J. Univ. Kuwait (Sci.) **8,** 1981

On the distribution of a sum of non-identical independent binomials

M. T. ABDULNASSER AND A. M. KHIDR

Department of Mathematics, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

A theorem on the distribution of W_n^l, the sum of the independent random variables $\{X_{n,i}\}_{i=1}^n$, where $X_{n,i}^l$ is $b(l,i/n)$, $l=1, 2, \ldots$, is proved. The exact distribution is found for $l=1$, and a probability-generating formula is obtained for $l=2, 3, \ldots$.

INTRODUCTION

In industry, training programmes are conducted with the aim of training new workers to do a particular job repeatedly every day. It is assumed that a particular trainee will show progress proportional to the number of days he attends the programme, otherwise his ability will be indifferent from one day to another. Industry is just one of the vast variety of situations where such programmes are conducted. Let n be the length of a programme in days and l the number of repetitions of the job per day a trainee has to do. If a trainee is responding to the instructions, it would be reasonable to assume the probability that he will do a single job right on the ith day is $p_{i,n}=i/n$ and the probability that he will do x jobs correctly out of l jobs on the ith day is $\binom{l}{x}(i/n)^x (1-i/n)^{l-x}$, $x=0, 1, 2, \ldots, l$. When a trainee is not responding to the instructions, $p_{i,n}$ will be a constant p, $0<p<1$.

In the latter case, the probability that he will do x jobs correctly out of l jobs on any day is $\binom{l}{x}p^x(1-p)^{l-x}$. To test whether a trainee is responding or not, we test whether $p_{i,n}$ is varying or sustaining a constant value p. This can be done by computing the total number of jobs that have been done correctly over the whole period of the programme. Let $X_{n,i}^l$ stand for the number of jobs done correctly out of l jobs on the ith day, $i=1, 2, \ldots n$, $l=1, 2, \ldots$, and

$$W_n^l = \sum_{i=1}^{n} X_{n,i}^l, \quad l \leq w_n^l \leq nl.$$

In case $p_{i,n}=p$, $0<p<1$, the distribution of W_n^l will be $b(nl,p)$. Otherwise, it will be the sum of independent random variables $\{X_{n,i}^l\}_{i=1}$, where $X_{n,i}^l$ is $b(l,i/n)$, $i=1, 2, \ldots n$, and $l=1, 2, \ldots$. In this case, let

$$V_n^l = \sum_{i=1}^{l} W_n^{1,j}, \quad l \leq v_n^l \leq nl,$$

where $W_n^{1,1}, W_n^{1,2}, \ldots, W_n^{1,l}$ is a random sample of size l from the distribution of W_n^1.

THEOREMS

In this section, we will prove the following theorems.

Theorem 1. W_n^l equals V_n^l in distribution in the sense that $P(W_n^l=k)=P(V_n^l=k)$, $k=l$, $l+1, \ldots, nl$, and for $l=1, 2, \ldots$.

With $s(n,m)$, the Stirling numbers of the first kind defined by

$$x(x-1)(x-2)\ldots(x-n+1) = \sum_{m=0}^{n} s(n,m)\, x^m,$$

and satisfying the recurrence relation $s(n,m)=s(n-1,\, m-1)-(n-1)s(n-1,\, m)$ (see Riordan 1968), we prove:

Theorem 2. The distribution of W_n^1 is symmetric around its mean, and

$$P(W_n^1 = k) = (-1)^k \sum_{m=k}^{n} \frac{s(n+1,\, m+1-m)}{n^m} \binom{m}{k}, \quad k = 1, 2, \ldots, n.$$

A corollary to Theorems 1 and 2, using the result in Tsao (1956) is the following:

Corollary. For $l=2, 3, \ldots$, we have

$$P(W_n^l = k) = \sum_{i=1}^{n} P(W_n^1 = i). \quad P(W_n^{l-1} = k-i), \quad k = l, l+1, \ldots, nl.$$

We will use the probability-generating function technique to prove Theorem 1. The inclusion-exclusion principle will be used in proving Theorem 2, in which the Stirling numbers arise in a natural way. It is worth mentioning that the distribution of W_n^l is symmetric for only $l=1$. For $l\neq 1$, the mean of the distribution will not be in the middle of the range of W_n^l. Table 1 gives the distribution of W_n^1 for $n=1(1)\,17$, $k=1, 2, \ldots$, $[(n+1)/2]$, where $[a]$ stands for the greatest integer of a. For $k>[(n+1)/2]$, one can use the relation $P(W_n^1=k)=P(W_n^1=n-k+1)$.

