes an incorrect assumption of pairwise independence, leading to an approximated value of $Q(N, K)$ (Mosteller, 1962, Schwarz, 1988, Blom and Holst, 1989). MW2023 derived an exact formula for the conditional non-collision probability, and evaluated the precision of the exponentiation approximation. Their main findings are two-fold. First, the conditional non-collision probability decreases in a step-function form as more pairs are restricted to have distinct birthdays. Second, the exponentiation approximation slightly overestimates the true probability $Q(N, K)$.

This paper proposes a groupwise approach by sequentially handling all groups of an arbitrary size $J \in \{2, 3, \ldots, K\}$. The groupwise approach reduces to the pairwise approach when $J = 2$, and to the permutation approach when $J = K$. We characterize the conditional non-collision probability with $(N, K, J)$, and derive the following results. First, the conditional non-collision probability is monotonically decreasing in the number of conditioned groups if and only if $J = 2$. Second, focusing on $J \geq 3$, the conditional probability is monotonically increasing if and only if $K = J + 1$. Third, focusing on $J \geq 3$, the conditional probability is hump-shaped if and only if $K > J + 1$. These findings imply that MW2023’s first result (i.e., decreasing conditional probability) holds only under the pairwise approach. Non-collision restrictions among members of a group have a positive impact on the conditional non-collision probability, while those between members and non-members have a negative impact. The relative magnitude of these effects determines the shape of the conditional probability.

We also show that the exponentiation approximation results in extremely large negative bias for $J \in \{3, \ldots, K - 1\}$. It is in a stark contrast to MW2023’s second result (i.e., small positive bias resulting from the approximation). This paper is the first work that unifies the pairwise and permutation approaches, and we obtain a number of new insights on conditional probabilities of sequential events. These are a valuable contribution to the
literature of birthday paradox and more generally probability, statistics, and informatics.

The rest of this paper is organized as follows. In Section 2, the groupwise approach is formulated. In Section 3, our main results are stated and demonstrated. In Section 4, numerical examples which illustrate the main results are presented. In Section 5, brief concluding remarks are provided. Omitted technical details are collected in Appendices. We use the following notation throughout the paper. \(1(A)\) is the indicator function which equals 1 if event \(A\) occurs and 0 otherwise. \(|A|\) signifies the number of elements of set \(A\).

2 The groupwise approach

In this section, we propose the groupwise approach formally. We construct a canonical order of groups in Section 2.1, and formulate the conditional probability of no birthday collisions in Section 2.2.

2.1 Canonical order of groups

The family of all possible groups of size \(J\) selected from \(K\) individuals is written as \(\hat{B} = \{\{k_1, \ldots, k_J\} \subseteq \hat{K} \mid k_1 < \cdots < k_J\}\). The number of all possible groups is given by \(|\hat{B}| = K^{C_J} = \frac{K \cdot P_J}{J!}\). Let \(B_i = \{k_{i1}, \ldots, k_{iJ}\}\) be the \(i^{th}\) element of \(\hat{B}\) with \(i \in \{1, \ldots, K\}^{C_J}\). Let \(G(K, J)\) be a \(K^{C_J} \times J\) group membership matrix which results from aligning the \(K^{C_J}\) groups in a canonical order. The \((i, j)\)-element of \(G(K, J)\) signifies the \(j^{th}\) member of group \(B_i\) (i.e., \(k_{ij}\)), where \(i \in \{1, \ldots, K\}^{C_J}\) and \(j \in \{1, \ldots, J\}\).

The canonical order of groups is specified as follows. The first group is set to be \(B_1 = \{1, \ldots, J - 1, J\}\). While keeping the first \(J - 1\) members, we replace the last member with individuals \(J + 1, J + 2, \ldots, K\) sequentially; these steps yield \(\{B_2, \ldots, B_{K-J+1}\}\). The next group is constructed by replacing the second last member of \(B_1\) with individual \(J\): \(B_{K-J+2} = \{1, \ldots, J - 2, J, J + 1\}\). Proceed analogously until we reach the last group \(B_{K^{C_J}} = \{K - J + 1, \ldots, K\}\).

Generic formulation of the canonical order of groups is tedious, but practical computation
is straightforward as the canonical order is built in many statistical software packages. The group membership matrix \( G(K, J) \) can be computed by typing `combn(K, J)` in R with the `combinat` package, or typing `nchoosek(1:K, J)` in Matlab.\(^2\) For illustration, we present an example with \( K = 5 \) and \( J = 3 \):\(^3\)

\[
G(5, 3) = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 2 & 2 & 3 \\
2 & 2 & 2 & 3 & 3 & 3 & 4 & 4 \\
3 & 4 & 5 & 5 & 5 & 4 & 5 & 5 \\
\end{bmatrix}^\top.
\]

(4)

When \( J = 2 \), our group construction coincides with MW2023’s pair construction for any \( K \in \{2, \ldots, N\} \).

2.2 Conditional probability of no birthday collisions

Let \( D_B \) be an event that there are no birthday collisions within group \( B \in \mathcal{B} \). Note that

\[
Q(N, K) = \Pr \left( \bigcap_{i=1}^{KCJ} D_{B_i} \right) = \Pr(D_{B_1}) \prod_{i=2}^{KCJ} \Pr \left( D_{B_i} \left| \bigcap_{j=1}^{i-1} D_{B_j} \right. \right),
\]

(6)

where \( Q(N, K) \) is defined in (2); (5) holds since all \( K \) individuals have distinct birthdays if and only if there are no collisions for all \( KCJ \) groups; (6) follows by the definition of conditional probability; the groups are aligned in the canonical order. Define:

\[
Q_i(N, K, J) = \Pr \left( D_{B_i} \left| \bigcap_{j=1}^{i-1} D_{B_j} \right. \right), \quad i \in \{1, \ldots, KCJ\},
\]

(7)

where it is understood that \( Q_1(N, K, J) = \Pr(D_{B_1}) = NP_J / N^J \). In words, \( Q_i(N, K, J) \) is the probability for the \( i^{\text{th}} \) group to have distinct birthdays conditional on that all previous

\(^2\)For Python, \( G(K, J) \) can be computed by using the `arange` function from the `numpy` library and the `combinations` function from the `itertools` library.
groups do. Substitute (7) into (6) to get

\[ Q(N, K) = \prod_{i=1}^{KCJ} Q_i(N, K, J) \equiv Q^{**}(N, K). \] (8)

We call (8) the \textit{groupwise approach}, since it exploits all possible groups of size \(J\). Recall from (3) that \(Q^*(N, K)\) is an exact solution for \(Q(N, K)\) when one takes the permutation approach. Similarly, \(Q^{**}(N, K)\) in (8) is an exact solution for \(Q(N, K)\) when one takes the groupwise approach. Our primary goal is to characterize \(Q_i(N, K, J)\) with \((N, K, J)\).

