Ann. Inst. Statist. Math.
Vol. 40, No. 1, 77-91 (1988)

OCCUPANCY WITH TWO TYPES OF BALLS

KAZUO NISHIMURA AND MASAAKI SIBUYA

Department of Mathematics, Keio University, Hiyoshi, Yokohama 223, Japan

(Received February 13, 1987; revised August 5, 1987)

Abstract. The classical occupancy problem is extended to the case where two types of balls are thrown. In particular, the probability that no urn contains both types of balls is studied. This is a birthday problem in two groups of boys and girls to consider the coincidence of a boy's and a girl's birthday. Let N_1 and N_2 denote the numbers of balls of each type thrown one by one when the first collision between the two types occurs in one of m urns. Then N_1N_2/m is asymptotically exponentially distributed as m tends to infinity.

This problem is related to the security evaluation of authentication procedures in electronic message communication.

Key words and phrases: Urn models, collisions, birthday problem, 2×2 occupancy distribution, Stirling numbers of the second kind, compound binomial distribution, exponential distribution, Rayleigh distribution, cryptography.

1. Introduction

Suppose that balls are thrown at random and independently into one of m urns with the same probability $1/m$. Further, suppose that there are two types of balls to be thrown, say, n_1 white balls and n_2 red balls. As the result there are four types of urns with and without white and red balls. The numbers of urns of these types are represented by a 2×2 contingency table, Table 1.

On the other hand, each ball enters an urn with or without balls of the different color. In the former case "collision between two different colors" occurs. Corresponding to Table 1, the numbers of balls of four types are denoted as shown in Table 2.

The purpose of this report is to first study, in Sections 2 and 3, the joint and marginal distributions of these numbers. Of utmost concern is the number S of urns with balls of both colors, and the probability that $S=0$. This is the probability that $Y_1=Y_2=0$, namely there is no "collision" of balls of different colors within a single urn. If balls are thrown one by one, the numbers N_1 and N_2 of white and red balls, respectively, at the first occurrence

Table 1. Numbers of urns of four types.

urns		contains red balls: yes	contains red balls: no	total
contains white balls	yes	S	R_1	$T_1 = S + R_1$
	no	R_2	$R_3 = m - R_1 - T_2$	$m - T_1$
total		$T_2 = S + R_2$	$m - T_2$	m

$1 \leq T_i \leq \min(n_i, m)$, $i = 1, 2$.

Table 2. Numbers of balls of four types.

balls		collides with red: yes	collides with red: no	white total
collides with white	yes	$Y_2 \setminus Y_1$	$n_1 - Y_1$	n_1
	no	$n_2 - Y_2$	—	—
red total		n_2	—	$n_1 + n_2$

Each entry corresponds to that in Table 1.

of collision are "waiting time" for the collision. And the probability that $S=0$, which depends on m, n_1 and n_2, is the probability that $N_1 > n_1$ or $N_2 > n_2$. In Section 4, after the evaluation of this probability, it is shown that $N_1 N_2 / m$ is asymptotically exponentially distributed.

Suppose that there are two groups, say, n_1 boys and n_2 girls. Assume that their birthdays are independent and uniformly distributed on 365 days. The event $S>0$ means that there is at least one birthday which a boy and a girl have in common. The classical birthday problem, Feller (1968) and Johnson and Kotz (1977), relates to collision within the same color, and is known because of the high probability of a common birthday. Our "birthday problem in two groups" has also high probabilities of (3.4) or (3.5) with $m=365$, as shown in Table 3.

This modified birthday problem stemmed from cryptography. To authenticate a message to be sent through an electronic communication network, the sender compresses the sequence of fragments of the message into a short message, called a digest, using a hash function. The digest is encrypted and sent with the original message as a signature. An opponent, knowing the original and the digest tries to tamper with the original by changing some parts of the fragments at random keeping the signature unchanged, Davies and Price (1980) and Mueller-Schloer (1983). The urns are possible hashed

Table 3(a). Birthday problem in two groups of n_1 boys and n_2 girls.

Probability of coincidence of boys' and girls' birthdays											
$n_1 \backslash n_2$	5	10	15	20	25	30	35	40	45	50	55
5	0.066										
10	0.128	0.240									
15	0.186	0.337	0.460								
20	0.240	0.422	0.561	0.666							
25	0.290	0.496	0.642	0.746	0.820						
30	0.337	0.561	0.709	0.807	0.872	0.915					
35	0.381	0.617	0.763	0.853	0.909	0.944	0.965				
40	0.422	0.666	0.807	0.888	0.935	0.963	0.978	0.987			
45	0.460	0.709	0.843	0.915	0.954	0.975	0.987	0.993	0.996		
50	0.496	0.746	0.872	0.935	0.967	0.983	0.992	0.996	0.998	0.999	
55	0.530	0.779	0.896	0.951	0.977	0.989	0.995	0.998	0.999	0.999	1.000

$n_1 = n_2$	10	11	12	13	14	15	16	17	18	19	20
Probability	0.240	0.282	0.326	0.371	0.416	0.460	0.504	0.547	0.589	0.628	0.666

Table 3(b). Classical birthday problem.

Probability of coincidence of birthdays in groups of n persons											
n	10	20	21	22	23	24	25	30	40	50	60
Probability	0.117	0.411	0.444	0.476	0.507	0.538	0.569	0.706	0.891	0.970	0.994

fragments of texts, and the balls are randomly modified and hashed texts. The two types represent forward and backward compression starting from the ends to meet in the middle. The modeling is discussed in an accompanying note, Nishimura and Sibuya (1987).

Occupancy problems have been extensively studied. See, for example, Johnson and Kotz (1977), Kolchin *et al.* (1978) and Fang (1985), among others. However, the generalization of the above-mentioned direction has not been thoroughly studied. Popova (1968) obtained limit distributions of the joint distribution of R_1, R_2 and R_3 in the more general case of nonuniform throw-in probabilities.

