

ON THE BIRTHDAY PROBLEM: SOME GENERALIZATIONS AND APPLICATIONS

P. N. RATHIE and P. ZÖRNIC

Received 3 October 2001

We study the birthday problem and some possible extensions. We discuss the unimodality of the corresponding exact probability distribution and express the moments and generating functions by means of confluent hypergeometric functions $U(-; -; -)$ which are computable using the software Mathematica. The distribution is generalized in two possible directions, one of them consists in considering a random graph with a single attracting center. Possible applications are also indicated.

2000 Mathematics Subject Classification: 60C05, 05A10, 05C05.

1. Introduction. In the present paper, we study the birthday problem which may be formulated in various manners: consider, for example, a smoker who draws matches from a matchbox that contains initially n (unused) matches. Whenever he needs a match, he draws one at random and returns the used match into the box. He continues drawing matches until a used one is encountered for the first time. The probability $p_k^{(n)}$ that the first used match is encountered in the k th draw is given by (see, e.g., [5, equation (7.1), page 47])

$$p_k^{(n)} = \frac{n!(k-1)}{(n-k+1)!n^k} \quad \text{for } k = 2, \dots, n+1. \quad (1.1)$$

For $n = 365$, (1.1) represents the probability that among k persons can be found two persons with the same birthday (in this case the matches correspond to the 365 days of the year).

The distribution (1.1) also plays an important role in the study of characteristics of random mappings (random graphs): we consider a random mapping T of a finite set $X = \{1, 2, \dots, n\}$ into itself, assuming that T assigns independently to each $x \in X$ its unique image $y \in X$ with probability $1/n$ for all $y \in X$. Denote by G the directed graph with vertex set X , containing an arrow from x to y if and only if $T(x) = y$. The set of successors of the element $x \in X$ in T is defined as

$$S_T(x) = \{x, T(x), T^2(x), \dots, T^{n-1}(x)\}, \quad (1.2)$$

where $T^0(x) = x$ and $T^k(x) = T(T^{k-1}(x))$ for $k \geq 1$. Denoting by $s_T(x)$ the number of elements in $S_T(x)$, the probability that a vertex x has k successors is given by

$$P(s_T(x) = k) = \frac{(n-1)!k}{(n-k)!n^k} = p_{k+1}^{(n)} \quad (1.3)$$

(see [8, equation (3.2), page 1047]).

The distribution (1.1) and diverse extensions have been intensively studied (see, e.g., [1, 3, 4, 5, 7, 8, 10, 11, 12, 13, 17, 20, 21, 22, 23, 24, 25, 26, 27]).

In the next section, we study some basic properties of distribution (1.1) and indicate possible applications. Moments and generating functions are determined in Section 3, using the well-known U -function. In Section 4, we study a very general type of distribution defined by means of directed weighted trees, which generalizes (1.1). Among others, these results yield a recurrence relation for the moments of distribution (1.1). In Section 5, we generalize (1.1), considering the distribution of the number of successors in a random graph with a single attracting center.

2. Some basic properties of the distribution. The following proposition proves that the distribution in (1.1) is unimodal. (To simplify the notation, we write p_k instead of $p_k^{(n)}$ in the following.)

PROPOSITION 2.1. *The probabilities p_k satisfy*

$$p_2 < \cdots < p_{k_0} \geq p_{k_0+1} > \cdots > p_{n+1}, \quad (2.1)$$

where k_0 is the smallest integer greater than or equal to $1/2 + \sqrt{n+1/4}$.

Moreover, $p_{k_0} = p_{k_0+1}$ if and only if $1/2 + \sqrt{n+1/4}$ is an integer.

PROOF. We have $p_{k-1} < p_k$ if and only if

$$\frac{n!(k-2)}{(n-k+2)!n^{k-1}} < \frac{n!(k-1)}{(n-k+1)!n^k}. \quad (2.2)$$

Multiplying this inequality by $(n-k+2)!n^k/n!$ results in the equivalent condition $(k-2)n < (k-1)(n-k+2)$ which can be written as $k^2 - 3k - (n-2) < 0$ or $k < 3/2 + \sqrt{n+1/4}$. Hence, $p_{k-1} < p_k$ if $k \leq k_0$. Analogously, we get $p_k > p_{k+1}$ for $k \geq k_0$ if $1/2 + \sqrt{n+1/4}$ is not an integer. If this number is an integer, we get $p_{k_0} = p_{k_0+1}$ and $p_k > p_{k+1}$ for $k \geq k_0 + 1$. \square