We are also interested in a consequence of approximating (6) as follows.

\[ Q(N, K) \approx \Pr(D_{B_1}) \prod_{i=2}^{KCJ} \Pr(D_{B_i}) \]

\[ = \{\Pr(D_{B_i})\}^{KCJ} \]

\[ = \left(\frac{NP_J}{NJ}\right)^{KCJ} \equiv Q^{ap}(N, K, J), \] (11)

where (9) is an approximation which replaces the conditional probability \(\Pr(D_{B_i} \mid \cap_{j=1}^{i-1}D_{B_j})\) in (6) with the unconditional probability \(\Pr(D_{B_i})\); (10) follows by Assumption (1). \(Q^{ap}(N, K, J)\) defined in (11) is called the \textit{exponentiation approximation} of \(Q(N, K)\), since the unconditional groupwise probability of non-collisions is raised by the number of groups. Note that the outcome of the exponentiation approximation, \(Q^{ap}(N, K, J)\), depends on group size \(J \in \{2, \ldots, K\}\), while the true probability \(Q(N, K)\) does not.

A common mistake of students or even teachers is to believe that (9) should hold exactly, as they are sometimes confused between conditional and unconditional probabilities. If this misunderstanding occurs, one reaches an incorrect conclusion that \(Q(N, K) = Q^{ap}(N, K, J)\). Equation (9) would hold if there were independence between an event that all previous groups have distinct birthdays and an event that a present group has distinct birthdays. The independence does not hold actually, as the information of the previous groups having no collisions affects the probability of the present group having no collisions. In what follows, we evaluate the sign and magnitude of the bias resulting from the exponentiation approximation.
3 Main results

In this section, we present our main results characterizing the groupwise approach. Before stating our theorems, it is of use to summarize the basic assumptions.

**Assumption 1.** (i) The calendar size is \( N \in \{2,3,\ldots\} \). (ii) The class size is \( K \in \{2,3,\ldots, N\} \). (iii) The group size is \( J \in \{2,3,\ldots, K\} \). (iv) The set of calendar days is \( \mathcal{N} = \{1,\ldots, N\} \), and the set of individuals is \( \mathcal{K} = \{1,\ldots, K\} \). (v) \( \Pr(\cap_{k=1}^{K} A_{k,n}) = 1/N^K \) for all \( \{n_1,\ldots,n_K\} \in \mathcal{N}^K \), where \( A_{k,n} \) is an event that the birthday of individual \( k \in \mathcal{K} \) is day \( n \in \mathcal{N} \). (vi) The \( K C_J \) groups are aligned in the canonical order.

Assumption 1 is maintained throughout the paper. In Section 3.1, we present a main theorem which characterizes the conditional groupwise probability of no birthday collisions. In Section 3.2, we characterize the shapes of the conditional non-collision probability. In Section 3.3, we analyze the consequence of the exponentiation approximation.

3.1 Characterizing the conditional non-collision probability

The following is a main theorem which characterizes the conditional groupwise probability of no birthday collisions, \( \{Q_i(N,K,J)\}_{i=1}^{K C_J} \), with arbitrary \((N,K,J)\).

**Theorem 1.** Under Assumption 1, we have that

\[
Q_1(N,K,J) = \frac{N P_J}{N^J},
\]

\[
Q_i(N,K,J) = \frac{N - J + 1}{N}, \quad \forall i \in \{2,\ldots, K - J + 1\},
\]

\[
Q_i(N,K,J) = \frac{N - J + 1 - \sum_{\ell=1}^{K-J} \ell \times 1 \{i \in \mathcal{M}_\ell(K,J)\}}{N - J + 2 - \sum_{\ell=1}^{K-J} \ell \times 1 \{i \in \mathcal{M}_\ell(K,J)\}},
\]

\[
\forall i \in \left\{ K - J + 2,\ldots, \frac{1}{2}(K - J + 1)(K - J + 2) \right\},
\]

\[
Q_i(N,K,J) = 1, \quad \forall i \in \left\{ \frac{1}{2}(K - J + 1)(K - J + 2) + 1,\ldots, K C_J \right\},
\]
where \( Q_i(N, K, J) \) is defined in (7) and

\[
\mathcal{M}_1(K, J) = \{K - J + 2, \ldots, 2(K - J) + 1\},
\]

\[
\mathcal{M}_\ell(K, J) = \left\{ K - J + 2 + \sum_{h=1}^{\ell-1} (K - J - h + 1), \ldots, 2(K - J) + 1 + \sum_{h=1}^{\ell-1} (K - J - h) \right\},
\]

\[\ell \in \{2, \ldots, K - J\}.\]

\textit{Proof.} Proof of Theorem 1 is presented in Appendix A.1.

Theorem 1 delivers a number of useful implications on the groupwise approach. First, the groupwise approach with any group size \( J \in \{2, \ldots, K\} \) leads to \( Q^{**}(N, K) = \frac{N}{P_K/N^K} \), which coincides with the solution of the permutation approach appearing in (3).

\textbf{Corollary 1.} Under Assumption 1, we have that \( Q^*(N, K) = Q^{**}(N, K) = \frac{N}{P_K/N^K} \), where \( Q^*(N, K) \) is characterized in (3) and \( Q^{**}(N, K) \) is defined in (8).

\textit{Proof.} Proof of Corollary 1 is presented in Appendix A.2.

\textbf{Remark 1.} Corollary 1 indicates that the groupwise approach with any group size \( J \in \{2, \ldots, K\} \) yields the same solution as the permutation approach. In particular, when \( J = K \), there is only one group (i.e., the whole class) and \( Q^{**}(N, K) = Q_1(N, K, K) = \frac{N}{P_K/N^K} \) by (12). This scenario is literally identical to the permutation approach. Thus, the groupwise approach contains the permutation approach as a special case with \( J = K \).

When \( J = 2 \), Theorem 1 is simplified as follows.

\textbf{Corollary 2.} Under Assumption 1, we have that

\[
Q_i(N, K, 2) = \frac{N - 1 - \sum_{\ell=1}^{K-2} \ell \times 1 \{i \in \mathcal{M}_\ell(K, 2)\}}{N - \sum_{\ell=1}^{K-2} \ell \times 1 \{i \in \mathcal{M}_\ell(K, 2)\}} \quad \forall i \in \{1, \ldots, K C_2\},
\]

(17)

where \( \mathcal{M}_\ell(K, 2) \) is defined in accordance with (16):

\[
\mathcal{M}_1(K, 2) = \{K, \ldots, 2K - 3\},
\]
\[ M_\ell(K, 2) = \left\{ K + \sum_{h=1}^{\ell-1} (K - 1 - h), \ldots, 2K - 3 + \sum_{h=1}^{\ell-1} (K - 2 - h) \right\}, \quad \ell \in \{2, \ldots, K - 2\}. \]

**Proof.** Equations (12)-(13) imply that \( Q_i(N, K, 2) = (N - 1)/N \) for all \( i \in \{1, \ldots, K - 1\} \).

Equation (14) implies that
\[
Q_i(N, K, 2) = \frac{N - 1 - \sum_{\ell=1}^{K-2} \ell \times 1 \{i \in M_\ell(K, 2)\}}{N - \sum_{\ell=1}^{K-2} \ell \times 1 \{i \in M_\ell(K, 2)\}}, \quad \forall i \in \{K, \ldots, K C_2\}.
\]

Hence, (17) is true.  \( \square \)

**Remark 2.** Corollary 2 coincides with MW2023’s Theorem 1, which means that the groupwise approach contains the pairwise approach as a special case with \( J = 2 \). Thus, we have shown that the groupwise approach unifies the permutation and pairwise approaches.

