Research Report

KSTS/RR-87/002
31 Mar. 1987

Occupancy with two types of balls

by

Kazuo Nishimura
Masaaki Sibuya

Kazuo Nishimura
Masaaki Sibuya

Department of Mathematics
Faculty of Science and Technology
Keio University

Hiyoshi 3-14-1, Kohoku-ku
Yokohama, 223 Japan

Department of Mathematics
Faculty of Science and Technology
Keio University

Hiyoshi 3-14-1, Kohoku-ku, Yokohama, 223 Japan

KSTS/RR-87/002
March 31, 1987

Occupancy with Two Types of Balls

by Kazuo NISHIMURA
and Masaaki SIBUYA

Department of Mathematics, Keio University

January 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. Let N_1 and N_2 denote the number of balls of each type 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, 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 number of balls in each urn is disregarded, and the numbers of urns of these types are represented by a 2 × 2 contingency table (Table 1).

Table 1. Numbers of urns of four types.

white \ red	with	without	sum
with	S	R_1	$T_1 = S + R_1$
without	R_2	$R_3 = m - R_1 - T_2$	$m - T_1$
sum	$T_2 = S + R_2$	$m - T_2$	m

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

The purpose of this report is to first study, in Section 2, the joint, marginal and conditional 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 probability is also expressed, in Section 3, in terms of the number of "collisions" 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 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_1N_2/m is asymptotically exponentially distributed.

This study started from a problem of 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 to attempt forgery keeping the signature unchanged, Davies and Price (1980), and Mueller-Schloer (1983). The urns are possible hashed 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, Sevast'yanov and Chistyanov (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 $\left\{ {n \atop m} \right\}$, $1 \leq m \leq n$, are defined by the polynomial identity

$$(1.1) \qquad x^n = \sum_{m=1}^{n} \left\{ {n \atop m} \right\} x^{(m)} , \quad \text{where} \quad x^{(m)} = x(x-1)\ldots(x-m+1) .$$

They are also expressed by using the difference operator Δ as

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

and satisfy the recurrence relation

$$(1.2) \qquad \left\{ {n \atop m} \right\} = m \left\{ {n-1 \atop m} \right\} + \left\{ {n-1 \atop m-1} \right\} ,$$

$\left\{ {0 \atop 0} \right\}$ 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. Distributions of the number of urns.

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;

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

$$(2.2)\qquad \Pr[S+R_1+R_2 = u] = \left\{ {n_1+n_2 \atop u} \right\} \frac{m^{(u)}}{m^{n_1+n_2}}, \qquad 1 \leq u \leq \min(m, n_1+n_2).$$

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,

$$(2.3)\qquad \Pr[S=s \mid T_1=t_1,\ T_2=t_2] = \binom{t_1}{s}\binom{m-t_1}{t_2-s}\Big/\binom{m}{t_2} = \binom{t_2}{s}\binom{m-t_2}{t_1-s}\Big/\binom{m}{t_1},$$

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

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

Combining (2.1) and (2.3) the joint distribution of (S,R_1,R_2), which can be called a "2 × 2 occupancy distribution", is obtained as follows:

$$\Pr[(S,R_1,R_2) = (s,r_1,r_2);\ m,n_1,n_2]$$

(2.4)
$$= \frac{1}{m^{n_1+n_2}} \left\{ {n_1 \atop r_1+s} \right\} \left\{ {n_2 \atop r_2+s} \right\} \frac{m!\,(r_1+s)!\,(r_2+s)!}{s!\,r_1!\,r_2!\,(m-r_1-r_2-s)!} ,$$

$$0 \leqq s,r_1,r_2;\ r_1+r_2+s \leqq m;\ 1 \leqq r_1+s \leqq n_1;\ 1 \leqq r_2+s \leqq n_2 .$$