Beyond the index k_0 above it is interesting to determine the index k_1 ($k_1 \geq k_0$) which maximizes the difference

$$\Delta_k = p_k - p_{k+1} = \frac{(n-1)!}{(n-k+1)!n^k} (k^2 - k - n). \quad (2.3)$$

This index is an equivalent to a turning point of a distribution in the continuous case. Similar to the above consideration, it can be shown that $\Delta_k < \Delta_{k+1}$ if

$$(3n + 1)k - k^3 - n \tag{2.4}$$

is positive and that $\Delta_k > \Delta_{k+1}$ if (2.4) is negative. We get the following proposition.

PROPOSITION 2.2. *The index k_1 is the smallest integer k such that (2.4) is nonpositive.*

For large numbers n , the indices k_0 and k_1 satisfy $k_0 \approx \sqrt{n}$ and $k_1 < \sqrt{3n}$ (see (2.4)), that is, there is an increasing degree of antisymmetry. If n converges to ∞ , p_{k_0} (and thus all p_k) converges to 0 since for large n ,

$$\begin{aligned} p_{k_0} &= \frac{n!(k_0 - 1)}{(n - k_0 + 1)!n^{k_0}} = \frac{n \cdots (n - k_0 + 2)}{n^{k_0}} (k_0 - 1) \\ &\leq \frac{n^{k_0 - 1}}{n^{k_0}} (k_0 - 1) = \frac{k_0 - 1}{n} \approx \frac{1}{\sqrt{n}}. \end{aligned} \tag{2.5}$$

The distribution (1.1) might have practical applications. Among others, it could be useful to estimate the size of a population. Suppose, for example, that biologists want to estimate the number of fishes of a certain species in a lake. They could proceed as follows: catching a fish repeatedly, mark it every time and put it back into the water. The number k of fishes (of the considered species) that had to be caught until the first marked fish was encountered represents the test statistics. Let n denote the (unknown) size of the fish population, the observed value k occurs with probability p_k (see (1.1)). By means of the maximum likelihood method, the number of fishes can be estimated by the value of n that maximizes p_k (k fixed).

EXAMPLE 2.3. We assume that $k = 101$; for example, when the 101th fish is caught, a previously marked fish is encountered for the first time. This event occurs with probability

$$p_{101}^{(n)} = \frac{(n - 1)!100}{(n - 100)!n^{100}} \tag{2.6}$$

if n denotes the size of the fish population. Generally, $p_k^{(n)}$ is a unimodal function in n for every fixed k and

$$\frac{p_k^{(n+1)}}{p_k^{(n)}} = \frac{n}{n + 2 - k} \left(\frac{n}{n + 1} \right)^{k-1}, \tag{2.7}$$

that is,

$$p_k^{(n+1)} (\geq) p_k^{(n)} \iff \frac{n}{n+2-k} \left(\frac{n}{n+1} \right)^{k-1} (\geq) 1. \tag{2.8}$$

This implies $p_{101}^{(5017)} < p_{101}^{(5016)} > p_{101}^{(5015)}$, that is, $p_{101}^{(n)}$ is maximized for $n = 5016$ which is the estimator for the size of the fish population.

As the above calculations illustrate, the population size is approximately given by $n = k^2/2$ for a sufficiently large population. In this case we have (see [8, page 1047])

$$\frac{(n-1)!k}{(n-k)!n^k} \approx \frac{k}{n} e^{-k^2/(2n)}. \tag{2.9}$$

The right-hand side of this relation is maximized for $n = k^2/2$ when k is fixed.

The above method represents an alternative to the estimating procedure for a fish population described in [9, Chapter 14].

Further applications of the probability distribution (1.1) are conceivable in quantitative linguistics, in the framework of testing the hypothesis that there is a tendency in language to repeat already used text elements (see [29]).