PROOF OF THEOREMS

Theorem 1. The p.g.f., $\psi(t)$, of W_n^l is

$$\psi(t) = E(t^{W_n^l}) = \prod_{i=1}^{n} E(t^{X_{n,i}^l}) = \prod_{i=1}^{n}\left(\frac{n-i}{n}+\frac{i}{n}\,t\right)^l.$$

On the other hand, $\phi(t)$, the p.g.f. of V_n^l, is

$$\phi(t) = E(t^{V_n^l}) = [E(t^{W_n^1})]^l = [\prod_{i=1}^{n} E(t^{X_{n,i}^1})]^l$$

$$= \left[\prod_{i=1}^{n}\left(\frac{n-i}{n}+\frac{i}{n}\,t\right)\right]^l = \psi(t).$$

Thus Theorem 1 is proved.

Theorem 2. To employ the inclusion principle (Parzen 1960, p. 76), we define the event E_i as the event $X^1_{n,i}=1$, $i=1, 2, \ldots, n$, and the sum

$$P(n,k) = \sum_{1\le i_1<i_2<\ldots i_k\le n} P(E_{i_1}, E_{i_2}\ldots E_{i_k})$$

$$= \sum_{1\le i_1<i_2\ldots<i_k<n} P(X^1_{n,i_1}=1,\ldots, X^1_{n,i_k}=1)$$

$$= (1/n^k)\sum_{1\le i_1<i_2\ldots\le i_k\le n} i_1, i_2\ldots i_k$$

$$= (-1)^k\frac{s(n+1, n+1-k)}{n^k}.$$

With $P(n,k)$ as above, we get

$$P(W^1_n\ge k) = \sum_{m=k}^{n}(-1)^{m-k}\binom{m-1}{k-1}P(n,m)$$

$$= \sum_{m=k}^{n}(-1)^{m-k}(-1)^m\frac{s(n+1, n+1-m)}{n^m}\binom{m-1}{k-1}$$

$$= (-1)^k\sum_{m=k}^{n}\binom{n-1}{k-1}\frac{s(n+1, n+1-m)}{n^m},$$

from which we get

$$P(W^1_n=k) = P(W^1_n\ge k)-P(W^1_n\ge k+1)$$

$$= (-1)^k s(n+1, n+1-k)/n^k+(-1)^k\sum_{m=k+1}^{n}\left\{\frac{s(n+1, n+1-m)}{n^m}\binom{m-1}{k-1}\right.$$

$$\left.+\frac{s(n+1, n+1-m)}{n^m}\binom{m-1}{k}\right\}$$

$$= (-1)^k s(n+1, n+1-k)/n^k+\sum_{m=k+1}^{n}\frac{s(n+1, n+1-m)}{n^m}\binom{m}{k},$$

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

To show that $P(W^1_n=k)$ is a probability function, we sum over k and obtain

$$\sum_{k=1}^{n}(-1)^k\sum_{m=k}^{n}\frac{s(n+1, n+1-m)}{n^m}\binom{m}{k}$$

$$= \sum_{m=1}^{n}\frac{s(n+1, n+1-m)}{n^m}\sum_{k=1}^{m}(-1)^k\binom{m}{k}$$

$$= \sum_{m=1}^{n}\frac{s(n+1, n+1-m)}{n^m}(-1).$$

For $n+1-m=t$, the above becomes

$$-\frac{1}{n^{n+1}}\sum_{t=1}^{n} s(n+1,\, t)\, n^{t}$$
$$= -\frac{1}{n^{n+1}}\left[\{n(n-1)(n-2)\ldots(n-n)\} - \frac{1}{n^{n+1}}\left(-s(n+1,0)n^{0} - s(n+1,\, n+1)\, n^{n+1}\right)\right]$$
$$= 1, \text{ since } s(n,0) = 0, \text{ and } s(n+1,\, n+1) = 1.$$

This proves the second part of Theorem 2.

It can be shown easily that the mean of W_n^1 is equal to $(n+1)/2$ and the variance of W_n^1 is equal to $(n^2-1)/6n$.