We call group \( i \in \{1, \ldots, K C_J\} \) a relevant group if the conditional probability of distinct birthdays is less than 1 (i.e., \( Q_i(N, K, J) < 1 \)). Similarly, we call group \( i \) an irrelevant group if \( Q_i(N, K, J) = 1 \). As seen from (17), all \( K C_J \) pairs are relevant when \( J = 2 \). For \( J \geq 3 \), irrelevant groups start to appear as more groups are restricted to have no collisions.

**Corollary 3.** Under Assumption 1, the following are true. (i) Groups \( \{1, \ldots, i^*(K - J)\} \) are relevant groups, where \( i^*(A) = A + 2 C_2 \) with \( A \) being a non-negative integer. (ii) Groups \( \{i^*(K - J) + 1, \ldots, K C_J\} \) are irrelevant groups.

**Proof.** Equations (12)-(14) imply that \( Q_i(N, K, J) < 1 \) for all \( i \in \{1, \ldots, i^*(K - J)\} \). Equation (15) implies that \( Q_i(N, K, J) = 1 \) for all \( i \in \{i^*(K - J) + 1, \ldots, K C_J\} \).  \( \square \)

An intuition behind Corollary 3 is as follows. Due to the canonical ordering of groups, \( B_{i^*(K - J)} = \{1, \ldots, J - 2, K - 1, K\} \) and it is the first group to which both individuals \( K - 1 \) and \( K \) belong. Any pair of individuals jointly belongs to at least one of \( \{B_1, \ldots, B_{i^*(K - J)}\} \).

Hence, each of \( \{B_{i^*(K - J) + 1}, \ldots, B_{K C_J}\} \) must have distinct birthdays if there are no collisions in all of \( \{B_1, \ldots, B_{i^*(K - J)}\} \). One can better understand this logic by referring to example (4), where \( i^*(K - J) = 6 \) and \( B_6 = \{1, 4, 5\} \). Finally, it is worth noting that the number of relevant groups \( i^*(K - J) \) is determined by \( K - J \) but not by \( K \) and \( J \) separately.
3.2 Shapes of the conditional non-collision probability

In this section, we investigate the shapes of the conditional groupwise probability of no birthday collisions, \( \{Q_i(N, K, J)\}_{i=1}^{K} \). A striking fact is that \( \{Q_i(N, K, J)\} \) is monotonically decreasing if and only if the pairwise approach is taken.

**Theorem 2.** Under Assumption 1, \( \{Q_i(N, K, J)\}_{i=1}^{K} \) is monotonically decreasing in group index \( i \) if and only if \( J = 2 \).

**Proof.** Proof of Theorem 2 is presented in Appendix A.3. \( \square \)

To avoid potential confusion, let us clarify what we mean by “monotonicity”. First, we cannot say \( \{Q_i(N, K, 2)\} \) is strictly monotonically decreasing, as there are constant parts; for example, \( Q_i(N, K, 2) = (N - 1)/N \) for all \( i \in \{1, \ldots, K - 1\} \). Second, we rule out the possibility that \( \{Q_i(N, K, 2)\} \) is entirely constant; observe \( Q_1(N, K, 2) = (N - K + 1)/(N - K + 2) < Q_1(N, K, 2) \). The same clarification applies for the rest of the paper.

**Remark 3.** MW2023 showed in their Theorem 1 that \( \{Q_i(N, K, 2)\}_{i=1}^{K} \) is monotonically decreasing. Theorem 2 above includes it, and additionally shows that \( \{Q_i(N, K, J)\} \) is monotonically decreasing only if \( J = 2 \). It is a notable refinement of MW2023’s theorem.

To explain an intuitive reason why \( \{Q_i(N, K, 2)\} \) is monotonically decreasing, consider an example with \( N = K = 4 \). In this case, pairs are aligned as \( \mathcal{B}_1 = \{1, 2\}, \mathcal{B}_2 = \{1, 3\}, \mathcal{B}_3 = \{1, 4\}, \mathcal{B}_4 = \{2, 3\}, \mathcal{B}_5 = \{2, 4\}, \mathcal{B}_6 = \{3, 4\} \). The conditional pairwise probability of no birthday collisions is indeed decreasing:

\[
Q_i(4, 4, 2) = \begin{cases} 
0.75 & \text{if } i \in \{1, 2, 3\}, \\
0.667 & \text{if } i \in \{4, 5\}, \\
0.5 & \text{if } i = 6.
\end{cases}
\]

The reason why \( Q_3(4, 4, 2) > Q_4(4, 4, 2) \) is that, conditional on \( \cap_{j=1}^{3} D_{\mathcal{B}_j} \), the members of group 4 (i.e., individuals 2-3) must have different birthdays from individual 1, who is not a member of group 4. Similarly, \( Q_5(4, 4, 2) > Q_6(4, 4, 2) \) since the members of group 6 (i.e.,...
individuals 3-4) must have different birthdays from the non-members (i.e., individuals 1-2) conditional on $$\bigcap_{j=1}^{5} D_B$$. Generally, if members of a group are required to have different birthdays from non-members, then a collision among the members is more likely to occur.

For $$J \geq 3$$, $$\{Q_i(N, K, J)\}$$ can take various shapes depending on the values of $$(N, K, J)$$. A key condition which determines the shape of $$\{Q_i(N, K, J)\}$$ is whether $$K = J + 1$$ or $$K > J + 1$$. When $$K = J + 1$$, $$\{Q_i(N, K, J)\}$$ is monotonically increasing.

**Theorem 3.** Impose Assumption 1. Assume further that $$J \geq 3$$. Then, $$\{Q_i(N, K, J)\}_{i=1}^{K \cap C}$$ is monotonically increasing in group index $$i$$ if and only if $$K = J + 1$$.

**Proof.** Proof of Theorem 3 is presented in Appendix A.4.

An intuitive reason for the increasing conditional probability of no collisions is that each group is close to the whole class since $$K = J + 1$$. Consider an example with $$(N, K, J) = (5, 5, 4)$$, then $$B_1 = \{1, 2, 3, 4\}$$, $$B_2 = \{1, 2, 3, 5\}$$, $$B_3 = \{1, 2, 4, 5\}$$, $$B_4 = \{1, 3, 4, 5\}$$, and $$B_5 = \{2, 3, 4, 5\}$$. The conditional non-collision probability is increasing as expected:

$$Q_i(5, 5, 4) = \begin{cases} 
0.192 & \text{if } i = 1, \\
0.4 & \text{if } i = 2, \\
0.5 & \text{if } i = 3, \\
1 & \text{if } i \in \{4, 5\}.
\end{cases}$$

The reason why $$Q_1(5, 5, 4) < Q_2(5, 5, 4)$$ is that individuals 1-3 are members of groups 1-2 and hence guaranteed to have distinct birthdays each other conditional on $$D_{B_1}$$. Similarly, $$Q_2(5, 5, 4) < Q_3(5, 5, 4)$$ because, conditional on $$D_{B_1} \cap D_{B_2}$$, individuals 1, 2, and 4 have distinct birthdays each other, and individual 5 has a different birthday from individuals 1-2. Generally, if more members of a group are carried over to a subsequent group, the conditional non-collision probability increases.