3. Moments and generating functions. In order to express the moments we make use of the well-known confluent hypergeometric function $U(r; r + n; x)$ defined by (see, e.g., [19, page 116] and [28, page 41])

$$U(r; r + n; x) = \sum_{k=0}^{n-1} \frac{r^{(k)} (n-k)^{(k)}}{k!} x^{-k-r}, \quad n, r = 1, 2, \dots, x > 0, \tag{3.1}$$

where $\alpha^{(k)} = \alpha(\alpha + 1) \cdots (\alpha + k - 1)$ denotes the rising factorial. Computation of this function is easily done by using the software Mathematica. For the m th rising factorial moment ($m = 0, 1, \dots$), we get

$$\begin{aligned} E[X^{(m)}] &= E[X(X+1) \cdots (X+m-1)] \\ &= \sum_{k=2}^{n+1} k(k+1) \cdots (k+m-1) \frac{n!(k-1)}{(n-k+1)!n^k} \\ &= \sum_{k=2}^{n+1} \frac{(k+m-1)!}{(k-2)!} \frac{n!}{(n-k+1)!n^k} \\ &= \frac{(m+1)!}{n} \sum_{k=0}^{n-1} \frac{(k+m+1)!(n-1)!}{(m+1)!k!(n-k-1)!n^k} \\ &= \frac{(m+1)!}{n} \sum_{k=0}^{n-1} \frac{(m+2)^{(k)} (n-k)^{(k)}}{k!} \left(\frac{1}{n} \right)^k \\ &= (m+1)!n^{m+1}U(m+2; m+n+2; n). \end{aligned} \tag{3.2}$$

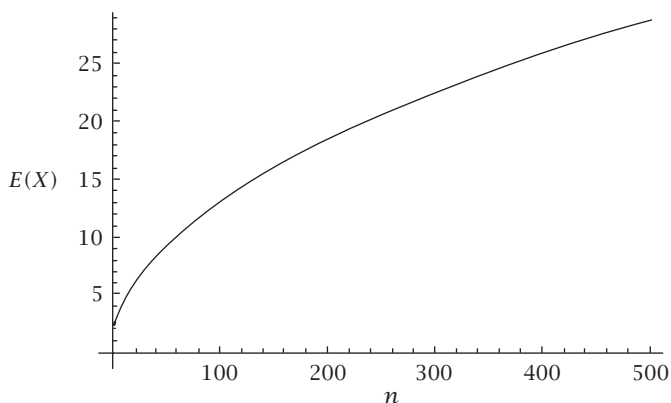


FIGURE 3.1. The dependence of the mean $E(X)$ on n .

For $m = 0$, we obtain

$$nU(2;n+2;n) = 1, \tag{3.3}$$

a result which is not readily available in the literature on special functions. For $m = 1$, we get the mean

$$E(X) = 2n^2U(3;n+3;n). \tag{3.4}$$

Alternatively,

$$E(X) = \frac{n!}{n^n} \sum_{k=0}^n \frac{n^k}{k!} = nU(1;n+2;n) = \left(\frac{e}{n}\right)^n \Gamma(n+1, n), \tag{3.5}$$

where $\Gamma(\alpha, x)$ is the incomplete gamma function [28, page 97]. An alternative exact and asymptotic result for a general case, for the expected number of repetitions necessary for one of the alternatives to occur a certain number of times, was given by Klamkin and Newman [17].

Thus, we get

$$2nU(3;n+3;n) = U(1;n+2;n), \tag{3.6}$$

another result not easily available.

The dependence of the mean $E(X)$ on n is illustrated in Figure 3.1. It may be noted that for $n = 365$, $E(X) = 24,6166$.

The variance of X is given by

$$\begin{aligned} V(X) &= E[X(X+1)] - E[X] - (E(X))^2 \\ &= 6n^3U(4;n+4;n) - 2n^2U(3;n+3;n) - 4n^4U^2(3;n+3;n). \end{aligned} \tag{3.7}$$

The moments about the origin can be calculated using the relation (see [16, page 6])

$$x^n = \sum_{k=0}^n \bar{S}(n, k)x^{(k)}, \tag{3.8}$$

where $\bar{S}(n, k)$ denote the Stirling numbers of the fourth kind. We obtain the r th moments of X as a linear combination of the polynomial $U(-; -; n)$:

$$\begin{aligned} \mu'_r &= E(X^r) = \sum_{k=0}^r \bar{S}(r, k)E[X^{(k)}] \\ &= \sum_{k=0}^r \bar{S}(r, k)(k+1)!n^{k+1}U(k+2; k+n+2; n). \end{aligned} \tag{3.9}$$

Since $U(-; -; n)$ can be expressed as a linear combination of incomplete gamma functions, so is μ'_r in a complicated manner. Dwass [4] considered higher asymptotic moments of X . Holst [10] commented that moments of order statistics from the gamma distribution can be used to get higher moments.

