Some Combinatorial Identities and Explanations Based on Occupancy Model

Tan Mingshu

Department of Mathematics
Chongqing Three Gorges University
Chongqing 404000, P.R. China
Tan_mingshu@163.com

Abstract

Some special random variables in occupancy model that balls are distributed into m urns are investigated. The number of occupied urns and the minimal number of balls in all urns are discussed. Some combinatorial identities and their explanations related to the binomial coefficient and Stirling number are derived. Several new infinite summation combinatorial identities on the binomial coefficients are obtained.

1 Introduction

It is an important method to find and prove some combinatorial identities based on probabilistic models[1−8]. In this paper, some classical occupancy models are investigated again by discussing some interesting random variables, and several combinatorial identities or combinatorial explanations are given.

2 Occupied Number in Occupancy Model

In this section, suppose that n balls fall into m urns, we consider the random variable X the number of occupied urns.

Set $X_i = \begin{cases} 1 & \text{the } i\text{th urn is occupied} \\ 0 & \text{otherwise} \end{cases}$, then

$$X = X_1 + X_2 + \cdots + X_m.$$

2.1 Undistinguishable balls and distinguishable urns

Model 1 n like balls are distributed randomly into m distinguishable urns. If vacant urn is permitted and the capacity of urns is unlimited, then there are $\binom{n+m-1}{m-1}$ ways. Furthermore, If no vacant urn exists and $n \geq m$, then there are $\binom{n-1}{m-1}$ ways.

Clearly,

$$P(X = k) = \frac{\binom{n-1}{n-k}\binom{m}{k}}{\binom{n+m-1}{m-1}}, \quad k = 1, 2, \cdots, m.$$

Therefore,

$$\sum_{k=1}^{m} P(X = k) = \sum_{k=1}^{m} \frac{\binom{n-1}{n-k}\binom{m}{k}}{\binom{n+m-1}{m-1}} = 1.$$

This derives Vandermode identity

$$\sum_{k=1}^{m} \binom{n-1}{n-k}\binom{m}{k} = \binom{n+m-1}{n}. \tag{1}$$

Namely, we give a combinatorial explanation for (1).

Furthermore, if we draw n balls without replacement from an urn containing m red balls and $n-1$ white balls, considering A_i event that i red balls are drawn. By $P\left(\bigcup_{i=0}^{m} A_i\right) = 1$, (1) is obtained also.

Because

$$P(X_i = 1) = \frac{\binom{n+m-1}{n} - \binom{n+m-2}{n}}{\binom{n+m-1}{n}} = \frac{n}{n+m-1}, \quad i = 1, 2, \cdots, m.$$

Then the mean $EX = \frac{mn}{n+m-1}$. By $EX = \sum_{k=1}^{m} kP\{X = k\}$,

$$\sum_{k=1}^{m} k\binom{n-1}{n-k}\binom{m}{k} = \frac{mn}{n+m-1}\binom{n+m-1}{n}. \tag{2}$$

In fact, (2) is equivalent to Vandermode involution formula

$$\sum_{k=1}^{m} \binom{n-1}{n-k}\binom{m-1}{k-1} = \binom{n+m-2}{n-1}.$$

In addition,

$$E(X^2) = E(X_1 + X_2 + \cdots + X_m)^2 = m(m-1)E(X_1X_2) + mE(X_1^2),$$

$$P(X_1 X_2 = 1) = \frac{\binom{n+m-1}{n} - 2\binom{n+m-2}{n} + \binom{n+m-3}{n}}{n+m-1} = \frac{n(n-1)}{(n+m-1)(n+m-2)},$$

$$P(X_1^2 = 1) = P(X_1 = 1) = \frac{n}{n+m-1},$$

Hence

$$E(X^2) = \frac{mn(mn-1)}{(n+m-1)(n+m-2)}.$$

Thus we obtain

$$\sum_{k=1}^{m} k^2 \binom{n-1}{n-k}\binom{m}{k} = \binom{n+m-1}{n} \frac{mn(mn-1)}{(n+m-1)(n+m-2)}. \tag{3}$$

For example, $n = 5, m = 3$, the right and left are 105.
Now, we study the random variable X in another way.