When $$K > J + 1$$, $$\{Q_i(N, K, J)\}_{i=1}^{i \cap (K-J)}$$ is hump-shaped.

**Theorem 4.** Impose Assumption 1. Assume further that $$J \geq 3$$, then the following are equivalent:
(i) $K > J + 1$.

(ii) $\{Q_i(N,K,J)\}_{i=1}^{i^*(K-J)}$ is monotonically increasing on the range $i \in \{1, \ldots, K - J + 2\}$, monotonically decreasing on the range $i \in \{K - J + 2, \ldots, i^*(K - J)\}$, and the maximum value is given by $Q_i(N,K,J) = (N - J)/(N - J + 1)$ for all $i \in M_1(K,J)$,

where $i^*(K-J) = K-J+2C_2$ as defined in Corollary 3; $M_1(K,J) = \{K - J + 2, \ldots, 2(K - J) + 1\}$ as defined in (16).

Proof. Proof of Theorem 4 is presented in Appendix A.5. 

The hump-shaped curve represents a balance between two opposite effects. Non-collision conditions among members have a positive impact on the conditional non-collision probability, while those between members and non-members have a negative impact. The former effect is dominant under the groupwise approach with $K = J + 1$, making $\{Q_i(N,K,J)\}$ monotonically increasing (Theorem 3). The latter effect is dominant under the pairwise approach, making $\{Q_i(N,K,2)\}$ monotonically decreasing (Theorem 2). Under the groupwise approach with $K > J + 1$, the former dominates the latter for $i \in \{1, \ldots, K - J + 2\}$ and vice versa for $i \in \{K - J + 2, \ldots, i^*(K - J)\}$.

To observe a hump-shaped curve, suppose that $(N,K,J) = (5,5,3)$. Then, relevant groups are $B_1 = \{1, 2, 3\}$, $B_2 = \{1, 2, 4\}$, $B_3 = \{1, 2, 5\}$, $B_4 = \{1, 3, 4\}$, $B_5 = \{1, 3, 5\}$, and $B_6 = \{1, 4, 5\}$ as shown in (4). The conditional non-collision probability is given by

$$Q_i(5,5,3) = \begin{cases} 
0.48 & \text{if } i = 1, \\
0.6 & \text{if } i \in \{2, 3\}, \\
0.667 & \text{if } i \in \{4, 5\}, \\
0.5 & \text{if } i = 6, \\
1 & \text{if } i \in \{7, 8, 9, 10\}.
\end{cases}$$  \hfill (18)

We observe $Q_1(5,5,3) < Q_2(5,5,3)$ since individuals 1-2 are members of groups 1-2 and hence guaranteed to have distinct birthdays each other conditional on $D_{B_1}$. By contrast,
\(Q_5(5,5,3) > Q_6(5,5,3)\) as each member of group 6 must have a different birthday from all non-members conditional on \(\cap_{i=1}^{5} D_i\); a non-collision between individuals 1 and 4 as well as a non-collision between individuals 1 and 5 are guaranteed, but individuals 4-5 should also have distinct birthdays each other while avoiding a collision with the non-members.

### 3.3 The exponentiation approximation

Theorems 2-3 have direct implications on the exponentiation approximation. By (8), the true probability of no birthday collisions is \(Q(N,K) = \prod_{i=1}^{K} Q_i(N,K,J)\). By (11) and (12), the exponentiation approximation amounts to \(Q_{ap}(N,K,J) = \{Q_1(N,K,J)\}^{K}C_J\). Hence, if \(Q_1(N,K,J)\) is the maximum (resp. minimum) value of \(\{Q_i(N,K,J)\}^{K}C_J\), then the exponentiation approximation must overestimate (resp. underestimate) the true probability.

Combining this insight and Theorems 2-3, we can establish the following corollary.

**Corollary 4.** Impose Assumption 1, then the following are true. (i) If \(J = 2\), then \(Q_{ap}(N,K,J) > Q(N,K)\) for any \(N \in \{3,4,\ldots\}\) and \(K \in \{3,4,\ldots,N\}\). (ii) If \(J \geq 3\) and \(K = J + 1\), then \(Q_{ap}(N,K,J) < Q(N,K)\) for any \(N \in \{4,5,\ldots\}\), \(K \in \{4,5,\ldots,N\}\), and \(J = K - 1\).

**Proof.** (i) Theorem 2 implies that, if \(J = 2\), then \(\{Q_i(N,K,J)\}^{K}C_J\) is monotonically decreasing and hence its maximum is \(Q_1(N,K,J)\). Hence, \(Q_{ap}(N,K,J) > Q(N,K)\).

(ii) Theorem 3 implies that, if \(J \geq 3\) and \(K = J + 1\), then \(\{Q_i(N,K,J)\}^{K}C_J\) is monotonically increasing and hence its minimum is \(Q_1(N,K,J)\). Hence, \(Q_{ap}(N,K,J) < Q(N,K)\). \(\Box\)

**Remark 4.** The bias resulting from the exponentiation approximation is positive if \(J = 2\), and it is negative if \(J \geq 3\) and \(K = J + 1\). The former result is identical to MW2023’s Corollary 3, and the latter result is new to the literature.

When \(J \geq 3\) and \(K > J + 1\), we know from Theorem 4 that \(\{Q_i(N,K,J)\}^{K-J+2}_{i=1}\) is monotonically increasing and \(\{Q_i(N,K,J)\}^{K-J}_{i=1}\) is monotonically decreasing. Hence, the minimum of \(\{Q_i(N,K,J)\}^{K}C_J\) is given by either \(Q_1(N,K,J) = N^J / N^J\) or \(Q_{i*}(K-J)(N,K,J) = (N-K+1)/(N-K+2)\), and their relative magnitude depends on \((N,K,J)\). If \(Q_1(N,K,J)\)
is the minimum like in example (18), then the exponentiation approximation surely leads to negative bias. Otherwise, it is not clear whether the bias is positive or negative. This issue will be inspected via numerical examples in the next section.

4 Numerical examples

We provide some numerical examples to illustrate the main results derived in Section 3. In Section 4.1, we compute the conditional groupwise probability of no collisions. In Section 4.2, we compare the true non-collision probability and its exponentiation approximation.

4.1 Conditional probability of no birthday collisions

Consider a simple example with calendar size $N = 6$, class size $K \in \{4, 5, 6\}$, and group size $J \in \{2, 3, \ldots, K - 1\}$. The conditional groupwise probability of no birthday collisions, $\{Q_i(N, K, J)\}_{i=1}^{KCJ}$, can be calculated from Theorem 1. By (8), the non-collision probability based on the groupwise approach is $Q^{**}(N, K) = \prod_{i=1}^{KCJ} Q_i(N, K, J)$. We have that $Q^{**}(6, 4) = 0.278$, $Q^{**}(6, 5) = 0.093$, and $Q^{**}(6, 6) = 0.015$ for all $J \in \{2, \ldots, K - 1\}$. These are identical to solutions of the permutation approach (3): $Q^*(6, 4) = 0.278$, $Q^*(6, 5) = 0.093$, and $Q^*(6, 6) = 0.015$. Hence, we have verified Corollary 1, which states that the permutation and groupwise approaches lead to the same solution: $Q^*(N, K) = Q^{**}(N, K)$.