The Stirling numbers of the second kind which are denoted by $\left\{ {n \atop m} \right\}$, $1 \le m \le n$, are defined by the polynomial identity

$$x^n = \sum_{m=1}^{n} \left\{ {n \atop m} \right\} x^{(m)}, \quad \text{where} \quad x^{(m)} = x(x-1)\cdots(x-m+1) .$$

They are also expressed by using the forward difference operator Δ as

$$\left\{ {n \atop m} \right\} = \Delta^m 0^n / m! \,, \tag{1.1}$$

and satisfy the recurrence relation

$$\left\{ {n \atop m} \right\} = m \left\{ {n-1 \atop m} \right\} + \left\{ {n-1 \atop m-1} \right\}, \tag{1.2}$$

$\left\{ {0 \atop 0} \right\}$ being one by convention. See, for example, Jordan (1950), Riordan (1958), Johnson and Kotz (1977) and Knuth (1967–1981). The notation of the Stirling numbers differs in the literature. Here the notation of Knuth, which emphasizes the similarity to binomial coefficients, is followed. Univariate discrete distributions including the Stirling numbers of the first and the second kinds have been surveyed by Sibuya (1986).

2. Joint distributions of the numbers of urns and balls

In Table 1 the marginal distributions of T_1, T_2 and $S+R_1+R_2=T_1+T_2-S$ follow the classical occupancy distributions;

$$\Pr[T_i = t] = \left\{ {n_i \atop t} \right\} \frac{m^{(t)}}{m^{n_i}} \,, \quad 1 \le t \le \min(m, n_i) \,, \quad i = 1, 2 \,; \tag{2.1}$$

$$\Pr[S + R_1 + R_2 = u] = \left\{ {n_1 + n_2 \atop u} \right\} \frac{m^{(u)}}{m^{n_1+n_2}} \,, \tag{2.2}$$

$$1 \le u \le \min(m, n_1 + n_2) \,.$$

Under the condition that the marginals m, T_1 and T_2, and therefore $m-T_1$ and $m-T_2$ are given, the entries of the 2×2 table follow the hypergeometric distributions. For example,

$$\begin{aligned} \Pr[S = s \mid T_1 = t_1, T_2 = t_2] \\ = \binom{t_1}{s}\binom{m-t_1}{t_2-s} \Big/ \binom{m}{t_2} = \binom{t_2}{s}\binom{m-t_2}{t_1-s} \Big/ \binom{m}{t_1} \,, \end{aligned} \tag{2.3}$$

$$\max(0, t_1 + t_2 - m) \le s \le \min(t_1, t_2) \,.$$

There are several models leading conditionally to a 2×2 table, with different joint distributions, and the above-mentioned is just another type of model.

Combining (2.1) and (2.3) and assigning $t_i = r_i + s$, $i = 1, 2$, the joint distribution of (S, R_1, R_2), which can be called a "2×2 occupancy distribution", is obtained as follows:

$$\text{(2.4)}\qquad \Pr[(S, R_1, R_2) = (s, r_1, r_2); m, n_1, n_2]$$
$$= \frac{1}{m^{n_1+n_2}} \left\{ {n_1 \atop r_1+s} \right\} \left\{ {n_2 \atop r_2+s} \right\} \frac{m!(r_1+s)!(r_2+s)!}{s!r_1!r_2!(m-r_1-r_2-s)!},$$
$$0 \le s, r_1, r_2; \quad r_1+r_2+s \le m; \quad 1 \le r_1+s \le n_1; \quad 1 \le r_2+s \le n_2 .$$

Suppose n_i balls randomly occupy t_i urns, and first separately consider the cases $i=1$ and $i=2$. Under this condition, select s urns at random from the occupied t_i ones. It is shown that these s urns contain Y_i balls with the probability

$$\text{(2.5)}\qquad \Pr[Y_i = y_i | T_i = t_i = r_i + s; n_i, s]$$
$$= \binom{n_i}{y_i} \left\{ {y_i \atop s} \right\} \left\{ {n_i - y_i \atop r_i} \right\} \Big/ \binom{t_i}{s} \left\{ {n_i \atop t_i} \right\} .$$

In Tables 1 and 2 the events $Y_i=y_i$, $i=1, 2$, occur for a set of s urns common to both sets of t_1 and t_2 urns. Multiplying (2.4) and (2.5) with $i=1$ and $i=2$, we obtain the following joint probability function of the five free random variables in Tables 1 and 2:

$$\text{(2.6)}\qquad \Pr[(S, R_1, R_2, Y_1, Y_2) = (s, r_1, r_2, y_1, y_2); m, n_1, n_2]$$
$$= \frac{1}{m^{n_1+n_2}} \binom{n_1}{y_1} \binom{n_2}{y_2} \left\{ {y_1 \atop s} \right\} \left\{ {n_1 - y_1 \atop r_1} \right\} \left\{ {y_2 \atop s} \right\} \left\{ {n_2 - y_2 \atop r_2} \right\} \frac{m!\, s!}{(m-r_1-r_2-s)!} .$$

For confirmation (2.6) is obtained by another method. Under the condition $T_1=t_1$, Y_2 follows the binomial distribution $\mathrm{Bn}(n_2, t_1/m)$. Since T_1 follows (2.1), the joint probability function of T_1 and Y_2 is

$$\text{(2.7)}\qquad \Pr[(T_1, Y_2) = (t_1, y_2); m, n_1, n_2]$$
$$= \frac{1}{m^{n_1+n_2}} \binom{n_2}{y_2} \left\{ {n_1 \atop t_1} \right\} m^{(t_1)} t_1^{y_2} (m-t_1)^{n_2-y_2} .$$

Under the condition $(T_1, Y_2)=(t_1, y_2)$, S and R_2 are independent and follow (2.1) with modified parameters:

$$\text{(2.8)}\qquad \Pr[(S, R_2) = (s, r_2) | (T_1, Y_2) = (t_1, y_2); m, n_1, n_2]$$
$$= \left\{ {y_2 \atop s} \right\} \frac{t_1^{(s)}}{t_1^{y_2}} \left\{ {n_2 - y_2 \atop r_2} \right\} \frac{(m-t_1)^{(r_2)}}{(m-t_1)^{n_2-y_2}} .$$

The distribution of Y_1 given $T_1=r_1+s=t_1$ is (2.5). Multiplying (2.8), (2.7) and (2.5) with $i=1$, we again obtain (2.6).