$$P(X = k) = \binom{m}{k} P\Big(X_1 = 1, \cdots, X_k = 1, X_{k+1} = 0, \cdots, X_m = 0\Big). \tag{4}$$

By the inclusion-exclusion principle,

$$P\Big(X_1 = 1, X_2 = 1, \cdots, X_k = 1, X_{k+1} = 0, X_{k+2} = 0, \cdots, X_m = 0\Big)$$

$$= \sum_{i=0}^{m-k} (-1)^i \binom{m-k}{i} P(X_1 = 1, X_2 = 1, \cdots, X_{k+i} = 1).$$

In general,

$$P\Big(X_1 = 1, X_2 = 1, \cdots, X_l = 1\Big) = \frac{\binom{n-l+m-1}{m-1}}{\binom{n+m-1}{m-1}}.$$

Thus

$$P(X = k) = \binom{m}{k} \sum_{i=0}^{m-k} (-1)^i \binom{m-k}{i} \frac{\binom{n-k-i+m-1}{m-1}}{\binom{n+m-1}{m-1}} = \frac{\binom{n-1}{k-1}\binom{m}{k}}{\binom{n+m-1}{m-1}}.$$

That is, for any $m(k \le m \le n)$,

$$\sum_{i=0}^{m-k} (-1)^i \binom{m-k}{i} \binom{n-k-i+m-1}{m-1} = \binom{n-1}{k-1}. \tag{5}$$

For example, (5) holds when $n = 5, k = 2$ for any $2 \le m \le 5$, the right and left are 4 .

2.2 Distinguishable balls and distinguishable urns
Model 2 n Distinguishable balls are distributed randomly into m distinguishable urns. If the capacity of urns is unlimited and no vacant urn exists and

$n \geq m$, then there are $S(n, m)m!$ ways, where $S(n, k)$ is the Stirling number of the second.

By the definition of $S(n, m)$,

$$\sum_{k=1}^{m} S(n, k)(m)_k = m^n, \quad P(X = k) = \frac{S(n, k)(m)_k}{m^n}, \quad k = 1, 2, \cdots, m,$$

$$P(X_i = 1) = 1 - (1 - \frac{1}{m})^n, \quad X = X_1 + X_2 + \cdots + X_m.$$

Where $(x)_k = x(x - 1) \cdots (x - k + 1)$.

The mean

$$EX = \sum_{k=1}^{m} k \frac{S(n, k)(m)_k}{m^n} = m\left(1 - (1 - \frac{1}{m})^n\right).$$

$$\sum_{k=1}^{m} kS(n, k)(m)_k = m^{n+1} - m(m - 1)^n. \tag{6}$$

$$\begin{aligned}
P(X_1 X_2 = 0) &= 1 - P(X_1 = 0) - P(X_2 = 0) + P(X_1 = 0, X_2 = 0) \\
&= 1 - 2P(X_1 = 0) + P(X_1 = 0, X_2 = 0) \\
&= \frac{m^n - 2(m - 1)^n + (m - 2)^n}{m^n}. \\
E(X^2) &= E(X_1 + X_2 + \cdots + X_m)^2 \\
&= m(m - 1)E(X_1 X_2) + mE(X_1^2) \\
&= m(m - 1)\frac{m^n - 2(m - 1)^n + (m - 2)^n}{m^n} + m\left(1 - \frac{(m - 1)^n}{m^n}\right)
\end{aligned}$$

We have the combinatorial identity

$$\sum_{k=1}^{m} k^2 S(n, k)(m)_k = m(m-1)\left(m^n - 2(m-1)^n + (m-2)^n\right) + m\left(m^n - (m-1)^n\right). \tag{7}$$

For example, $m = 5, m = 3$,

$$\sum_{k=1}^{3} k^2 S(5, k)(3)_k = 6\left(3^5 - 2 \times 2^5 + 1\right) + 3\left(3^5 - 2^5\right) = 1713.$$

Now, we study the random variable X in another view as above.

$$P(X = k) = \binom{m}{k} \sum_{i=0}^{m-k} (-1)^i \binom{m - k}{i} P\left(X_1 = 1, X_2 = 1, \cdots, X_{k+i} = 1\right).$$