In Figure 1, we plot $\{Q_i(N, K, J)\}_{i=1}^{KCJ}$. Panel (a) depicts $\{Q_i(6, 4, 2)\}$, and Panel (c) depicts $\{Q_i(6, 5, 2)\}$. These panels coincide with MW2023’s Figure 1, which confirms our Corollary 2; the groupwise approach with $J = 2$ is equivalent to the pairwise approach. Focusing on $J \geq 3$, Corollary 3 implies that irrelevant groups start to emerge at $i \in \{4, 7, 11\}$ when $K - J \in \{1, 2, 3\}$, respectively. Indeed, $Q_i(N, K, J) = 1$ for all $i \geq 4$ in Panels (b), (e), and (i) of Figure 1 ($K - J = 1$), for all $i \geq 7$ in Panels (d) and (h) ($K - J = 2$), and for all $i \geq 11$ in Panel (g) ($K - J = 3$). Hence, we have confirmed the validity of Corollary 3.

Figure 1 highlights a difference between the pairwise approach ($J = 2$) and the groupwise approach ($J \geq 3$); $\{Q_i(N, K, J)\}$ is monotonically decreasing in group index $i \in \{4, 7, 11\}$ when $K - J \in \{1, 2, 3\}$, respectively.
Figure 1: The conditional groupwise probability of no birthday collisions, $Q_i(N, K, J)$

This figure plots the conditional groupwise probability of no birthday collisions $\{Q_i(N, K, J)\}_{i=1}^{K-C_J}$, where the calendar size is $N = 6$, the class size is $K \in \{4, 5, 6\}$, and the group size is $J \in \{2, \ldots, K-1\}$. The groups are aligned in the canonical order as described in Section 2.1.

$\{1, \ldots, K-C_J\}$ in Panels (a), (c), and (f), while it is not in the other panels. This contrast is consistent with Theorem 2, which states that $\{Q_i(N, K, J)\}$ is monotonically decreasing if
and only if $J = 2$. Focusing on $J \geq 3$, Theorem 3 states that $\{Q_i(N, K, J)\}$ is monotonically increasing if and only if $K = J+1$. In fact, monotonically increasing curves appear in Panel (b) with $(K, J) = (4, 3)$, Panel (e) with $(K, J) = (5, 4)$, and Panel (i) with $(K, J) = (6, 5)$.

Given $J \geq 3$, Theorem 4 states that $\{Q_i(N, K, J)\}$ is hump-shaped if and only if $K > J+1$. This statement is supported by Panels (d), (g), and (h), where $(K, J) \in \{(5, 3), (6, 3), (6, 4)\}$, respectively. Theorem 4 also asserts that $\{Q_i(N, K, J)\}^{K-J+2}_{i=1}$ is monotonically increasing, $\{Q_i(N, K, J)\}^{i^*(K-J)}_{i=K-J+2}$ is monotonically decreasing, and the maximum of $\{Q_i(N, K, J)\}^{i^*(K-J)}_{i=1}$ is given by $Q_i(N, K, J) = (N - J)/(N - J + 1)$ for all $i \in \{K - J + 2, \ldots, 2(K - J) + 1\}$, where $i^*(K - J) = K - J + 2C_2$. When $(N, K, J) = (6, 5, 3)$, these conjectures mean that $\{Q_i(6, 5, 3)\}_{i=1}^4$ must be increasing, $\{Q_i(6, 5, 3)\}_{i=4}^6$ be decreasing, and the maximum be $Q_i(6, 5, 3) = 0.75$ at $i \in \{4, 5\}$. These conjectures are all supported by Panel (d). Similar argument applies for Panels (g) and (h) as well.

Another implication from Theorem 4 is that, when $J \geq 3$ and $K > J+1$, the minimum value of $\{Q_i(N, K, J)\}$ must arise at either $i = 1$ or $i = i^*(K - J)$. In Panel (d), $Q_1(6, 5, 3) = 0.556$ and $Q_{i^*(5-3)}(6, 5, 3) = 0.667$ with $i^*(5-3) = 6$, hence the minimum is given by the former. In Panel (g), $Q_1(6, 6, 3) = 0.556$ and $Q_{i^*(6-3)}(6, 6, 3) = 0.5$ with $i^*(6-3) = 10$, hence the minimum is given by the latter. In Panel (h), the minimum is $Q_1(6, 6, 4) = 0.278$.

### 4.2 True probability and the exponentiation approximation

We now compare the true non-collision probability $Q(N, K)$ and its exponentiation approximation $Q^{ap}(N, K, J)$. Suppose that the calendar size is $N \in \{5, 6, 7\}$, the class size is $K \in \{4, \ldots, N\}$, and the group size is $J \in \{3, \ldots, K - 1\}$. For each $(N, K, J)$, we compute $Q(N, K)$ from (3) and $Q^{ap}(N, K, J)$ from (11). In Table 1, we report $Q(N, K)$, $Q^{ap}(N, K, J)$, and their natural logs. We begin by reviewing some basic properties. First, $Q(N, K)$ increases as $N$ increases; for example, $Q(5, 4) = 0.192$, $Q(6, 4) = 0.278$, and $Q(7, 4) = 0.350$. It is a reasonable result since the large calendar size should make a non-collision more likely. Second, $Q(N, K)$ decreases as $K$ increases; compare the previous values with $Q(5, 5) = 0.038$, $Q(6, 5) = 0.093$, and $Q(7, 5) = 0.150$. It is also reasonable since the larger class size should
make a non-collision less likely. The punchline of the birthday paradox is that $Q(N, K)$ is decreasing in $K$ more sharply than people imagine (i.e., a birthday collision often occurs even in a seemingly small class).

Table 1: True probability of no birthday collisions and exponentiation approximation

<table>
<thead>
<tr>
<th>$N$</th>
<th>$K$</th>
<th>$J$</th>
<th>$Q(N, K)$</th>
<th>$Q^{ap}(N, K, J)$</th>
<th>$\ln Q(N, K)$</th>
<th>$\ln Q^{ap}(N, K, J)$</th>
<th>$1{Q_1 &lt; Q_{i^*(K-J)}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5</td>
<td>4</td>
<td>0.19200</td>
<td>0.05308</td>
<td>-1.650</td>
<td>-2.936</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3</td>
<td>0.03840</td>
<td>0.00065</td>
<td>-3.260</td>
<td>-7.340</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>3</td>
<td>0.27778</td>
<td>0.09526</td>
<td>-1.281</td>
<td>-2.351</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>3</td>
<td>0.09259</td>
<td>0.00280</td>
<td>-2.380</td>
<td>-5.878</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>3</td>
<td>0.01543</td>
<td>0.00001</td>
<td>-4.171</td>
<td>-11.756</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>3</td>
<td>0.01543</td>
<td>0.00000</td>
<td>-4.171</td>
<td>-14.277</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>3</td>
<td>0.34985</td>
<td>0.14051</td>
<td>-1.050</td>
<td>-1.962</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>3</td>
<td>0.14994</td>
<td>0.00740</td>
<td>-1.898</td>
<td>-4.906</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>3</td>
<td>0.04284</td>
<td>0.00005</td>
<td>-3.150</td>
<td>-9.812</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>3</td>
<td>0.04284</td>
<td>0.00000</td>
<td>-3.150</td>
<td>-15.754</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>3</td>
<td>0.00612</td>
<td>0.00000</td>
<td>-5.096</td>
<td>-17.172</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>3</td>
<td>0.00612</td>
<td>0.00000</td>
<td>-5.096</td>
<td>-36.758</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>3</td>
<td>0.00612</td>
<td>0.00000</td>
<td>-5.096</td>
<td>-39.848</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>3</td>
<td>0.00612</td>
<td>0.00000</td>
<td>-5.096</td>
<td>-22.052</td>
<td>1</td>
</tr>
</tbody>
</table>