The joint probability function (2.6) can be rewritten using the difference operator expression (1.1) of the Stirling numbers of the second kind. Further its probability generating function is written as follows:

$$
\begin{aligned}
(2.9)\qquad & g(\sigma, \rho_1, \rho_2, \eta_1, \eta_2; m, n_1, n_2) \\
&= \Sigma \Pr[(S, R_1, R_2, Y_1, Y_2) \\
&\qquad = (s, r_1, r_2, y_1, y_2); m, n_1, n_2]\, \sigma^s \rho_1^{r_1} \rho_2^{r_2} \eta_1^{y_1} \eta_2^{y_2} \\
&= [m^{-n_1-n_2}(1 + \rho_1 \Delta_v + \rho_2 \Delta_z + \sigma \Delta_u \Delta_w)^m \\
&\quad \cdot (u\eta_1 + v)^{n_1}(w\eta_2 + z)^{n_2}]_{u=v=w=z=0} .
\end{aligned}
$$

The extended exponential generating function of the family of joint probability functions (2.6) is defined by

$$
\begin{aligned}
&\phi(\sigma, \rho_1, \rho_2, \eta_1, \eta_2; \nu_1, \nu_2; m) \\
&\quad = \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} \frac{(m\nu_1)^{n_1}}{n_1!} \frac{(m\nu_2)^{n_2}}{n_2!} \sum_s \sum_{r_1} \sum_{r_2} \sum_{y_1} \sum_{y_2} \sigma^s \rho_1^{r_1} \rho_2^{r_2} \eta_1^{y_1} \eta_2^{y_2} \\
&\qquad \cdot \Pr[(S, R_1, R_2, Y_1, Y_2) = (s, r_1, r_2, y_1, y_2)] ,
\end{aligned}
$$

Johnson and Kotz (1977, p.63). Since the exponential generating function of $\left\{ {n \atop m} \right\}$, $n=m, m+1,\ldots$ is

$$
\sum_n \left\{ {n \atop m} \right\} \frac{z^n}{n!} = \frac{(e^z - 1)^m}{m!} , \qquad m = 1, 2, \ldots ,
$$

it is shown that

$$
\begin{aligned}
(2.10)\qquad & \phi(\sigma, \rho_1, \rho_2, \eta_1, \eta_2; \nu_1, \nu_2; m) \\
&\qquad = \{1 + \rho_1(e^{\nu_1} - 1) + \rho_2(e^{\nu_2} - 1) + \sigma(e^{\nu_1\eta_1} - 1)(e^{\nu_2\eta_2} - 1)\}^m .
\end{aligned}
$$

This expression can be also obtained by (2.9).

3. Marginal distributions

Various marginal probability functions and probability generating functions can be obtained from (2.6) or (2.9). Typical functions obtained are summarized in Table 4, and others can be obtained from those in the table. Some have been previously shown in Section 2.

The corresponding extended generating functions are obtained from (2.10) by replacing some of σ, ρ_1, ρ_2, η_1 and η_2 with one. For example, the extended generating function corresponding to (2.4) is

$$
\begin{aligned}
(3.1)\qquad & \phi(\sigma, \rho_1, \rho_2; \nu_1, \nu_2; m) \\
&\qquad = \{1 + \rho_1(e^{\nu_1} - 1) + \rho_2(e^{\nu_2} - 1) + \sigma(e^{\nu_1} - 1)(e^{\nu_2} - 1)\}^m .
\end{aligned}
$$

Popova (1968) obtained a more general expression for the probabilities of (R_1, R_2, R_3) in Table 1, for the case of non-uniform throw-in probabilities. She showed that the random vector $(S, n_1 - R_1, n_2 - R_2)$ is asymptotically

independent and Poisson if $0<c_1\le n_i^2/m\le c_2<\infty$, $i=1, 2$, and that the random vector is asymptotically normal if $0<c_1\le n_i/m\le c_2<\infty$, $i=1, 2$.

The factorial moments are obtained from the probability generating function. For example,

$$g(\sigma) = \frac{1}{m^{n_1+n_2}}[\{(1+\Delta_v)(1+\Delta_z)+(\sigma-1)\Delta_v\Delta_z\}^m v^{n_1} z^{n_2}]_{v=z=0}$$
$$= \frac{1}{m^{n_1+n_2}} \sum_l \frac{(\sigma-1)^l}{l!} m^{(l)}[\nabla_v^l (m+v)^{n_1} \nabla_z^l (m+z)^{n_2}]_{v=z=0},$$

where ∇ denotes backward difference. Therefore, the factorial moments of S are

$$\mathrm{E}[S^{(l)}] = \frac{1}{m^{n_1+n_2}} m^{(l)} \nabla^l m^{n_1} \nabla^l m^{n_2}. \tag{3.2}$$