The central moments μ_r are obtained from the relation

$$\mu_r = E(X - \mu)^r = \sum_{k=0}^r (-1)^k \binom{r}{k} \mu^k \mu'_{r-k}. \tag{3.10}$$

Finally, the probability generating function $P(s)$ and the moment generating function $M_X(t)$ can be expressed as

$$P(s) = E(s^X) = \sum_{k=2}^{n+1} p_k s^k = \sum_{k=2}^{n+1} \frac{n!(k-1)s^k}{(n-k+1)!n^k} = nU\left(2; n+2; \frac{n}{s}\right), \tag{3.11}$$

$$M_X(t) = E(e^{tX}) = \sum_{k=2}^{n+1} \frac{n!(k-1)e^{tk}}{(n-k-1)!n^k} = nU(2; n+2; ne^{-t}), \tag{3.12}$$

respectively, and the characteristic function is obtained from (3.12), substituting it for t , where $i = \sqrt{-1}$.

4. A general type of distribution. Let T be a directed tree with the root 0 such that all arcs are leaving from the root, that is, 0 is the unique source and $\delta^-(v) = 1$ for every vertex $v \neq 0$ ($\delta^-(v)$ denotes the indegree of v , that is, the number of arcs entering into v). We associate a positive weight with all arcs of T , such that the sum of the weights of all arcs leaving a vertex is 1. Furthermore, we associate a weight $w(v)$ with every vertex v of T such that $w(0) = 1$ and

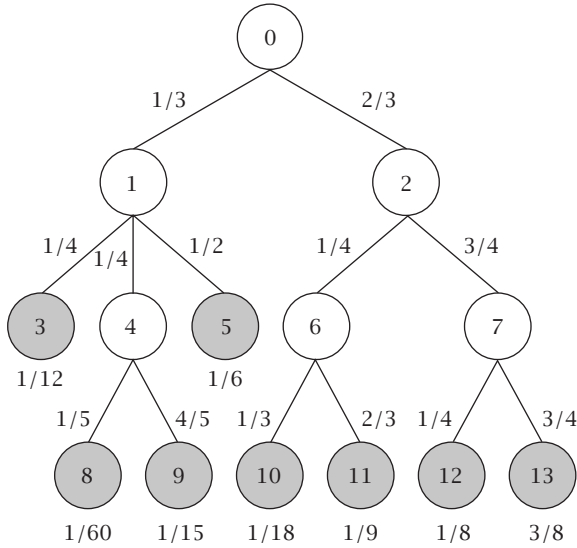


FIGURE 4.1. Example of a directed weighted tree.

for every vertex $v \neq 0$, we define $w(v)$ as the product of the weights of the arcs on the unique path from 0 to v . An example of a directed weighted tree is shown in Figure 4.1, where the vertex weights are only indicated when the vertex is a leaf (represented by shaded circles).

PROPOSITION 4.1. *For every directed weighted tree of the above type, the weights of the leaves sum up to 1.*

Then result can be easily proved by induction.

Obviously, every tree of the above type defines a discrete probability distribution when the weights of the leaves are interpreted as the probabilities. We consider now the tree in Figure 4.2 with the vertices $0, 1, \dots, n, \ell_2, \dots, \ell_{n+1}$, where $\ell_i, i = 2, \dots, n + 1$, represent the leaves.

The weights of the leaves (probabilities) are

$$p_k = \begin{cases} 1 - c_1, & \text{for } k = 2, \\ c_1 c_2 \cdots c_{k-2} (1 - c_{k-1}), & \text{for } 3 \leq k \leq n, \\ c_1 c_2 \cdots c_{n-1}, & \text{for } k = n + 1, \end{cases} \tag{4.1}$$

which obviously sum up to 1. For $c_i = (n - i)/n, i = 1, \dots, n - 1$, we obtain the probability distribution (1.1) as a special case. In this case a vertical (horizontal) arc represents the selection of an unused (used) match; the vertex i can be interpreted as the stage of the system with i used matches in the box, while ℓ_i represents the encounter of a used match for the first time in the i th draw.