Calendar size: $N \in \{5, 6, 7\}$. Class size: $K \in \{4, \ldots, N\}$. Group size: $J \in \{3, \ldots, K-1\}$. The true non-collision probability is $Q(N, K) = N!P_K/N^K$, and its exponentiation approximation is $Q^{ap}(N, K, J) = (NP_J/N^J)^{KC_J}$. This table reports $Q(N, K)$, $Q^{ap}(N, K, J)$, and their natural logs. We also report $1\{Q_1(N, K, J) < Q_{i^*(K-J)}(N, K, J)\}$ to see which of $Q_1$ and $Q_{i^*(K-J)}$ is the minimum value of $Q_i(N, K, J)^{KC_J}$, where $i^*(K - J) = K - J + 2C_2$.

For all cases considered in Table 1, the exponentiation approximation heavily underestimates the true non-collision probability. Observe $Q(5, 4) = 0.1920$ versus $Q^{ap}(5, 4, 3) = 0.0531$; $Q(5, 5) = 0.0384$ versus $Q^{ap}(5, 5, 3) = 0.0007$ and $Q^{ap}(5, 5, 4) = 0.0003$. In some cases
like $N = K = 7$, comparing $Q(N, K)$ and $Q^{ap}(N, K, J)$ is visually hard since both of them are close to 0. Comparison becomes easier after taking natural logs: $\ln Q(7, 7) = -5.096$ versus $\ln Q^{ap}(7, 7, 3) = -17.17$, $\ln Q^{ap}(7, 7, 4) = -36.76$, $\ln Q^{ap}(7, 7, 5) = -39.85$, and $\ln Q^{ap}(7, 7, 6) = -22.05$. Due to its substantial negative bias, the exponentiation approximation loses a practical value when $J \in \{3, \ldots, K-1\}$.

Corollary 4.(ii) asserts that, when $J \geq 3$ and $K = J + 1$, $Q_1(N, K, J)$ is the minimum value of $\{Q_i(N, K, J)\}_{i=1}^{K}$. and hence $Q(N, K) > Q^{ap}(N, K, J)$. The condition that $K = J + 1$ holds for 9 out of the 19 cases considered in Table 1, and the negative bias arises for all of them as expected. In 8 out of the remaining 10 cases, $K > J + 1$ and $Q_1(N, K, J)$ is the minimum of $\{Q_i(N, K, J)\}$. Not surprisingly, the negative bias arises for all these cases. The last two cases are outside the scope of Corollary 4: $(N, K, J) \in \{(6, 6, 3), (7, 7, 3)\}$, where $Q_{\gamma(K-J)}(N, K, J)$ is the minimum of $\{Q_i(N, K, J)\}$. Even for these cases, the exponentiation approximation produces large negative bias. The large negative bias under the groupwise approach with $J \in \{3, \ldots, K-1\}$ is in a stark contrast to small positive bias under the pairwise approach, where the latter was revealed analytically and numerically by MW2023. In summary, the exponentiation approximation achieves a reasonable accuracy only when $J = 2$, and one should never use it when $J \geq 3$.

5 Conclusion

In the existing literature of the birthday paradox, the non-collision probability is computed via the permutation or pairwise approach. The permutation approach is a direct approach which uses permutations of all $K$ individuals. The pairwise approach sequentially computes the conditional probability for each pair of individuals to have distinct birthdays given that all previous pairs do. In this paper, we have proposed the groupwise approach by sequentially handling all groups of an arbitrary size $J \in \{2, 3, \ldots, K\}$. The groupwise approach reduces to the pairwise approach when $J = 2$, and to the permutation approach when $J = K$.

We have characterized the conditional groupwise probability of non-collisions with calendar size $N$, class size $K$, and group size $J$. The groupwise approach yields many useful
implications. First, the conditional non-collision probability is monotonically decreasing in the number of conditioned groups if and only if $J = 2$. Second, focusing on $J \geq 3$, the conditional probability is monotonically increasing if and only if $K = J + 1$. Third, focusing on $J \geq 3$, the conditional probability is hump-shaped if and only if $K > J + 1$. Non-collision restrictions among members of a group have a positive impact on the conditional non-collision probability, while those between members and non-members have a negative impact. The shape of the conditional non-collision probability is determined by a balance between these two effects. We have also shown that the exponentiation approximation, which imposes an incorrect assumption that all groups are independent, results in extremely large negative bias for $J \in \{3, \ldots, K - 1\}$.

This paper is the first work that unifies the pairwise and permutation approaches. Motegi and Woo (2023) showed that, if $J = 2$, then the conditional non-collision probability is monotonically decreasing, and the exponentiation approximation produces small positive bias. We have replicated their results as a special case with $J = 2$, and have established substantial extensions. The new results with $J \in \{3, \ldots, K - 1\}$ help us better understand the conditional probability of sequential events, hence they are a valuable contribution to probability, statistics, and informatics.

Acknowledgements

We thank John W. Dennis, Yuki Sugimoto, Yuto Tada, Sejun Woo, and seminar participants at Kobe University for helpful comments and discussions.

References


Appendices

In this appendix, we present technical details omitted from the main body of the paper. Impose Assumption 1 throughout the appendix.

A.1 Proof of Theorem 1

Let \( R_i(N, K, J) \) be the number of all birthday profiles of \( K \) individuals such that there are no collisions in the first through \( i \)th groups:

\[
R_i(N, K, J) = \left\{ \{n_1, \ldots, n_K\} \in \mathbb{N}^K \left| \bigcap_{j=1}^{i} D_{B_j} \text{ occurs} \right. \right\}, \quad i \in \{1, \ldots, K_C J\}. \tag{A.1}
\]

Recall that \( B_1 = \{1, \ldots, J\} \) by Assumption 1.(vi). Under \( D_{B_1} \), individual 1 can pick \( N \) days; individual 2 can pick \( N-1 \) days; individual \( J \) can pick \( N-J+1 \) days; each of the remaining \( K-J \) individuals can pick \( N \) days. Hence, we have that

\[
R_1(N, K, J) = N P_J N^{K-J}. \tag{A.2}
\]

Equation (A.2) implies that \( Q_1(N, K, J) = R_1(N, K, J)/N^K = N P_J/N^J \), confirming (12).