In general,

$$P\left(X_1 = 1, X_2 = 1, \cdots, X_l = 1\right) = \frac{1}{m^n} \sum_{j=0}^{l} (-1)^j \binom{l}{j} (m - j)^n.$$

Applying the same argument similar to (4), we have

$$P(X = k) = \binom{m}{k} \sum_{i=0}^{m-k} (-1)^i \binom{m-k}{i} \frac{1}{m^n} \sum_{j=0}^{k+i} (-1)^j \binom{k+i}{j} (m-j)^n.$$

That is, for any $m(k \le m \le n)$, we obtain the same explicit expression on the Stirling number of the second as in [8]

$$S(n, k) = \frac{1}{k!} \sum_{i=0}^{m-k} \sum_{j=0}^{k+i} (-1)^{i+j} \binom{m-k}{i} \binom{k+i}{j} (m-j)^n. \tag{8}$$

For example, (8) holds when $n = 5, k = 2$ for $m = 2, 3, 4, 5$, the right and left is 10 .

Case $n = 5, k = 3$, for $m = 3$, or $m = 4$ shown as the following:

$$
\begin{aligned}
S(5, 3) &= \frac{1}{3!} \sum_{i=0}^{3-3} \sum_{j=0}^{3+i} (-1)^{i+j} \binom{3-3}{i} \binom{3+i}{j} (3-j)^5 \\
&= \frac{1}{3!} \sum_{j=0}^{3} (-1)^j \binom{3}{j} (3-j)^5 = \frac{1}{3!} (3^5 - 3 \times 2^5 + 3) = 25. \\
S(5, 3) &= \frac{1}{3!} \sum_{i=0}^{4-3} \sum_{j=0}^{3+i} (-1)^{i+j} \binom{4-3}{i} \binom{3+i}{j} (4-j)^5 \\
&= \frac{1}{3!} \Big[\sum_{j=0}^{3} (-1)^j \binom{3}{j} (4-j)^5 + \sum_{j=0}^{4} (-1)^{1+j} \binom{4}{j} (4-j)^5 \Big] \\
&= \frac{1}{3!} [4^5 - 3 \times 3^5 + 3 \times 2^5 - 1 - 4^5 + 4 \times 3^5 - 6 \times 2^5 + 4] = 25.
\end{aligned}
$$

3 The Minimal Number of Balls in Urns

In this section, we consider model that n balls are distributed into m urns, and the random variable Y the minimal number of balls in urns.

3.1 Undistinguishable balls and distinguishable urns

In this case, $[x]$ denotes the integral part of x, then

$$P(Y = k) = \frac{\sum_{i=1}^{m-1} \binom{n-km-1}{m-i-1} \binom{m}{i}}{\binom{n+m-1}{m-1}} = \frac{\binom{n-km+m-1}{m-1} - \binom{n-km-1}{m-1}}{\binom{n+m-1}{m-1}}, k = 0, 1, 2, \cdots, [\frac{n-1}{m}].$$

$P(Y = [\frac{n}{m}]) = \frac{1}{\binom{n+m-1}{n}}$ if $m|n$.

3.2 Undistinguishable balls and undistinguishable urns

This case is corresponding to integer partition. Let $p(n, m)$ be the number of integer n is divided into sum of m positive integers. For example, $5 = 1 + 4 = 2 + 3$, then $p(5, 2) = 2$.

$$P(Y = k) = \frac{\sum\limits_{i=1}^{m-1} p(n - km, m - i)}{\sum\limits_{k=1}^{m} p(n, k)}.$$

Thus,

$$\sum_{k=0}^{\lceil \frac{n-1}{m} \rceil} \sum_{i=1}^{m-1} p(n - km, m - i) + \lambda_{m,n} = \sum_{k=1}^{m} p(n, k). \tag{9}$$

Where $\lceil x \rceil$ is the minimal integer that is not less than x, $\lambda_{m.n} = \begin{cases} 1 & m \mid n, \\ 0 & m \nmid n \end{cases}$.