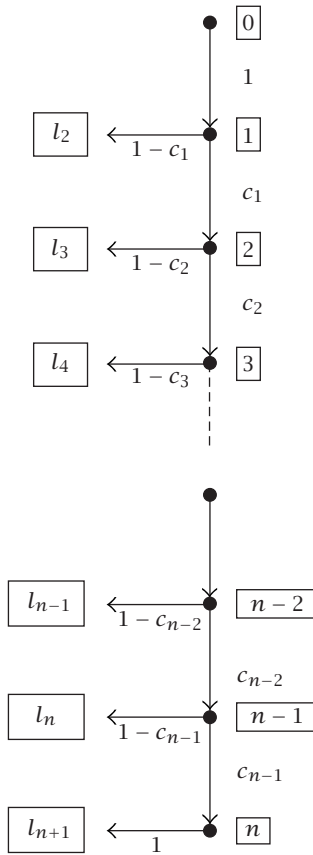


FIGURE 4.2. A tree representing a probability distribution.

The moments about the origin of distribution (4.1) can be expressed as

$$\begin{aligned}
 \mu'_r &= E(X^r) \\
 &= 2^r(1 - c_1) + \sum_{i=3}^n c_1 c_2 \cdots c_{i-2} (1 - c_{i-1}) i^r + (n + 1)^r c_1 c_2 \cdots c_{n-1} \quad (4.2) \\
 &= 1 + \sum_{i=1}^n [(i + 1)^r - i^r] c_1 c_2 \cdots c_{i-1}.
 \end{aligned}$$

Using (4.2), we finally state some interrelations between the moments about the origin of distribution (1.1). Thus, we obtain

$$\mu'_r = 1 + \sum_{i=1}^n [(i + 1)^r - i^r] \frac{(n - 1)!}{(n - i)! n^{i-1}}. \quad (4.3)$$

Alternatively the μ'_r can be written as (see (1.1))

$$\mu'_r = \sum_{k=2}^{n+1} \frac{(n-1)!(k-1)k^r}{(n-k+1)!n^{k-1}} = \sum_{i=1}^n \frac{(n-1)!i(i+1)^r}{(n-i)!n^i}. \tag{4.4}$$

Defining $b_i = (n-1)/(n-i)!n^{i-1}$ and $T_r = \sum_{i=1}^n b_i i^r$, relations (4.3) and (4.4) read

$$\mu'_r = 1 + \sum_{i=1}^n [(i+1)^r - i^r] b_i = 1 + \sum_{i=1}^n \sum_{j=0}^{r-1} \binom{r}{j} i^j b_i = 1 + \sum_{j=0}^{r-1} \binom{r}{j} T_j, \tag{4.5}$$

$$\mu'_r = \frac{1}{n} \sum_{i=1}^n i(i+1)^r b_i = \frac{1}{n} \sum_{i=1}^n \sum_{j=0}^r \binom{r}{j} i^{j+1} b_i = \frac{1}{n} \sum_{j=0}^r \binom{r}{j} T_{j+1}. \tag{4.6}$$

Combining (4.5) and (4.6) yields the following recurrence relations for T_r and μ'_r , respectively:

$$T_{r+1} = n - rT_r + \sum_{j=0}^{r-1} \left[n \binom{r}{j} - \binom{r}{j-1} \right] T_j \quad \text{for } r = 0, 1, \dots, \tag{4.7}$$

$$\mu'_r = (n+1)\mu'_{r-1} + T_{r-1} - T_r \quad \text{for } r = 1, 2, \dots$$

For small values of r , (4.7) gives

$$T_1 = n, \quad T_2 = nT_0, \quad T_3 = 2n^2 - nT_0, \quad T_4 = n(1+3n)T_0 - 3n^2,$$

$$\mu'_1 = T_0 + 1, \quad \mu'_2 = T_0 + 2n + 1, \quad \mu'_3 = (3n+1)T_0 + 3n + 1, \tag{4.8}$$

$$\mu'_4 = (2n+1)T_0 + 8n^2 + 4n + 1,$$

where T_0 can be computed as (see Section 3)

$$T_0 = \mu'_1 - 1 = E(X) - 1 = \frac{n!}{n^n} \sum_{k=0}^{n-1} \frac{n^k}{k!} \tag{4.9}$$

$$= 2n^2 U(3; n+3; n) - 1.$$

5. Generalization of the distribution by means of an attracting center. We now generalize the random mapping considered in the introduction, assuming that the vertex 1 is an attracting center, that is, T assigns independently to each $x \in X$ the image 1 with probability q and the image y with probability