From the joint distribution the other marginal and conditional distributions are also obtained:

$$\Pr[S=s;\ m,n_1,n_2]$$

(2.5)
$$= \frac{1}{m^{n_1+n_2}\, s!} \sum_{t_1} \sum_{t_2} \left\{ {n_1 \atop t_1} \right\} \left\{ {n_2 \atop t_2} \right\} t_1^{(s)}\ t_2^{(s)}\ m^{(t_1+t_2-s)} ,$$

$$0 \leqq s \leqq \min\,(n_1,n_2,m) ,$$

where the summations run over $s \leqq t_1 \leqq n_1$, $s \leqq t_2 \leqq n_2$, and $t_1+t_2 \leqq m + s$.

$$\Pr[(S,R_1,R_2) = (s,r_1,r_2) \mid S+R_1+R_2 = u]$$

(2.6)
$$= \frac{\left\{ {n_1 \atop r_1+s} \right\} \left\{ {n_2 \atop r_2+s} \right\}}{\left\{ {n_1+n_2 \atop u} \right\}} \frac{(r_1+s)^{(s)}(r_2+s)^{(s)}}{s!} ,$$

$$0 \leqq s,r_1,r_2\ ;\quad s+r_1 \leqq \min(n_1,u),\ s+r_2 \leqq \min(n_2,u),$$

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

$$\Pr[S=s \mid S+R_1+R_2 = u]$$

(2.7)
$$= \frac{1}{\left\{ {n_1+n_2 \atop u} \right\} s!} \sum_{t_1,t_2} \left\{ {n_1 \atop t_1} \right\} \left\{ {n_2 \atop t_2} \right\} t_1^{(s)} t_2^{(s)} ,$$

$$0 \leqq s \leqq \min(n_1,n_2,u) ,\quad u \leqq \min(n_1+n_2,m),$$

where the summation runs over $1 \leq t_1 \leq n_1$, $1 \leq t_2 \leq n_2$ and $t_1+t_2 = s+u$.

Since the exponential generating function of $\left\{ {n \atop m} \right\}$, $n = m, m+1, \ldots$, is

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

the exponential generating function of the family of joint probabilities of (2.4), for $n_i = 1,2, \ldots$, is

$$\begin{aligned} (2.8) \quad & \phi(w_1,w_2;\, z_0,z_1,z_2) \\ &= \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} \frac{(mw_1)^{n_1}}{n_1!} \frac{(mw_2)^{n_2}}{n_2!} \sum_s \sum_{r_1} \sum_{r_2} z_0{}^s z_1{}^{r_1} z_2{}^{r_2} \\ &\qquad \times \Pr[(S,R_1,R_2) = (s,r_1,r_2)] \\ &= \{1 + z_1(e^{w_1}-1) + z_2(e^{w_2}-1) + z_0(e^{w_1}-1)(e^{w_2}-1)\}^m . \end{aligned}$$

The generating function of the probabilities of (R_1,R_2,R_3) for nonuniform throw-in probabilities was obtained by Popova (1968).

The exponential generating function of the family of the marginal distributions (2.5) of S is obtained by putting $z_1=z_2=1$ in (2.8):

$$(2.9) \qquad \phi(w_1,w_2;z) = \{e^{w_1}+e^{w_2}-1 + z(e^{w_1}-1)(e^{w_2}-1)\}^m .$$

Further, by differentiating (2.9) the moments of S are obtained:

$$\begin{aligned} (2.10) \quad & E[S] = m(1-(1-\tfrac{1}{m})^{n_1})(1-(1-\tfrac{1}{m})^{n_2}) , \\ & E[S(S-1)] = m(m-1)\{1-2(1-\tfrac{1}{m})^{n_1} + (1-\tfrac{2}{m})^{n_1}\} \\ &\qquad \times \{1-2(1-\tfrac{1}{m})^{n_2} + (1-\tfrac{2}{m})^{n_2}\} . \end{aligned}$$

Of particular interest is the special value

$$\Pr[S=0\ ;\ m,n_1,n_2]$$

$$(2.11)\qquad = \frac{1}{m^{n_1+n_2}} \sum_{t_1}\sum_{t_2} \left\{ {n_1 \atop t_1} \right\} \left\{ {n_2 \atop t_2} \right\} m^{(t_1+t_2)}$$

$$= \frac{1}{m^{n_1+n_2}} \sum_{v=2}^{m} m^{(v)} \sum_{t_1+t_2=v} \left\{ {n_1 \atop t_1} \right\} \left\{ {n_2 \atop t_2} \right\} ,$$

of (2.5). In the following section another aspect of (2.11) is discussed.