For example, $n = 5, m = 3$ and $n = 6, m = 3$.

$$\sum_{i=1}^{2} p(5, 3 - 1) + \sum_{i=1}^{2} p(2, 3 - i) = (2 + 1) + (1 + 1) = 5.$$

$$\sum_{i=1}^{2} p(6, 3 - 1) + \sum_{i=1}^{2} p(3, 3 - i) + 1 = (3 + 1) + (1 + 1) + 1 = 6.$$

4 Sequential Distribution Model

Model 3   Assume that like balls are sequentially distributed at random into m distinguishable urns of unlimited capacity.

Three random variables investigated respectively. Some infinite summation combinatorial identities related to binomial coefficients are obtained.

let W_k be the number of balls required to be distributed until k urns are occupied[1, p.245].

$$P(W_k = n) = \frac{\binom{m-1}{k-1}\binom{n-2}{k-2}}{\binom{m+n-1}{n-1}},$$

then

$$\sum_{n=k}^{\infty} \frac{\binom{n-2}{k-2}}{\binom{m+n-1}{n-1}} = \frac{1}{\binom{m-1}{k-1}} (m \geq k \geq 2), \tag{10}$$

i.e.

$$\sum_{n=k}^{\infty} \frac{\binom{n-1}{k-1}}{\binom{m+1+n}{n}} = \frac{1}{\binom{m}{k}} (m \geq k \geq 1). \tag{11}$$

Let V_l be the number of balls required to be distributed until the minimal number of balls in all m urns reaches l respectively. Then

$$P(V_l = n) = \frac{\binom{n-(l-1)m-2}{m-2} m}{\binom{n+m-1}{m}} \frac{l}{m} = \frac{\binom{n-lm+m-2}{m-2} l}{\binom{n+m-1}{m}} (n \geq ml, l \geq 1).$$

Thus

$$\sum_{n=ml}^{\infty} \frac{\binom{n-lm+m-2}{m-2}}{\binom{n+m-1}{m}} = \frac{1}{l}. \tag{12}$$

This is (10) when $l = 1$ and $m = k$. That is the case all urns are not vacant. For example, $m = 2$,

$$\sum_{n=2l}^{\infty} \frac{l}{n(n+1)/2} = 1.$$

$m = 3$,

$$\sum_{n=3l}^{\infty} \frac{(n-3l+1) \times 6l}{n(n+1)(n+2)} = 6l \sum_{n=3l}^{\infty} \frac{1}{n(n+2)} - 18l^2 \sum_{n=3l}^{\infty} \frac{1}{n(n+1)(n+2)}$$

$$= 6l\frac{1}{2}\left(\frac{1}{3l} + \frac{1}{3l+1}\right) - 18l^2 \frac{1}{2 \times 3l(3l+1)} = 1.$$

$m = 4$,

$$12l \sum_{n=4l}^{\infty} \frac{(n-4l+2)(n-4l+1)}{n(n+1)(n+2)(n+3)}$$

$$= 12l \sum_{n=4l}^{\infty} \frac{1}{n(n+3)} - 48l^2 \Big(\sum_{n=4l}^{\infty} \frac{1}{n(n+1)(n+3)} + \sum_{n=4l}^{\infty} \frac{1}{n(n+2)(n+3)} \Big)$$

$$+ 12 \times 16l^3 \sum_{n=4l}^{\infty} \frac{1}{n(n+1)(n+2)(n+3)}$$

$$= 12l \times \frac{1}{3}\left(\frac{1}{4l} + \frac{1}{4l+1} + \frac{1}{4l+2}\right) - 48l^2 \Big(\sum_{n=4l}^{\infty} \frac{1}{n(n+2)} - \sum_{n=4l}^{\infty} \frac{1}{(n+1)(n+3)} \Big)$$

$$+ 12 \times 16l^3 \times \frac{1}{3 \times 4l(4l+1)(4l+2)}$$

$$= 12l \times \frac{1}{3}\left(\frac{1}{4l} + \frac{1}{4l+1} + \frac{1}{4l+2}\right) - 48l^2 \Big(\frac{1}{2}\left(\frac{1}{4l} + \frac{1}{4l+1}\right) - \frac{1}{2}\left(\frac{1}{4l+1} + \frac{1}{4l+2}\right)\Big)$$

$$+ 12 \times 16l^3 \times \frac{1}{3 \times 4l(4l+1)(4l+2)} = 1.$$

In addition, case $m = 5$ is true by computing.