Another example:

$$g(\eta_2) = \frac{1}{m^{n_1+n_2}}\left[\sum_l \frac{(\eta_2-1)^l}{l!} n_2^{(l)}(1+E_z^{-1}E_w\Delta_v)^m w^l \cdot (m+w+z)^{n_2-l} v^{n_1}\right]_{w=v=z=0}$$
$$= \frac{1}{m^{n_1+n_2}}\left[\sum_l \frac{(\eta_2-1)^l}{l!} n_2^{(l)}(m+w+z)^{n_2-l} \cdot \sum_j \left\{{l \atop j}\right\}(1+E_w\Delta_v)^m w^{(j)} v^{n_1}\right]_{w=v=z=0},$$

where E denotes the shift operator: $\Delta = E-1$ and $\nabla = 1-E^{-1}$. Therefore,

$$\mathrm{E}[Y_2^{(l)}] = \frac{n_2^{(l)}}{m^{n_1+l}}\left[\sum_j \left\{{l \atop j}\right\} \sum_k \frac{m^{(k)}}{k!} k^{(j)} \Delta_v^k v^{n_1}\right]_{v=0} \tag{3.3}$$
$$= \frac{n_2^{(l)}}{m^l} \sum_j \left\{{l \atop j}\right\} \frac{1}{m^{n_1}} \nabla^j m^{n_1},$$

which is obtained more easily from $\mathrm{E}[Y_2^{(l)} \mid T_1=t_1]$.

Of particular interest is the probability of the event $S=0$, which is equivalent to $Y_1=0$ or $Y_2=0$:

$$\Pr[S=0; m, n_1, n_2] = \frac{1}{m^{n_1+n_2}} \sum_{t_1} \sum_{t_2} \left\{{n_1 \atop t_1}\right\}\left\{{n_2 \atop t_2}\right\} m^{(t_1+t_2)} \tag{3.4}$$
$$= \frac{1}{m^{n_1+n_2}} \sum_{v=2}^{m} m^{(v)} \sum_{t_1+t_2=v} \left\{{n_1 \atop t_1}\right\}\left\{{n_2 \atop t_2}\right\}.$$

Table 4. Probability functions and their generating functions in occupancy with two types of balls.

Probability functions	
$p(s, r_1, r_2, y_1, y_2)$	$= \frac{1}{m^{n_1+n_2}} \binom{n_1}{y_1} \left\{ {y_1 \atop s} \right\} \left\{ {n_1 - y_1 \atop r_1} \right\} \binom{n_2}{y_2} \left\{ {y_2 \atop s} \right\} \left\{ {n_2 - y_2 \atop r_2} \right\} \frac{m!\, s!}{(m - r_1 - r_2 - s)!}$
$p(s, r_1, t_2, y_1)$	$= \frac{1}{m^{n_1+n_2}} \binom{n_1}{y_1} \left\{ {y_1 \atop s} \right\} \left\{ {n_1 - y_1 \atop r_1} \right\} \left\{ {n_2 \atop t_2} \right\} \left({t_2 \atop s} \right) \frac{m!\, s!}{(m - r_1 - t_2)!}$
$p(s, t_1, y_1, y_2)$	$= \frac{1}{m^{n_1+n_2}} \binom{n_1}{y_1} \left\{ {y_1 \atop s} \right\} \left\{ {n_1 - y_1 \atop t_1 - s} \right\} \binom{n_2}{y_2} \left\{ {y_2 \atop s} \right\} m^{(t_1)} (m - t_1)^{n_2 - y_2} s!$
$p(s, t_1, t_2)$	$= \frac{1}{m^{n_1+n_2}} \left\{ {n_1 \atop t_1} \right\} \left({t_1 \atop s} \right) \left\{ {n_2 \atop t_2} \right\} \left({t_2 \atop s} \right) \frac{m!\, s!}{(m - t_1 - t_2 + s)!}$
$p(s, r_1, y_2)$	$= \frac{1}{m^{n_1+n_2}} \left\{ {n_1 \atop r_1 + s} \right\} \left({r_1 + s \atop s} \right) \binom{n_2}{y_2} \left\{ {y_2 \atop s} \right\} m^{(r_1+s)} (m - r_1 - s)^{n_2 - y_2} s!$
$p(t_1, r_2, y_1)$	$= \frac{1}{m^{n_1+n_2}} \binom{n_1}{y_1} \sum_s \left\{ {y_1 \atop s} \right\} \left\{ {n_1 - y_1 \atop t_1 - s} \right\} \left\{ {n_2 \atop r_2 + s} \right\} \left({r_2 + s \atop s} \right) \frac{m!\, s!}{(m - t_1 - r_2)!}$
$p(t_1, r_2, y_2)$	$= \frac{1}{m^{n_1+n_2}} \left\{ {n_1 \atop t_1} \right\} \binom{n_2}{y_2} \left\{ {n_2 - y_2 \atop r_2} \right\} \frac{m!}{(m - t_1 - r_2)!}\, t_1^{y_2}$
$p(r_1, y_1, y_2)$	$= \frac{1}{m^{n_1+n_2}} \binom{n_1}{y_1} \left\{ {n_1 - y_1 \atop r_1} \right\} \binom{n_2}{y_2} \sum_s \left\{ {y_1 \atop s} \right\} \left\{ {y_2 \atop s} \right\} m^{(r_1+s)} (m - r_1 - s)^{n_2 - y_2} s!$
$p(r_1, r_2)$	$= \frac{1}{m^{n_1+n_2}} \sum_s \left\{ {n_1 \atop r_1 + s} \right\} \left({r_1 + s \atop s} \right) \left\{ {n_2 \atop r_2 + s} \right\} \left({r_2 + s \atop s} \right) \frac{m!\, s!}{(m - r_1 - r_2 - s)!}$
$p(t_1, t_2)$	$= \frac{1}{m^{n_1+n_2}} \left\{ {n_1 \atop t_1} \right\} \left\{ {n_2 \atop t_2} \right\} m^{(t_1)} m^{(t_2)}$
$p(r_1, y_1)$	$= \frac{1}{m^{n_1+n_2}} \binom{n_1}{y_1} \left\{ {n_1 - y_1 \atop r_1} \right\} \sum_{t_2} \left\{ {n_2 \atop t_2} \right\} \frac{m!}{(m - r_1 - t_2)!}\, t_2^{y_1}$
$p(t_1, y_2)$	$= \frac{1}{m^{n_1+n_2}} \left\{ {n_1 \atop t_1} \right\} \binom{n_2}{y_2} m^{(t_1)} (m - t_1)^{n_2 - y_2} t_1^{y_2}$
$p(y_1, y_2)$	$= \frac{1}{m^{n_1+n_2}} \binom{n_1}{y_1} \binom{n_2}{y_2} \sum_{r_1} \sum_s \left\{ {y_1 \atop s} \right\} \left\{ {y_2 \atop s} \right\} \left\{ {n_1 - y_1 \atop r_1} \right\} m^{(r_1+s)} (m - r_1 - s)^{n_2 - y_2} s!$
$p(s)$	$= \frac{1}{m^{n_1+n_2}} \sum_{t_1} \sum_{t_2} \left\{ {n_1 \atop t_1} \right\} \left({t_1 \atop s} \right) \left\{ {n_2 \atop t_2} \right\} \left({t_2 \atop s} \right) \frac{m!\, s!}{(m - t_1 - t_2 + s)!}$
$p(t_1)$	$= \frac{1}{m^{n_1}} \left\{ {n_1 \atop t_1} \right\} m^{(t_1)}$
$p(y_1)$	$= \frac{1}{m^{n_1+n_2}} \binom{n_1}{y_1} \sum_{t_2} \left\{ {n_2 \atop t_2} \right\} m^{(t_2)} (m - t_2)^{n_1 - y_1} t_2^{y_1}$