3. Distribution of the number of collisions.

In the previous sections we discussed the numbers of the four types of urns, disregarding the number of balls in the urns. In this section we study mainly the balls disregarding the details of the urns.

Suppose that the balls are thrown one by one. If a ball enters an urn already occupied by other balls we say "a collision occurred". The term has been used in the "hashing table technique" for computer-search of data. In our set-up of Section 1 we are concerned with collisions between balls of different colors. Let the number of urns occupied by the white balls be $T_1 = t_1$. Under this condition, Y_2 among n_2 red balls collide with white balls, that is they enter one of t_1 urns occupied by white balls, with the binomial probability

$$(3.1)\qquad \Pr[Y_2=y\,|\,T_1=t_1] = \binom{n_2}{y} \left(\frac{t_1}{m}\right)^{y} \left(1-\frac{t_1}{m}\right)^{n_2-y} ,$$

and T_1 follows (2.1).

Thus, the unconditional distribution of the number Y_2 of red balls which collide with white balls follows "the binominal distribution compounded by the occupancy distribution" :

$$\Pr[Y_2=y] = \frac{n_2^{(y)}}{m^{n_1+n_2}} \frac{1}{y!} \sum_t \left\{ {n_1 \atop t} \right\} m^{(t)} t^y (m-t)^{n_2-y} , \tag{3.2}$$

$$0 \leq y \leq n_2 ,$$

where the summation runs over $1 \leq t \leq \min(n_1, m)$. The number Y_1 of white balls which collide with red balls has the distribution (3.1) with the roles of n_1 and n_2 exchanged, which is different from (3.1) unless $n_1 = n_2$. The conditional r-th factorial moments of Y_2 in (3.1) are known to be

$$E[Y_2^{(r)} | T_1 = t_1] = n_2^{(r)} (t_1/m)^r ,$$

and those of T_1 in (2.1) are

$$E[T_1^{(r)}] = (m^{(r)}/m^{n_1}) \nabla^r m^{n_1} , \tag{3.3}$$

where ∇ denotes the backward difference. Therefore unconditionally,

$$E[Y_2^{(r)}] = \frac{n_2^{(r)}}{m^r} \sum_j \left\{ {r \atop j} \right\} \frac{m^{(j)}}{m^{n_1}} \nabla^j m^{n_1} . \tag{3.4}$$

For example,

$$E[Y_2] = n_2(1-(1-1/m)^{n_1}) , \tag{3.5}$$

$$E[Y_2^{(2)}] = n_2^{(2)}(1-(2-\frac{1}{m})(1-\frac{1}{m})^{n_1} + (1-\frac{1}{m})(1-\frac{2}{m})^{n_1}) ,$$

and

$$\mathrm{Var}[Y_2] = n_2(1+\frac{n_2-1}{m})(1-\frac{1}{m})^{n_1} + n_2(n_2-1)(1-\frac{1}{m})(1-\frac{2}{m})^{n_1}$$

$$- n_2^2 (1-\frac{1}{m})^{2n_1} .$$

The event $S = 0$ is identical with the events $Y_2 = 0$ and $Y_1 = 0$, and

$$(3.6) \qquad \Pr[Y_2=0] = E^{T_1}[(1 - T_1/m)^{n_2}]$$

$$= \frac{1}{m^{n_1}} \sum_t \left\{ {n_1 \atop t} \right\} m^{(t)} (1-\frac{t}{m})^{n_2}$$

$$= \frac{1}{m^{n_2}} \sum_t \left\{ {n_2 \atop t} \right\} m^{(t)} (1-\frac{t}{m})^{n_1}$$

To see the equivalence of the last two expressions and (2.11), develop $(m-t)^{n_2}$ or $(m-t)^{n_1}$ using (1.1).

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 1. The probability $\Pr[S=0; m,n_1,n_2]$ of (2.11) or (3.6) 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.

Lower bounds.