For \( i \geq 2 \), we have that

\[
Q_i(N, K, J) = \frac{\Pr \left( \bigcap_{j=1}^{i} D_{B_j} \right)}{\Pr \left( \bigcap_{j=1}^{i-1} D_{B_j} \right)} \tag{A.3}
\]

\[
= \frac{R_i(N, K, J)}{N^K} \bigg/ \frac{R_{i-1}(N, K, J)}{N^K} \tag{A.4}
\]

\[
= \frac{R_i(N, K, J)}{R_{i-1}(N, K, J)}, \quad i \in \{2, \ldots, K_C J\}, \tag{A.5}
\]

where (A.3) follows by (7) and the definition of conditional probability; (A.4) follows by (A.1). Equation (A.5) implies that \( \{Q_i(N, K, J)\}_{i=2}^{K_C J} \) is characterized once \( \{R_i(N, K, J)\}_{i=1}^{K_C J} \) is characterized.

To characterize \( R_2(N, K, J) \), recall that \( B_2 = \{1, \ldots, J-1, J+1\} \) by Assumption 1.(vi). Under \( D_{B_1} \cap D_{B_2} \), individuals \( \{1, \ldots, J-1\} \) can pick \( \{N, \ldots, N-J+2\} \) days, respectively; individual \( J \) can pick \( N-J+1 \) days as she belongs to \( B_1 \) but not to \( B_2 \); individual \( J+1 \) can pick \( N-J+1 \) days as she belongs to \( B_2 \) but not to \( B_1 \); each of individuals \( \{J+2, \ldots, K\} \) can pick \( N \) days as they belong to neither \( B_1 \) nor \( B_2 \). Hence, we have that \( R_2(N, K, J) = \)
\[ N P_J(N - J + 1)^{N - J - 1}. \] Using (A.2) and (A.5), we have that

\[ Q_2(N, K, J) = \frac{R_2(N, K, J)}{R_1(N, K, J)} = \frac{N P_J(N - J + 1)^{N - J - 1}}{N P_J N^{K - J}} = \frac{N - J + 1}{N}. \tag{A.6} \]

Similar argument applies up to \( B_{K-J+1} = \{1, \ldots, J - 1, K\} \):

\[ R_i(N, K, J) = N P_J(N - J + 1)^{i-1} N^{K - J + 1 - i}, \quad i \in \{2, \ldots, K - J + 1\}. \tag{A.7} \]

Hence, we have that

\[ Q_i(N, K, J) = \frac{N P_J(N - J + 1)^{i-1} N^{K - J + 1 - i}}{N P_J(N - J + 1)^{i-2} N^{K - J + 2 - i}} = \frac{N - J + 1}{N}, \quad i \in \{3, \ldots, K - J + 1\}, \tag{A.8} \]

where the first equality holds by (A.5) and (A.7). Equations (A.6) and (A.8) match (13).

To derive (14), observe the following facts. First, individual \( J \) is fixed as the \( (J-1)^{th} \) member for \( B_{K-J+2} = \{1, \ldots, J - 2, J, J + 1\} \) through \( B_{2(K-J)+1} = \{1, \ldots, J - 2, J, K\} \). Second, individual \( J + 1 \) is fixed as the \( (J-1)^{th} \) member for \( B_{2(K-J)+2} = \{1, \ldots, J - 2, J + 1, J + 2\} \) through \( B_{3(K-J)} = \{1, \ldots, J - 2, J + 1, K\} \). Finally, individual \( K - 1 \) is the \( (J-1)^{th} \) member for \( B_{0.5(K-J+1)(K-J)+2} \), where

\[ B_{0.5(K-J+1)(K-J)+2} = \{1, \ldots, J - 2, K - 1, K\}. \tag{A.9} \]

To exploit these facts, we collect group indices \( K - J + 2 \) through \( 0.5(K-J+1)(K-J+2) \) into \( K - J \) divisions as shown in (16). First, \( M_1(K, J) \) consists of group indices \( K - J + 2 \) through \( 2(K-J)+1 \), in which the \( (J-1)^{th} \) member is always individual \( J \). Second, \( M_2(K, J) \) consists of group indices \( 2(K-J) + 2 \) through \( 3(K-J) \), in which the \( (J-1)^{th} \) member is always individual \( J + 1 \). Finally, \( M_{K-J}(K, J) \) is equal to group index \( 0.5(K-J+1)(K-J+2) \), in which the \( (J-1)^{th} \) member is individual \( K - 1 \). Note that \( |M_\ell(K, J)| = K - J - \ell + 1 \) for all \( \ell \in \{1, \ldots, K - J\} \).

Focusing on \( i \in M_1(K, J) \), the only incremental restriction added at step \( i \) relative to step \( i - 1 \) is that the last member of \( B_i \) must have a different birthday from individual \( J \). The number of birthdays available for the last member of \( B_i \) is \( N - J + 1 \) at step \( i - 1 \) and \( N - J \) at step \( i \). Hence, we have that \( R_i(N, K, J) / R_{i-1}(N, K, J) = (N - J) / (N - J + 1) \) for any \( i \in M_1(K, J) \). Focusing on \( i \in M_2(K, J) \), the only incremental restriction added at step \( i \) relative to step \( i - 1 \) is that the last member of \( B_i \) must have a different birthday from individual \( J + 1 \). The number of birthdays available for the last member of \( B_i \) is \( N - J \) at step \( i - 1 \) and \( N - J - 1 \) at step \( i \). Hence, \( R_i(N, K, J) / R_{i-1}(N, K, J) = (N - J - 1) / (N - J) \).
for any \( i \in \mathcal{M}_2(K, J) \). Similar argument applies up to \( i = \mathcal{M}_{K-J}(K, J) \), hence we obtain:

\[
\frac{R_i(N, K, J)}{R_{i-1}(N, K, J)} = \begin{cases} 
\frac{N-J}{N-J+1} & \text{if } i \in \mathcal{M}_1(K, J), \\
\frac{N-J-1}{N-J} & \text{if } i \in \mathcal{M}_2(K, J), \\
\vdots & \\
\frac{N-J+1-(K-J)}{N-J+2-(K-J)} & \text{if } i \in \mathcal{M}_{K-J}(K, J).
\end{cases} \tag{A.10}
\]

Using the indicator function, (A.10) can be rewritten in a single equation:

\[
\frac{R_i(N, K, J)}{R_{i-1}(N, K, J)} = \frac{N - J + 1 - \sum_{\ell=1}^{K-J} \ell \times 1 \{i \in \mathcal{M}_\ell(K, J)\}}{N - J + 2 - \sum_{\ell=1}^{K-J} \ell \times 1 \{i \in \mathcal{M}_\ell(K, J)\}},
\]

\[\forall i \in \left\{ K - J + 2, \ldots, \frac{1}{2}(K - J + 1)(K - J + 2) \right\}. \]

Hence, by (A.5), we have established (14).