Joint and marginal probability functions of the random variables S, R_1, R_2, Y_1 and Y_2 in Tables 1 and 2, and their generating functions. The first pgf is defined by (2.9).

(3.5)
$$\begin{aligned} \Pr[Y_2 = 0\,;\, m, n_1, n_2] &= \mathrm{E}^{T_1}[(1 - T_1/m)^{n_2}] \\ &= \frac{1}{m^{n_1}} \sum_t \left\{ {n_1 \atop t} \right\} m^{(t)} \left(1 - \frac{t}{m}\right)^{n_2} \\ &= \frac{1}{m^{n_1}} \sum_t \left\{ {n_2 \atop t} \right\} m^{(t)} \left(1 - \frac{t}{m}\right)^{n_1} . \end{aligned}$$

To check the equality of the last two expressions and (3.4), develop $(m-t)^{n_2}$ or $(m-t)^{n_1}$ using the definition of the Stirling numbers.

Table 4. (continued).

Corresponding pgf's expressed by difference operators
$g(\sigma, \rho_1, \rho_2, \eta_1, \eta_2) = \frac{1}{m^{n_1+n_2}}[(1+\rho_1\Delta_v+\rho_2\Delta_z+\sigma\Delta_u\Delta_w)^m(\eta_1 u+v)^{n_1}(\eta_2 w+z)^{n_2}]_{u=v=w=z=0}$
$g(\sigma, \rho_1, \tau_2, \eta_1) = \frac{1}{m^{n_1+n_2}}[(1+\rho_1\Delta_v+(1+\sigma\Delta_u)\tau_2\Delta_z)^m(\eta_1 u+v)^{n_1}z^{n_2}]_{u=v=z=0}$
$g(\sigma, \tau_1, \eta_1, \eta_2) = \frac{1}{m^{n_1+n_2}}[(E_z+\tau_1\Delta_v+\tau_1\sigma\Delta_u\Delta_w)^m(\eta_1 u+v)^{n_1}(\eta_2 w+z)^{n_2}]_{u=v=w=z=0}$
$g(\sigma, \tau_1, \tau_2) = \frac{1}{m^{n_1+n_2}}[(1+\tau_1\Delta_v+\tau_2\Delta_z+\tau_1\tau_2\sigma\Delta_v\Delta_z)^m v^{n_1}z^{n_2}]_{v=z=0}$
$g(\sigma, \rho_1, \eta_2) = \frac{1}{m^{n_1+n_2}}[(E_z+\rho_1\Delta_v+\sigma\Delta_v\Delta_w)^m v^{n_1}(\eta_2 w+z)^{n_2}]_{v=w=z=0}$
$g(\tau_1, \rho_2, \eta_1) = \frac{1}{m^{n_1+n_2}}[(1+\tau_1\Delta_v+\rho_2\Delta_z+\tau_1\Delta_u\Delta_z)^m(\eta_1 u+v)^{n_1}z^{n_2}]_{u=v=z=0}$
$g(\tau_1, \rho_2, \eta_2) = \frac{1}{m^{n_1+n_2}}[(1+\tau_1 E_w\Delta_v+\rho_2\Delta_z)^m v^{n_1}(\eta_2 w+z)^{n_2}]_{v=w=z=0}$
$g(\rho_1, \eta_1, \eta_2) = \frac{1}{m^{n_1+n_2}}[(E_z+\rho_1\Delta_v+\Delta_u\Delta_w)^m(\eta_1 u+v)^{n_1}(\eta_2 w+z)^{n_2}]_{u=v=w=z=0}$
$g(\rho_1, \rho_2) = \frac{1}{m^{n_1+n_2}}[(1+\rho_1\Delta_v+\rho_2\Delta_z+\Delta_v\Delta_z)^m v^{n_1}z^{n_2}]_{v=z=0}$
$g(\tau_1, \tau_2) = \frac{1}{m^{n_1}}[(1+\tau_1\Delta_v)^m v^{n_1}]_{v=0}\frac{1}{m^{n_2}}[(1+\tau_2\Delta_z)^m z^{n_2}]_{z=0}$
$g(\rho_1, \eta_1) = \frac{1}{m^{n_1+n_2}}[(E_z+\rho_1\Delta_v+\Delta_u\Delta_z)^m(\eta_1 u+v)^{n_1}z^{n_2}]_{u=v=z=0}$
$g(\tau_1, \eta_2) = \frac{1}{m^{n_1+n_2}}[(E_z+\tau_1 E_w\Delta_v)^m v^{n_1}(\eta_2 w+z)^{n_2}]_{v=w=z=0}$
$g(\eta_1, \eta_2) = \frac{1}{m^{n_1+n_2}}[(1+\Delta_v+\Delta_z+\Delta_u\Delta_w)^m(\eta_1 u+v)^{n_1}(\eta_2 w+z)^{n_2}]_{u=v=w=z=0}$
$g(\sigma) = \frac{1}{m^{n_1+n_2}}[(1+\Delta_v+\Delta_z+\sigma\Delta_v\Delta_z)^m v^{n_1}z^{n_2}]_{v=z=0}$
$g(\tau_1) = \frac{1}{m^{n_1}}[(1+\tau_1\Delta_v)^m v^{n_1}]_{v=0}$
$g(\eta_1) = \frac{1}{m^{n_1+n_2}}[(E_v+E_u\Delta_z)^m(\eta_1 u+v)^{n_1}z^{n_2}]_{u=v=z=0}$