Since $h(t) := (1 - t/m)^{n_2}$ in (3.6) is a convex function of t , a simple and good lower bound is

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

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

$$(4.2) \qquad h(E[T_1]) + \frac{1}{2} h''(E[T_1]) \operatorname{Var}[T_1]$$

$$= L_1(m,n_1,n_2) \times [1 + \frac{n_2(n_2-1)}{2m}(1-\frac{1}{m})^{n_1} \times \{1-(1-\frac{1}{m-1})^{n_1}$$

$$+ m((1-\frac{1}{m-1})^{n_1} - (1-\frac{1}{m})^{n_1})\}] =: L_0(m,n_1,n_2).$$

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

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

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)

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

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

There are other rough evaluations. Suppose T_1 in (3.6) takes the largest possible value with probability one, and

$$(4.5) \qquad (1 - \frac{n_1}{m})^{n_2} \quad \text{or} \quad (1 - \frac{n_2}{m})^{n_1}$$

or the maximum of these two is smaller than L_1. Taking a part of the summation in (2.11), we have

$$(4.6) \qquad m^{(n_1+n_2)} / m^{\,n_1+n_2} .$$

The bounds (4.5) and (4.6) are not satisfactory.

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 2 The Poisson distribution with mean $\lambda = n_1^2/2(m-n_1)$ is stochastically larger than the distribution of the number of collisions $C = n_1 - T_1$,

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

(cf. (2.1)).

The proof is given in Appendix.

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

$$(4.7) \qquad \exp\{-\frac{n_1 n_2}{m} + \lambda(\exp(\frac{n_2}{m}) - 1)\} =: U_1(m,n_1,n_2) .$$

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

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

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

$$(4.8) \qquad \frac{m^{(n_1)}}{m^{n_1}}\left(1 - \frac{n_1}{m}\right)^{n_2} + 1 - e^{-\lambda} =: U_2(m,n_1,n_2)$$

$$\leq (1 - \frac{n_1}{m})^{n_2} e^{-\lambda} + 1 - e^{-\lambda} .$$

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

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. For example, one white ball after five red balls; two white balls and two red balls alternately; only white balls after ten red balls; and so on. Let N_1 and N_2 denote the numbers of the white and red balls respectively when the first collision between the two colors occurs. That is, $N_1 + N_2$ is the waiting time of collisions, and Pr[S=0] is the probability that "$N_1 \geq n_1$ and $N_2 \geq n_2$ and one of the inequalities is strict." The evaluation of the preceding subsections leads to the following theorem.

Theorem. Let N_1 and N_2 denote the numbers of white and red balls respectively when the balls are thrown one by one according to a rule and the first collision between the two colors occurs. For any positive number $M > 0$, as $m \to \infty$

$$\Pr[N_1 N_2/m \le w] \to 1 - e^{-w} , \quad \text{for} \quad 0 < w < M .$$

Proof. The upper bound (4.7) is written as

$$U_1(m,n_1,n_2) = \exp\{ -\frac{n_1 n_2}{m} (1 + \frac{n_1}{2(m - n_1)} + O((\frac{n_1}{m})^2)\} ,$$

and as $m \to \infty$

$$\min(U_1(m,n_1,n_2), U_1(m,n_2,n_1)) \to \exp(-n_1 n_2/m)$$

since $\min(n_1,n_2)/m \to 0$. On the other hand the lower bound (4.1)

$$L_1(m,n_1,n_2) \to \exp(-n_1 n_2/m)$$

as $m \to \infty$. Thus $\Pr[N_1 N_2 > mw]$ tends to e^{-w} for any fixed $w > 0$ as $m \to \infty$.

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

$$2we^{-w^2} , \quad 0 < w < \infty .$$

Refer to Hirano (1986) for the Rayleigh distribution.

References

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

[2] Fang, Kai-Tai (1985) Occupancy problems, Encyclopedia of Statistical Sciences, Vol. 6, 402-406, John Wiley, New York.

[3] Hirano, K. (1986) Rayleigh distribution, Encyclopedia of Statistical Sciences, Vol. 7, 647-649, John Wiley, New York.

