A Note on a Binomial Identity

Do Sun Bai*

A binomial identity is proved and two examples of its application are displayed; one in connection with the probability distribution of the sample size random variable in a certain sequential sampling plan based on independent Bernoulli trials, and the other in connection with a generalization of the "Banach's match box problem."

Theorem For any nonnegative integers r_1 and r_2 and $0 \leqq p \leqq 1$ with $q = 1-p$, define $A_p(r_1, r_2)$ by

$$A_p(r_1,r_2) = p^{r_1+1} \sum_{k=0}^{r_2} \binom{k+r_1}{r_1} q^k + q^{r_2+1} \sum_{k=0}^{r_1} \binom{k+r_2}{r_2} p^k \tag{1}$$

Then

$$A_p(r_1,r_2) \equiv 1. \tag{2}$$

Proof The case $p=0$ or 1 is easy. Next, assume that $0<p<1$. We will use mathematical induction. It can easily be seen that (2) holds for $r_1, r_2 = 0$ or 1. Assume that $A_p(r_1,r_2)=1$ for some r_1 and r_2. Then by reasons of symmetry it is sufficient to show that $A_p(r_1, r_2+1)=1$.

$$A_p(r_1,r_2+1) = p^{r_1+1} \sum_{k=0}^{r_2+1} \binom{k+r_1}{r_1} q^k + q^{r_2+2} \sum_{k=0}^{r_1} \binom{k+r_2+1}{r_2+1} p^k$$

$$= \binom{r_1+r_2+1}{r_1} p^{r_1+1} q^{r_2+1} + [1 - q^{r_2+1} \sum_{k=0}^{r_1} \binom{k+r_2}{r_2} p^k] + q^{r_2+2} \sum_{k=0}^{r_1} \binom{k+r_2+1}{r_2+1} p^k$$

$$= 1 + \binom{r_1+r_2+1}{r_1} p^{r_1+1} q^{r_2+1} - q^{r_2+1} \sum_{k=0}^{r_1} p^k [\binom{k+r_2}{r_2} - q \binom{k+r_2+1}{r_2+1}]$$

Now,

$$\sum_{k=0}^{r_1} p^k [\binom{k+r_2}{r_2} - q\binom{k+r_2+1}{r_2+1}] = \sum_{k=0}^{r_1} [p^{k+1} \binom{k+r_2+1}{r_2+1} - p^k \binom{k+r_2}{r_2+1}]$$

* Department of Industrial Engineering, Korea Advanced Institute of Science, Seoul.

$$=p^{r_1+1}\binom{r_1+r_2+1}{r_2+1}-\binom{r_2}{r_2+1}$$

$$=p^{r_1+1}\binom{r_1+r_2+1}{r_2+1}.$$

Hence we have $A_p(r_1,r_2+1)=1$.

Example 1 (A sequential binomial sampling plan) Suppose that independent observations are to be taken sequentially on the Bernoulli random variable U with $P(U=1)=p$ and $P(U=0)=q=1-p$, $0<p<1$. The observations are to be stopped as soon as at least k_1 ones and at least k_2 zeroes are obtianed ([3], pp.144): that is when

$$\sum_{i=1}^{N} U_i=k_1 \text{ and } N-\sum_{i=1}^{N} U_i\geqq k_2,$$

or

$$\sum_{i=1}^{N} U_i\geqq k_1 \text{ and } N-\sum_{i=1}^{N} U_i=k_2,$$

where k_1 and k_2 are preassigned positive integers and N denotes the total number of observations required under this sampling plan. Then it can be shown that the probability a_n that $N=n$ is given by

$$a_n=P(N=n)=\binom{n-1}{k_1-1}p^{k_1}q^{n-k_1}+\binom{n-1}{k_2-1}p^{n-k_2}q^{k_2} \tag{3}$$

$$n=k_1+k_2,\ k_1+k_2+1,\cdots\cdots\cdots\cdots\cdots\cdots$$

To show that $\{a_n\}$ is really a probability distribution, we have to show that $\sum_n a_n=1$. Now,

$$\sum_n a_n=\sum_{n=k_1+k_2}^{\infty}\binom{n-1}{k_1-1}p^{k_1}q^{n-k_1}+\sum_{n=k_1+k_2}^{\infty}\binom{n-1}{k_2-1}p^{n-k_2}q^{k_2}$$

$$=2-[p^{k_1}\sum_{j=0}^{k_2-1}\binom{j+k_1-1}{k_1-1}q^j+q^{k_2}\sum_{j=0}^{k_1-1}\binom{j+k_2-1}{k_2-1}p^j]$$

$$=2-A_p(k_1-1,k_2-1)=2-1=1.$$

Example 2 (A generalization of the Banach's match box problem) Consider two boxes A and B containing r_1 and r_2 matches respectively.