$p = (1 - q)/(n - 1)$ for $y = 2, \dots, n$ (see, e.g., [14, page 191]; related models of random graphs are studied in [2, 15]). The probability that the attracting center 1 has k successors is given by

$$\begin{aligned} P(s(1) = k) &= P(T^r(x) \neq x, T(x), \dots, T^{r-1}(x) \quad (0 < r \leq k-1); \\ &\quad T^k(x) = T^j(x) \text{ for a } j \text{ with } 0 \leq j \leq k-1) \\ &= (1-q)(1-q-p)(1-q-2p) \cdots (1-q-(k-2)p)[q+(k-1)p] \\ &= [q+(k-1)p] \prod_{j=0}^{k-2} (1-q-jp). \end{aligned} \tag{5.1}$$

Since

$$1 - q - jp = p(n - 1 - j), \tag{5.2}$$

we get

$$\begin{aligned} P(s(1) = k) &= [q+(k-1)p] p^{k-1} \prod_{j=0}^{k-2} (n-1-j) \\ &= [q+(k-1)p] p^{k-1} \frac{(n-1)!}{(n-k)!} \quad \text{for } k = 1, \dots, n. \end{aligned} \tag{5.3}$$

We now determine the probability $P(s(x) = k)$, where $x \in \{2, \dots, n\}$, that is, x is not the attracting center. Two cases have to be distinguished depending on whether the attracting center is contained in the successor set or not. In the first case we get, assuming that $T^i(x) = 1$ for a fixed $i \in \{1, \dots, k-1\}$,

$$\begin{aligned} P(s(x) = k | T^i(x) = 1) &= (1-q-p)(1-q-2p) \cdots [1-q-(i-1)p] q [1-q-ip] \\ &\quad \cdots [1-q-(k-2)p] [q+(k-1)p] \\ &= [q+(k-1)p] q \prod_{j=1}^{k-2} (1-q-jp) \\ &= [q+(k-1)p] q p^{k-2} \prod_{j=1}^{k-2} (n-1-j) \quad (\text{see (5.2)}) \\ &= [q+(k-1)p] q p^{k-2} \frac{(n-2)!}{(n-k)!}. \end{aligned} \tag{5.4}$$

The probability that x has k successors different from 1 is given by

$$\begin{aligned}
 P(s(x) = k | T^i(x) \neq 1 \text{ for } i = 1, \dots, k-1) &= (1 - q - p)(1 - q - 2p) \cdots [1 - q - (k-1)p]kp \\
 &= kp \prod_{j=1}^{k-1} (1 - q - jp) \\
 &= kp p^{k-1} \prod_{j=1}^{k-1} (n-1-j) \\
 &= kp^k \frac{(n-2)!}{(n-k-1)!}.
 \end{aligned}
 \tag{5.5}$$

Finally, we get, for $x \neq 1$,

$$\begin{aligned}
 P(s(x) = k) &= P(s(x) = k | T^i(x) \neq 1 \text{ for } i = 1, \dots, k-1) \\
 &\quad + \sum_{i=1}^{k-1} P(s(x) = k | T^i(x) = 1) \\
 &= kp^k \frac{(n-2)!}{(n-k-1)!} + (k-1)[q + (k-1)p]qp^{k-2} \frac{(n-2)!}{(n-k)!} \\
 &= \{(n-k)kp^2 + (k-1)q[q + (k-1)p]\}p^{k-2} \frac{(n-2)!}{(n-k)!}.
 \end{aligned}
 \tag{5.6}$$

Rewriting the last term, we get

$$\begin{aligned}
 P(s(x) = k) &= \{k^2p(q-p) + k(np^2 + q^2 - 2qp) \\
 &\quad + q(p-q)\}p^{k-2} \frac{(n-2)!}{(n-k)} \quad \text{for } k = 1, \dots, n.
 \end{aligned}
 \tag{5.7}$$

Obviously the two distributions in (5.3) and (5.7) result in $p_{k+1}^{(n)}$ (see (1.1)) for $q = p = 1/n$.

The dependence of the distributions (5.3) and (5.7) on the choice of q is illustrated in Figures 5.1 and 5.2 for $n = 100$. It is observed that (5.7) has the same unimodal form for all q ($0 < q < 1$), while (5.3) is unimodal for $q < 0, 1$ and 1/2-modal for $q \geq 0, 1$.