[4] Johnson, N.L. and Kotz, S. (1977) Urn Models and Their Applications, John Wiley, New York.

[5] Jordan, Charles (Koroly) (1950, 1960) Calculus of Finite Difference, Chelsea Publishing Co., New York.

[6] Knuth, D.E. (1967-1981) The Art of Computer Programming, Vol. 1-3, Addison-Wesley, Reading, Mass.

[7] Kolchin, Sevast'yanov and Chistyakov (1978) Random Allocation, V.H. Wistons and Sons, Washington, D.C.

[8] Mueller-Schloer, C. (1983) DES-generated checksums for electronic signatures, Cryptologia 7, 257-273.

[9] Nishimura, K. and Sibuya, M. (1987) Probability to meet in the middle (in preparation).

[10] Popova, T.Yu (1968) Limit theorems in a model of distribution of particles of two types, Theory of Probability and Its Applications 13, 511-516.

[11] Riordan, John (1958) An Introduction to Combinatorial Analysis, John Wiley, New York.

[12] Sibuya, M. (1986) Stirling family of probability distributions, Japan. J. Appl. Statist. 15, (in press, in Japanese).

Appendix

Proposition 1 is straightforward from Lemma 1.

Lemma 1 The convolution of the Stirling numbers of the second kind

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

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

$$\text{(A1)} \qquad \sum_t \left\{ {k \atop t} \right\} \left\{ {n-k \atop v-t} \right\} < \sum_t \left\{ {j \atop t} \right\} \left\{ {n-j \atop v-t} \right\}$$

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

Proof It is sufficient to prove (A1) for the case $j = k-1 < k < n-k < n-k+1 = n-j$. Apply the recurrence formula (1.2) to $\left\{ {k \atop t} \right\}$ of the left-hand side and $\left\{ {n-k+1 \atop v-t} \right\}$ of the right-hand side. Then (A1) is equivalent to

$$\text{(A2)} \qquad \sum_t t \left\{ {k-1 \atop t} \right\} \left\{ {n-k \atop v-t} \right\} < \sum_t t \left\{ {k-1 \atop v-t} \right\} \left\{ {n-k \atop t} \right\} .$$

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

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

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

$$t \left\{ {k-1 \atop t} \right\} \left\{ {n-k \atop v-t} \right\} + (v-t) \left\{ {k-1 \atop v-t} \right\} \left\{ {n-k \atop t} \right\}$$

$$< (v-t) \left\{ {k-1 \atop t} \right\} \left\{ {n-k \atop v-t} \right\} + t \left\{ {k-1 \atop v-t} \right\} \left\{ {n-k \atop t} \right\} ,$$

and (A2) is proved.

For proving Proposition 2, another lemma is needed.

Lemma 2 For any positive integer $n \geqq 3$,

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

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

$$(m+1)\left\{ {n+1 \atop n-m} \right\}^2 - (m+2)\left\{ {n+1 \atop n-m+1} \right\}\left\{ {n+1 \atop n-m-1} \right\}$$

$$= (n-m)^2\left[(m+1)\left\{ {n \atop n-m} \right\}^2 - (m+2)\left\{ {n \atop n-m+1} \right\}\left\{ {n \atop n-m-1} \right\}\right]$$

$$+ (n-m+1)\left[m\left\{ {n \atop n-m} \right\}\left\{ {n \atop n-m-1} \right\} - (m+2)\left\{ {n \atop n-m+1} \right\}\left\{ {n \atop n-m-2} \right\}\right]$$

$$+ 2\left\{ {n \atop n-m} \right\}\left\{ {n \atop n-m-1} \right\} + (m+2)\left\{ {n \atop n-m+1} \right\}\left\{ {n \atop n-m-1} \right\}$$

$$+ \left[(m+1)\left\{ {n \atop n-m-1} \right\}^2 - (m+2)\left\{ {n \atop n-m} \right\}\left\{ {n \atop n-m-2} \right\}\right] .$$

All the terms are positive and Lemma 2 is proved.

To prove Proposition 2, put

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

and define the Poisson distribution function

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

with

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

Due to Lemma 2,

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

is decreasing in x , and

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

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