Δ denotes the forward difference operator, and E the shift operator: $\Delta = E - 1$.

It is intuitively true that n_1+n_2 being fixed $\Pr[S>0]$ will be maximized when n_1 and n_2 are equal. The following proposition confirms this. The proof is given in Appendix.

PROPOSITION 3.1. *The probability* $\Pr[S=0; m, n_1, n_2]$ *of* (3.4) *or* (3.5) *with* m *and* n_1+n_2 *fixed decreases when* $|n_1-n_2|$ *decreases.*

4. Bounds of Pr[S=0] and the asymptotic distribution of the waiting time

4.1 *Lower bounds*

Since $h(t):=(1-t/m)^{n_2}$ in (3.5) is a convex function of t, a simple and good lower bound is

$$\text{(4.1)} \qquad \begin{aligned} \mathrm{E}^{T_1}[h(T_1)] &\geq h(\mathrm{E}[T_1]) \\ &= (1 - 1/m)^{n_1 n_2} =: L_1(m, n_1, n_2) . \end{aligned}$$

Further, since $h(t)$ is bounded by the tangential parabola, a stronger version of (4.1) is

$$\text{(4.2)} \qquad \begin{aligned} & h(\mathrm{E}[T_1]) + \frac{1}{2} h''(\mathrm{E}[T_1])\ \mathrm{Var}[T_1] \\ &= L_1(m, n_1, n_2) \cdot \left[1 + \frac{n_2(n_2 - 1)}{2m}\left(1 - \frac{1}{m}\right)^{n_1}\right. \\ &\quad \left. \cdot \left\{1 - \left(1 - \frac{1}{m-1}\right)^{n_1} + m\left(\left(1 - \frac{1}{m-1}\right)^{n_1} - \left(1 - \frac{1}{m}\right)^{n_1}\right)\right\}\right] \\ &=: L_0(m, n_1, n_2) . \end{aligned}$$

This is a complicated but excellent bound and $\max(L_0(m, n_1, n_2), L_0(m, n_2, n_1))$ further improves the bound.

If $n_1 n_2$ is larger,

$$\text{(4.3)} \qquad \exp(-n_1 n_2/m)$$

is also a better bound than L_1, but (4.3) exceeds Pr[S=0] for smaller values of $n_1 n_2$. The expression (4.3) is a lower bound if $n_1 n_2$ is larger than m. It was only possible to check the range numerically. A modification of (4.3)

$$\text{(4.4)} \qquad \exp\left\{-\frac{n_1 n_2}{m}\left(1 + \frac{n_1 + n_2}{4m}\right)\right\}$$

is smaller than (4.3), but is shown numerically to be a lower bound in wider range $m>2$ and $n_1+n_2>2$.

4.2 *Upper bounds*

It is difficult to obtain a good and simple upper bound. It is known that the number of collisions is asymptotically Poisson, and this fact suggests a way.

PROPOSITION 4.1. *The Poisson distribution with mean* $\lambda=n_1^2/2(m-n_1+1)$

is stochastically larger than the distribution of the number of collisions $C=n_1-T_1$,

$$\Pr[C=c]=\left\{{n_1 \atop n_1-c}\right\}\frac{m^{(n_1-c)}}{m^{n_1}},$$

(*cf.* (2.1)).

The proof is given in Appendix.

Using this proposition, take the expectation of $h(T_1)=h(n_1-C)\leq\exp(-n_2\cdot(n_1-C)/m)$ and replace C with the Poisson variable to get

$$\exp\left\{-\frac{n_1n_2}{m}+\lambda\left(\exp\left(\frac{n_2}{m}\right)-1\right)\right\}=:U_1(m,n_1,n_2)\,. \tag{4.5}$$

This expression is asymmetrical in (n_1, n_2), and the minimum of $U_1(m, n_1, n_2)$ and $U_1(m, n_2, n_1)$ can be chosen. Notice that $h(t)<1$ and evaluate

$$\mathrm{E}[h(n_1-C)]\leq h(n_1)\Pr[C=0]+\Pr[C>0]\,,$$

replacing C in the last or both probabilities with the Poisson variable. Then, a simpler bound is

$$\begin{aligned}\frac{m^{(n_1)}}{m^{n_1}}\left(1-\frac{n_1}{m}\right)^{n_2}+1-e^{-\lambda}&=:U_2(m,n_1,n_2)\\&\leq\left(1-\frac{n_1}{m}\right)^{n_2}e^{-\lambda}+1-e^{-\lambda}\,.\end{aligned} \tag{4.6}$$

Rough upper bounds are also obtained by using the Chebyshev-type inequality on the distributions of S or Y_i.