Using the relation

$$\frac{(n-1)!k}{(n-k)!n^k} \approx \frac{k}{n} e^{-k^2/(2n)} \quad \text{for large } n,
 \tag{5.8}$$

(see Example 2.3) we obtain

$$P(s(1) = k) \approx [q + (k-1)p](pn)^{k-1} e^{-k^2/(2n)},
 \tag{5.9}$$

$$P(s(x) = k) \approx \{k^2p(q-p) + k(np^2 + q^2 - 2qp) + q(p-q)\} \frac{n}{n-1} (pn)^{k-2} e^{-k^2/(2n)}.
 \tag{5.10}$$

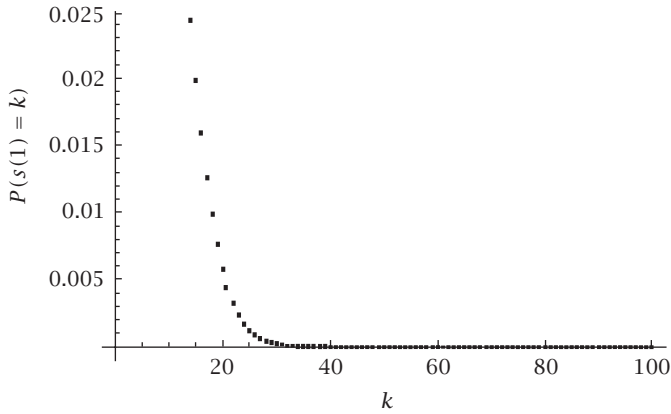


FIGURE 5.1. The distribution (5.3) for $n = 100$ and $q = 0, 1$ (1/2-modal).

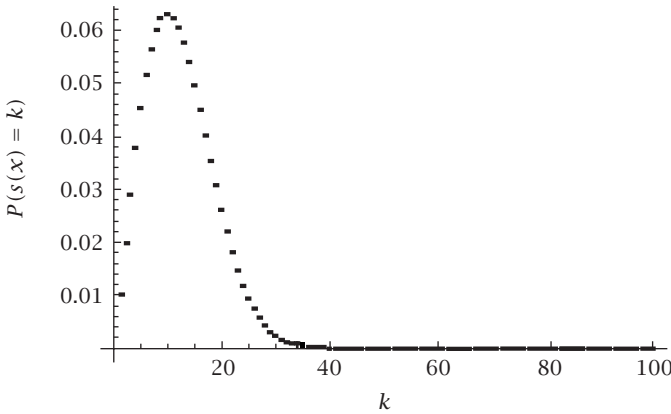


FIGURE 5.2. The distribution (5.3) or (5.7) for $n = 100$ and $q = 0, 1$ (unimodal).

For continuous values of k ($k > 0$), the right-hand sides in (5.9) and (5.10) represent the asymptotic densities of the number of successors of the attractor center 1 or a vertex $x \neq 1$, respectively.

The results of this section might be interesting in random-number generation. It may be assumed that in the generation of random numbers a sequence $x, T(x), T^2(x), \dots$ is constructed (see, e.g., [6] and [18, Section 3.1]), where $x \in X = \{1, 2, \dots, n\}$ and T is a mapping selected at random from all the n^n mappings from X into itself.

If all the n^n mappings are chosen with equal probability, $p_{k+1}^{(n)}$ gives the probability that the process begins to cycle after k iterations, that is, that the numbers $x, T(x), \dots, T^{k-1}(x)$ are pairwise different while $T^k(x) = T^i(x)$ for some $i \in \{0, 1, \dots, k-1\}$.

Formulas (5.3) and (5.7) give the probability of cycling after k iterations, when mappings T that assign frequently the image 1 are preferred.