In view of (A.9), \( \mathcal{B}_{0.5(K-J+1)(K-J+2)} \) is the first group to which both individuals \( K-1 \) and \( K \) belong. This means that any pair of individuals jointly belongs to at least one of \( \{ \mathcal{B}_1, \ldots, \mathcal{B}_{0.5(K-J+1)(K-J+2)} \} \). Hence, \( R_i(N, K, J) \) reaches its lower bound, \( N P_K \), when \( i = 0.5(K - J + 1)(K - J + 2) \). For \( i \geq 0.5(K - J + 1)(K - J + 2) + 1 \), no incremental restrictions are added because any pair of individuals is already required to have distinct birthdays. Hence, we have that

\[ R_i(N, K, J) = N P_K \quad \forall i \in \left\{ \frac{1}{2}(K - J + 1)(K - J + 2), \ldots, K C_J \right\}. \]

Hence, by (A.5), we obtain (15) and it completes the proof of Theorem 1.

### A.2 Proof of Corollary 1

For any group size \( J \in \{2, \ldots, K\} \), we have that

\[
\frac{1}{2}(K-J+1)(K-J+2) \prod_{i=K-J+2} Q_i(N, K, J)
\]

\[= \left( \frac{N - J}{N - J + 1} \right)^{K-J} \left( \frac{N - J - 1}{N - J} \right)^{K-J-1} \times \cdots \times \frac{N - K + 1}{N - K + 2} \tag{A.11} \]
\[ \frac{(N-J)(N-J-1) \times \cdots \times (N-K+1)}{(N-J+1)^{K-J}} = \frac{N-J^{P_{K-J}}}{(N-J+1)^{K-J}}, \quad (A.12) \]

where (A.11) follows by (14), (16), and the fact that \(|M_\ell(K, J)| = K - J - \ell + 1\) for all \(\ell \in \{1, \ldots, K - J\}\). Substitute (12), (13), (15), and (A.12) into (8) to obtain

\[ Q^*(N, K) = \frac{NP_J}{N^J} \left( \frac{N-J+1}{N} \right)^{K-J} \frac{N-J^{P_{K-J}}}{(N-J+1)^{K-J}} = \frac{NP_K}{N^K} = Q^*(N, K), \]

where the last equality follows by (3).

### A.3 Proof of Theorem 2

Equation (17) implies that \(\{Q_i(N, K, 2)\}_{i=1}^{K C_2}\) is decreasing in a step-function form; \(Q_i(N, K, 2)\) is equal to \((N-1)/N\) for \(i \in \{1, \ldots, K-1\}\), decreases to \((N-2)/(N-1)\) at \(i = K\), further decreases to \((N-3)/(N-2)\) at \(i = 2K-2\), and finally decreases to \((N-K+1)/(N-K+2)\) at \(i = K C_2\). Hence, \(\{Q_i(N, K, J)\}_{i=1}^{K C_J}\) is monotonically decreasing if \(J = 2\). When \(J \geq 3\), \(\{Q_i(N, K, J)\}\) cannot be monotonically decreasing. Equations (12)-(13) imply that

\[ Q_1(N, K, J) = \frac{NP_{J-1}}{N^{J-1}} \times Q_2(N, K, J) < Q_2(N, K, J). \quad (A.13) \]

Thus, \(\{Q_i(N, K, J)\}\) is monotonically decreasing if and only if \(J = 2\).

### A.4 Proof of Theorem 3

To prove the sufficient part of Theorem 3, suppose that \(K = J + 1\). Since \(J \geq 3\), we have by (A.13) that \(Q_1(N, K, J) < Q_2(N, K, J)\). Since \(K = J + 1\), (14) is simplified to

\[ Q_3(N, K, J) = \frac{(N-J+1)^2}{N(N-J)} \times Q_3(N, K, J) \]

\[ = \left\{ 1 - \frac{(J-2)(N-J-1)}{N(N-J)} \right\} Q_3(N, K, J) \]

\[ \leq Q_3(N, K, J), \quad (A.14) \]

where the first equality follows by (13); the second equality follows by direct division; the inequality follows by the assumption that \(J \geq 3\) and \(N-J \geq 1\); the inequality becomes an equality if and only if \((N, J) = (4, 3)\). Further, (15) is simplified to \(Q_i(N, K, J) = 1\)
for \( i \in \{4, \ldots, K\} \). Hence, \( \{Q_i(N, K, J)\}_{i=1}^{K+J} \) is monotonically increasing if \( K = J + 1 \), completing a proof of the sufficient part of Theorem 3.

To prove the necessary part, suppose that \( \{Q_i(N, K, J)\}_{i=1}^{K+J} \) is monotonically increasing. Observe that (14) has a strictly decreasing component if and only if \( K > J + 1 \). When \( K > J + 1 \), \( M_\ell(K, J) \) in (16) is well defined for at least \( \ell \in \{1, 2\} \). Equation (14) implies that \( \{Q_i(N, K, J)\}_{i=K-J+1}^{K+J+2} \) is decreasing in a step-function form, where \( i^*(K - J) = 0.5(K - J + 1)(K - J + 2) \) as defined in Corollary 3. The decay of the step function occurs each time group index \( i \) arrives at the first element of \( M_\ell(K, J) \) with \( \ell \in \{2, \ldots, K - J\} \). In particular, \( Q_{2(K-J)+1}(N, K, J) > Q_{2(K-J)+2}(N, K, J) \) as group \( 2(K - J) + 1 \) is the last element of \( M_1(K, J) \). Hence, it must be the case that \( K = J + 1 \) if \( \{Q_i(N, K, J)\}_{i=1}^{K+J} \) is monotonically increasing. It completes a proof of the necessary part of Theorem 3.

### A.5 Proof of Theorem 4

To prove that item \((i)\) implies item \((ii)\), suppose that \( K > J + 1 \). Since \( J \geq 3 \), we have by (A.13) that \( Q_1(N, K, J) < Q_2(N, K, J) \). Since \( N > J + 1 \), a strict version of (A.14) holds: \( Q_{K-J+1}(N, K, J) < Q_{K-J+2}(N, K, J) \). Hence, \( \{Q_i(N, K, J)\}_{i=1}^{K-J+2} \) is monotonically increasing. Since \( K > J + 1 \), \( M_\ell(K, J) \) in (16) is well defined for at least \( \ell \in \{1, 2\} \). Equation (14) implies that \( \{Q_i(N, K, J)\}_{i=K-J+1}^{K+J+2} \) is decreasing in a step-function form. The decay of the step function occurs each time group index \( i \) arrives at the first element of \( M_\ell(K, J) \) with \( \ell \in \{2, \ldots, K - J\} \). The maximum value of \( \{Q_i(N, K, J)\}_{i=1}^{i^*(K-J)} \) is given by \( \{Q_i(N, K, J)\}_{i=M_1(K,J)} \). By (14), \( Q_i(N, K, J) = (N - J)/(N - J + 1) \) for \( i \in M_1(K, J) \). Thus, item \((i)\) implies item \((ii)\) as desired. To prove the converse, impose item \((ii)\). Then, \( \{Q_i(N, K, J)\}_{i=K-J+1}^{i^*(K-J)} \) is monotonically decreasing. By Theorem 3, it must be the case that \( K > J + 1 \). Hence, item \((ii)\) implies item \((i)\) as desired.