4.3 *Asymptotic distribution of waiting time*

Suppose that white and red balls are thrown one by one according to some rule of choice fixed in advance. Let n_{1j} and n_{2j} be the numbers of white and red balls thrown up to the j-th step. The sequence $\{(n_{1j}, n_{2j})\}_{j=1}^{\infty}$, $n_{1j}+n_{2j}=j$, represents the rule of choice. We assume that after some finite steps $n_{1j}n_{2j}>0$. Let J denote the step number where the first collision between the two colors occurs, that is, J is the waiting time of collision, and put $(N_1, N_2)=(n_{1J}, n_{2J})$. The event $J>j$ is equivalent to the event $S=0$ at (n_{1j}, n_{2j}).

THEOREM 4.1. *Let* $\{(n_{1j}, n_{2j})\}_{j=1}^{\infty}$ *be a rule of choice of white and red balls, and let* $(N_1, N_2)=(n_{1J}, n_{2J})$ *be the numbers of white and red balls when the first collision of two colors occurs. For any positive* $M>0$, *as* $m\to\infty$

$$\Pr[N_1N_2/m \le w] \to 1 - e^{-w}, \quad \textit{for} \quad 0 < w < M.$$

PROOF. Since $n_{1j}n_{2j}$ is strictly increasing in j if $n_{1j}n_{2j}>0$, the event $N_1N_2/m>w$ is equivalent to the event $S=0$ for all (n_{1j}, n_{2j}) such that $n_{1j}n_{2j}\le mw$. Now, for a given w put $\omega(j)=n_{1j}n_{2j}/w$, for which

$$\Delta\omega(j)/\omega(j) = 1/n_{1j} \quad \text{or} \quad 1/n_{2j}.$$

Let $j=j(m)$ increase to infinity as $m\to\infty$ such that

$$\omega(j) \le m < \omega(j+1).$$

Then, provided that $n_{1j}, n_{2j}\to\infty$ $(j\to\infty)$,

$$\left|\frac{n_{1j}n_{2j}}{m} - w\right| \le \Delta\omega(j)/\omega(j) \to 0 \ (m \to \infty).$$

Otherwise assume, for example, $n_{2j}\le c$, $0<c<\infty$, then there exists finite k such that $n_{2k}=c$, and $\Delta\omega(j)/\omega(j)=1/n_{1j}$ for $j>k$. Thus, for any $\{(n_{1j}, n_{2j})\}_{j=1}^{\infty}$ such that $n_{1j}n_{2j}>0$ except for a smaller j,

$$n_{1j}n_{2j}/m \to w, \quad m \to \infty,$$

if j is increased as mentioned above.

The probability

$$\Pr[N_1N_2/m > w] = \Pr[S = 0; m, n_{1j}, n_{2j}],$$

where j satisfies $\omega(j)\le m<\omega(j+1)$, is bounded by $L_1(m, n_{1j}, n_{2j})$ and $\min(U_1(m, n_{1j}, n_{2j}), U_1(m, n_{2j}, n_{1j}))$. Since (4.5) is rewritten as

$$U_1(m, n_1, n_2) = \exp\left\{-\frac{n_1n_2}{m}\left(1 - \frac{n_1}{2(m-n_1+1)}\left(1 + \frac{n_2}{2m} + \cdots\right)\right)\right\},$$

$$\lim_{m\to\infty} L_1(m, n_{1j}, n_{2j}) = \lim_{m\to\infty} \min(U_1(m, n_{1j}, n_{2j}), U_1(m, n_{2j}, n_{ij}))$$

$$= \lim_{m\to\infty} \exp(-n_{1j}n_{2j}/m) = \exp(-w).$$

In the case where white and red balls are thrown alternately, the distribution of $N_1/\sqrt{m}$ or $N_2/\sqrt{m}$ is asymptotically the Rayleigh distribution with the probability density

$$2w \exp(-w^2), \quad 0 < w < \infty .$$

Refer to Hirano (1986) for the Rayleigh distribution.

Acknowledgements

The authors wish to thank the referees for their suggestions which substantially improved the paper.

Appendix

Proposition 3.1 is straightforward from Lemma A.1.

LEMMA A.1. *The convolution of the Stirling numbers of the second kind*

$$\sum_t \left\{ {n_1 \atop t} \right\} \left\{ {n_2 \atop m-t} \right\}, \quad m = 2, 3, \ldots ,$$

with $n_1+n_2=n$ fixed, decreases when $|n_1-n_2|$ decreases. That is,

$$\sum_t \left\{ {k \atop t} \right\} \left\{ {n-k \atop m-t} \right\} < \sum_t \left\{ {j \atop t} \right\} \left\{ {n-j \atop m-t} \right\}, \tag{A.1}$$

if $|n-2k|<|n-2j|$, provided that the right-hand side is positive.

PROOF. It is sufficient to prove (A.1) for the case $j=k-1<k\leq n-k<n-k+1=n-j$. Apply the recurrence formula (1.2) to $\left\{ {k \atop t} \right\}$ of the left-hand side and $\left\{ {n-k+1 \atop m-t} \right\}$ of the right-hand side. Then (A.1) is equivalent to

$$\sum_t t \left\{ {k-1 \atop t} \right\} \left\{ {n-k \atop m-t} \right\} < \sum_t t \left\{ {k-1 \atop m-t} \right\} \left\{ {n-k \atop t} \right\} . \tag{A.2}$$

Now, it is shown from (1.2) that the Stirling numbers of the second kind are TP_2 (Totally Positive 2), namely,

$$\left\{ {n_1 \atop m_2} \right\} \left\{ {n_2 \atop m_1} \right\} \leq \left\{ {n_1 \atop m_1} \right\} \left\{ {n_2 \atop m_2} \right\}, \quad \text{if} \quad n_1 < n_2 \quad \text{and} \quad m_1 < m_2 ,$$

and the strict inequality holds unless the right-hand side is zero. Therefore, if $t<m-t$,

$$t\left\{\begin{matrix} k-1 \\ t \end{matrix}\right\}\left\{\begin{matrix} n-k \\ m-t \end{matrix}\right\} + (m-t)\left\{\begin{matrix} k-1 \\ m-t \end{matrix}\right\}\left\{\begin{matrix} n-k \\ t \end{matrix}\right\}$$
$$< (m-t)\left\{\begin{matrix} k-1 \\ t \end{matrix}\right\}\left\{\begin{matrix} n-k \\ m-t \end{matrix}\right\} + t\left\{\begin{matrix} k-1 \\ m-t \end{matrix}\right\}\left\{\begin{matrix} n-k \\ t \end{matrix}\right\},$$

and (A.2) is proved.