Furthermore, let $V_{l,r}$ be the number of balls required to be distributed until the minimal number of balls in the appointed r $(2 \le r \le m-1)$ urns reaches l $(l \ge 1)$ respectively.
See V_l when $r = m$, and if $r = 1$ we have

$$\sum_{n=l}^{\infty} \frac{\binom{n-l+m-2}{m-2}}{\binom{n+m-1}{m}} = \frac{m}{l} \ (m \ge 2). \tag{13}$$

Then

$$P(V_{l,r} = n) = \frac{lr}{m} \sum_{k=lr}^{n} \frac{\binom{k-(l-1)r-2}{r-2} \binom{n-k+m-r-1}{m-r-1}}{\binom{n+m-1}{m}} \ (n \geq lr).$$

It follows that

$$\sum_{n=lr}^{\infty} \sum_{k=lr}^{n} \frac{\binom{k-(l-1)r-2}{r-2} \binom{n-k+m-r-1}{m-r-1}}{\binom{n+m-1}{m}} = \frac{m}{lr}. \tag{14}$$

For example, $m = 3, r = 2$.

$$\sum_{n=2l}^{\infty} \frac{6(n-2l+1)}{(n+2)(n+1)n} = 6 \sum_{n=2l}^{\infty} \frac{1}{(n+2)n} - 12l \sum_{n=2l}^{\infty} \frac{1}{(n+2)(n+1)n}$$

$$= 6 \times \frac{1}{2}(\frac{1}{2l} + \frac{1}{2l+1}) - 12l \times \frac{1}{2 \times 2l(2l+1)} = \frac{3}{2l} = \frac{m}{lr}.$$

$m = 4, r = 2$.

$$\sum_{n=2l}^{\infty} \sum_{k=2l}^{n} \frac{n-k}{\binom{n+3}{4}} = 12 \sum_{n=2l}^{\infty} \frac{n(n+1) - (4l-2)n + (4l^2 - 6l + 2)}{(n+3)(n+2)(n+1)n}$$

$$= 12\Big(\sum_{n=2l}^{\infty} \frac{1}{(n+3)(n+2)} - (4l-2) \sum_{n=2l}^{\infty} \frac{1}{(n+3)(n+2)(n+1)}$$

$$+ (4l^2 - 6l + 2) \sum_{n=2l}^{\infty} \frac{1}{(n+3)(n+2)(n+1)n} \Big)$$

$$= 12\Big(\frac{1}{2l+2} - (4l-2)\frac{1}{2(2l+1)(2l+2)} + \frac{4l^2 - 6l + 2}{3 \times 2l(2l+1)(2l+2)} \Big) = \frac{2}{l} = \frac{m}{lr}.$$

References

[1] Ch A.Charalambides, Combinatorial Methods in Discrete Distribution, New Jersey,John Wiley&Sons Inc., 2005.

[2] Arfwedson G, A probability distribution connected with Stirling's second class numbers, Skandinavisk Aktuarietidskrift, 34(1951),121-132.

[3] P.Erdös, J.Spencer, Probabilistic methods in combinatorics, New York, Academic Press, 1974.

[4] Ch A.Charalambides, A review of the Stirling numbers, their generalizations and statistical applications, Communications in Statistics Theory and Methods, 17(1988),2533-2595.

[5] H.Yamato, A probabilistic approach to Stirling numbers of the first kind, Communications in Statistics-Theory and Methods, 19(1990), 3915-3923.

[6] K.Nishimura, M.Sibuya, Extended Stirling family of discrete probability distributions, Communications in Statistics-Theory and Methods, 26(1997),1727-1744.

[7] Sun Ping, Wang Tianming, Probabilistic represents of Stirling number and its applications(In Chinese), Acta Mathematica Sinica (Chinese Series), 41(1998),281-290.

[8] Tan Mingshu, Some Combinatorial Identities on the Stirling Number Based on a Probabilistic Model, ARS Combinatoria, in press.

Received: December, 2010