Matches are drawn one at a time with probability p from box A and with probability $q=1-p$ from box B. The moment when, for the first time, it is discovered that a box is empty, the other box may contain k ($k=0, 1, \cdots, r_1(r_2)$) matches.

Therefore the event that box A will be found to be empty when box B contains exactly k matches (*i.e.*, r_2-k matches have already been used up from box B) is equivalent to the event that exactly r_2-k failures preceed $(r_1+1)th$ success in a sequence of independent Bernoulli trials with success probability p. The probability of this event is then given by

$$\binom{r_1+r_2-k}{r_1} p^{r_1}q^{r_2-k}p=\binom{r_1+r_2-k}{r_1} p^{r_1+1} q^{r_2-k} \tag{4}$$

where $k=0, 1, \cdots, r_2$.

Similarly, the probability that box B will be discovered empty when box A contains exactly k matches (*i.e.*, r_1-k matches have already been used up from box A) is given by

$$\binom{r_1-k+r_2}{r_2} p^{r_1-k}q^{r_2}q=\binom{r_1-k+r_2}{r_2} p^{r_1-k}q^{r_2+1} \tag{5}$$

where $k=0, 1, \cdots, r_1$.

Combining (4) and (5), we obtain the probability b_k that when one box is found to be empty, the other box will contain exactly k matches as

$$\begin{aligned} b_k&=\binom{r_1+r_2-k}{r_1} p^{r_1+1}q^{r_2-k}+\binom{r_1-k+r_2}{r_2} p^{r_1-k}q^{r_2+1}, \\ &\qquad k=0, 1, \cdots, \min(r_1,r_2), \qquad (6) \\ &=u(r_2-r_1)\binom{r_1+r_2-k}{r_1} p^{r_1+1}q^{r_2-k}+u(r_1-r_2)\binom{r_1-k+r_2}{r_2} p^{r_1-k}q^{r_2+1}, \\ &\qquad k=\min(r_1,r_2)+1, \cdots, \max(r_1,r_2), \end{aligned}$$

where $u(x)=1$ if $x>0$, and $u(x)=0$ if $x\leqq 0$.

Now,

$$\sum_k b_k=p^{r_1+1}\sum_{k=0}^{r_2}\binom{r_1+r_2-k}{r_1} q^{r_2-k}+q^{r_2+1}\sum_{k=0}^{r_1}\binom{r_1-k+r_2}{r_2} p^{r_1-k}$$

$$= p^{r_1+1} \sum_{j=0}^{r_2} \binom{j+r_1}{r_1} q^j + q^{r_2+1} \sum_{j=0}^{r_1} \binom{j+r_2}{r_2} p^j$$

$$= A_p(r_1, r_2) = 1.$$

Thus we see that $\{b_k\}$ is a probability distribution.

This is a generalization of the "Banach's match box problem ([1], [2])" (the special case when $p=\frac{1}{2}$ and $r_1=r_2=r$), and the well-known combinatorial identity derived from this problem,

$$\sum_{k=0}^{r} \binom{k+r}{r} \left(\frac{1}{2}\right)^k = 2^r \tag{7}$$

is an immediate consequence of the theorem with

$$A^{1/2}(r,r) = \sum_{k=0}^{r} \binom{k+r}{r} \left(\frac{1}{2}\right)^{r+k} = 1.$$

REFERENCES

[1] Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed., New York, John Wiley and Sons, Inc., 1968.

[2] Riordan, J., *Combinatorial Identities*, New York, John Wiley and Sons, Inc., 1968.

[3] Wilks, S. S., *Mathematical Statistics*, New York, John Wiley and Sons, Inc., 1962.