REFERENCES

- [1] G. C. Berresford, *The uniformity assumption in the birthday problem*, Math. Mag. **53** (1980), no. 5, 286–288.
- [2] Y. D. Burtin, *On a simple formula for random mappings and its applications*, J. Appl. Probab. **17** (1980), no. 2, 403–414.
- [3] M. Camarri and J. Pitman, *Limit distributions and random trees derived from the birthday problem with unequal probabilities*, Electron. J. Probab. **5** (2000), no. 1, 1–18.
- [4] M. Dwass, *More birthday surprises*, J. Combinatorial Theory **7** (1969), 258–261.
- [5] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. I*, John Wiley & Sons, New York, 1964.
- [6] P. Flajolet and A. M. Odlyzko, *Random mapping statistics*, Advances in Cryptology—EUROCRYPT '89 (Houthalen, 1989) (J.-J. Quisquater and J. Vandewalle, eds.), Lecture Notes in Comput. Sci., vol. 434, Springer, Berlin, 1990, pp. 329–354.
- [7] M. H. Gail, G. H. Weiss, N. Mantel, and S. J. O'Brien, *A solution to the generalized birthday problem with application to allozyme screening for cell culture contamination*, J. Appl. Probab. **16** (1979), no. 2, 242–251.
- [8] B. Harris, *Probability distributions related to random mappings*, Ann. Math. Statist. **31** (1960), 1045–1062.
- [9] M. Hollander and F. Proschan, *The Statistical Exorcist*, Marcel Dekker, New York, 1984.
- [10] L. Holst, *On birthday, collectors', occupancy and other classical urn problems*, Internat. Statist. Rev. **54** (1986), no. 1, 15–27.
- [11] ———, *The general birthday problem*, Random Structures Algorithms **6** (1995), no. 2-3, 201–208.
- [12] L. Holst and J. Hüsler, *Sequential urn schemes and birth processes*, Adv. in Appl. Probab. **17** (1985), no. 2, 257–279.
- [13] J. Hubert and T. V. Narayana, *A note on the birthday problem*, Bul. Inst. Politehn. Iași (N.S.) **15** (19) (1969), no. 1-2, 113–118.
- [14] J. Jaworski, *On a random mapping (T, P_j)* , J. Appl. Probab. **21** (1984), no. 1, 186–191.
- [15] ———, *Epidemic processes on digraphs of random mappings*, J. Appl. Probab. **36** (1999), no. 3, 780–798.
- [16] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Discrete Multivariate Distributions*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, 1997.
- [17] M. S. Klamkin and D. J. Newman, *Extensions of the birthday surprise*, J. Combinatorial Theory **3** (1967), 279–282.
- [18] D. E. Knuth, *The Art of Computer Programming, Vol 2: Seminumerical Algorithms*, 3rd ed., Addison-Wesley, Massachusetts, 1998.
- [19] Y. L. Luke, *The Special Functions and Their Approximations, Vol. I*, Mathematics in Science and Engineering, vol. 53, Academic Press, New York, 1969.
- [20] S. Mase, *Approximations to the birthday problem with unequal occurrence probabilities and their application to the surname problem in Japan*, Ann. Inst. Statist. Math. **44** (1992), no. 3, 479–499.
- [21] F. H. Mathis, *A generalized birthday problem*, SIAM Rev. **33** (1991), no. 2, 265–270.

- [22] J. Naus, *An extension of the birthday problem*, Amer. Statist. **22** (1968), 27-29.
- [23] ———, *Probabilities for a generalized birthday problem*, J. Amer. Statist. Assoc. **69** (1974), 810-815.
- [24] J. Riordan, *Combinatorial Identities*, John Wiley & Sons, New York, 1968.
- [25] P. F. Rust, *The effect of leap years and seasonal trends on the birthday problem*, Amer. Statist. **30** (1976), 197-198.
- [26] B. Saperstein, *The generalized birthday problem*, J. Amer. Statist. Assoc. **67** (1972), 425-428.
- [27] M. Sayrafiezadeh, *The birthday problem revisited*, Math. Mag. **67** (1994), no. 3, 220-223.
- [28] L. J. Slater, *Confluent Hypergeometric Functions*, Cambridge University Press, New York, 1960.
- [29] P. Zörnig, *A theory of distances between like elements in a sequence*, Glottometrika **8**, Quantitative Linguistics, vol. 32, Brockmeyer, Bochum, 1987, pp. 1-22.

P. N. Rathie: Departamento de Estatística, Instituto de Ciências Exatas, Universidade de Brasília, 70910-900 Brasília, DF, Brazil

E-mail address: rathie@unb.br

P. Zörnig: Departamento de Estatística, Instituto de Ciências Exatas, Universidade de Brasília, 70910-900 Brasília, DF, Brazil

E-mail address: peter@unb.br