For proving Proposition 4.1, another lemma is needed.

LEMMA A.2. *For any positive integer* $n \geq 3$,

$$(m+1)\left\{\begin{matrix} n \\ n-m-1 \end{matrix}\right\} \Big/ \left\{\begin{matrix} n \\ n-m \end{matrix}\right\}, \quad m = 0, 1, 2, \ldots, n-2,$$

is a strictly decreasing sequence.

PROOF. Proceed induction on n. If $n=3$, the sequence is $3, 2/3$. To advance the induction step from n to $n+1$, compute

$$(m+1)\left\{\begin{matrix} n+1 \\ n-m \end{matrix}\right\}^2 - (m+2)\left\{\begin{matrix} n+1 \\ n-m+1 \end{matrix}\right\}\left\{\begin{matrix} n+1 \\ n-m-1 \end{matrix}\right\}$$
$$= (n-m)^2\left[(m+1)\left\{\begin{matrix} n \\ n-m \end{matrix}\right\}^2 - (m+2)\left\{\begin{matrix} n \\ n-m+1 \end{matrix}\right\}\left\{\begin{matrix} n \\ n-m-1 \end{matrix}\right\}\right]$$
$$+ (n-m+1)\left[m\left\{\begin{matrix} n \\ n-m \end{matrix}\right\}\left\{\begin{matrix} n \\ n-m-1 \end{matrix}\right\}\right.$$
$$\left. - (m+2)\left\{\begin{matrix} n \\ n-m+1 \end{matrix}\right\}\left\{\begin{matrix} n \\ n-m-2 \end{matrix}\right\}\right]$$
$$+ 2\left\{\begin{matrix} n \\ n-m \end{matrix}\right\}\left\{\begin{matrix} n \\ n-m-1 \end{matrix}\right\} + (m+2)\left\{\begin{matrix} n \\ n-m+1 \end{matrix}\right\}\left\{\begin{matrix} n \\ n-m-1 \end{matrix}\right\}$$
$$+ \left[(m+1)\left\{\begin{matrix} n \\ n-m-1 \end{matrix}\right\}^2 - (m+2)\left\{\begin{matrix} n \\ n-m \end{matrix}\right\}\left\{\begin{matrix} n \\ n-m-2 \end{matrix}\right\}\right].$$

All the terms are positive and Lemma A.2 is proved.

To prove Proposition 4.1, put

$$f(x) := \left\{\begin{matrix} n \\ n-x \end{matrix}\right\} \frac{m^{(n-x)}}{m^n},$$

and define the Poisson distribution function

$$g(x) := e^{-\lambda}\lambda^x / x!$$

with

$$\lambda \geq f(1)/f(0) = n(n-1)/2(m-n+1) .$$

Due to Lemma A.2,

$$(x+1)f(x+1)/f(x) = \frac{x+1}{m-n+x+1} \left\{ \begin{matrix} n \\ n-x-1 \end{matrix} \right\} \Big/ \left\{ \begin{matrix} n \\ n-x \end{matrix} \right\}$$

is decreasing in x, and

$$(x+1)\frac{g(x+1)}{g(x)} = \lambda \geq \frac{f(1)}{f(0)} > (x+1)\frac{f(x+1)}{f(x)} .$$

Since $g(x)/f(x)$ is increasing, $g(x)$ is stochastically larger than $f(x)$ and the proof is complete.

REFERENCES

Davies, D. W. and Price, W. L. (1980). The application of digital signatures based on public key cryptosystems, *Proc. 5th Internat. Symp. Comput. Commun.*, IEEE, Oct. 27–30, 1980, 525–530.

Fang, K.-T. (1985). Occupancy problems, *Encyclopedia of Statistical Sciences*, (eds. S. Kotz and N. L. Johnson), **6**, 402–406, Wiley, New York.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed., Wiley, New York.

Hirano, K. (1986). Rayleigh distribution, *Encyclopedia of Statistical Sciences*, (eds. S. Kotz and N. L. Johnson), **7**, 647–649, Wiley, New York.

Johnson, N. L. and Kotz, S. (1977). *Urn Models and Their Applications*, Wiley, New York.

Jordan, C. (Koroly) (1950, 1960). *Calculus of Finite Difference*, Chelsea Publishing Co., New York.

Knuth, D. E. (1967–1981). *The Art of Computer Programming*, Vol. 1–3, Addison-Wesley, Reading, Massachusetts.

Kolchin, V. F., Sevast'yanov, B. A. and Chistyakov, V. H. (1978). *Random Allocation*, (tr. ed. A. V. Balakrishna), V. H. Wistons and Sons, Washington, D.C.

Mueller-Schloer, C. (1983). DES-generated checksums for electronic signatures, *Cryptologia*, **7**, 257–273.

Nishimura, K. and Sibuya, M. (1987). Probability to meet in the middle, KSTS-RR/87-006, Department of Mathematics, Keio University.

Popova, T. Y. (1968). Limit theorems in a model of distribution of particles of two types, *Theory Probab. Appl.*, **13**, 511–516.

Riordan, J. (1958). *An Introduction to Combinatorial Analysis*, Wiley, New York.

Sibuya, M. (1986). Stirling family of probability distributions, *Japan J. Appl. Statist.*, **15**, 131–146 (in Japanese).