To show the symmetry of the distribution, we start with

$$(-1)^{n-k+1}\sum_{m=n-k+1}^{n}\frac{s(n+1,\, n+1-m)}{n^m}\binom{m}{n-k+1},$$

which is the coefficient of x^{n-k+1} in $f(x)$,

$$f(x) = \sum_{m=n-k+1}^{n}\frac{s(n+1,\, n+1-m)}{n^m}(1-x)^m.$$

Put $n+1-m=t$. Then

$$f(x) = \sum_{t=1}^{k}\frac{s(n+1,\, t)}{n^{n+1-t}}(1-x)^{n+1-t}$$
$$= \frac{(1-x)^{n+1}}{n^{n+1}}\sum_{t=1}^{k} s(n+1,\, t)\,\{(1-x)n^{-1}\}^{-t}.$$

Use the substitution $1-x=(1+y)^{-1}$, $x=y\ (1+y)^{-1}$ and $y=x(1-x)^{-1}$. With this substitution, we get

$$f(y(1+y)^{-1}) = \{(1+y)n\}^{-(n+1)}\sum_{t=1}^{k} s(n+1,\, t)\,\{(1+y)n\}^{t}.$$

It can be seen from the definition of the Stirling numbers of the first kind that

$$(-1)^{n+1}\sum_{m=1}^{n+1} s(n+1,\, m)\,(-yn)^m = \sum_{m=1}^{n+1} s(n+1,\, m)\,\{(1+y)n\}^m.$$

Therefore

$$f[y(1+y)^{-1}] = \{(1+y)n\}^{-(n+1)}\Big[(-1)^{n+1}\sum_{m=1}^{n+1} s(n+1,\, m)\,(-yn)^m$$
$$-\sum_{m=k+1}^{n+1} s(n+1,\, m)\,\{(1+y)n\}^m\Big]$$

$$= (-1)^{n+1}\{(1+y)n\}^{-n+1)}\sum_{m=1}^{n+1} s(n+1, m)(-yn)^m$$

$$-\sum_{m=k+1}^{n+1} s(n+1, m)\{(1+y)n\}^{-(n+1-m)}$$

Substituting back for x, we get

$$f(x) = (-1)^{n+1}\{(1+x)/n\}^{n+1}\sum_{m=1}^{n+1} s(n+1, m)\{(-xn)/(1-x)\}^m$$

$$-\sum_{m=k+1}^{n+1} s(n+1, m)\{(1-x)n^{-1}\}^{n+1-m}.$$

The coefficient of x^{n-k+1} in the second sum of $f(x)$ above is equal to 0, and the coefficient of x^{n-k+1} in the first sum of $f(x)$ above is equal to

$$(-1)^{n+1}/n^{n+1}\sum_{m=1}^{n-k+1} n^m s(n+1, m)\binom{n+1-m}{n-k+1-m}(-1)^{m-k+1-m}(-1)^m.$$

Therefore

$$P(W_n^1 = n+1-k) = (-1)^k \sum_{m=1}^{n-k+1} \frac{s(n+1, m)}{n^{n+1-m}}\binom{n+1-m}{n+k+1-m}$$

$$= (-1)^k \sum_{n=k}^{n} \frac{s(n+1, n+1-m)}{n^m}\binom{m}{k}$$

$$= P(W_n^1 = k),$$

which completes the proof of Theorem 2.

Proof of corollary. In Tsao (1956), replace S by V_n^l, n by l, s by k, and k by n. With the use of Theorem 1, the result follows.

REFERENCES

Ash, R. 1972. Real analysis and probability. Academic Press.
Parzen, E. 1960. Modern probability theory and its applications. John Wiley.
Riordan, J. 1968. Combinatorial identities. John Wiley.
Tsao, C.K. 1956. Distribution of the sum in random samples from a discrete population. Ann. Math. Statist. **27:** 703–12.

(Received 28 April 1980)

Table 1. Distribution of W_n^1

$$P(W_n^1 = k),\ k = 1, 2, \ldots, \left[\frac{n+1}{2}\right]$$

$$P(W_n^1 = k) = P(W_n^1 = n-k+1),\ k > \left[\frac{n+1}{2}\right]$$

	1	2	3	4	5	6	7	8	9
1	1								
2	·5								
3	·222222	·555556							
4	·09375	·40625							
5	·0384	·2464	·4304						
6	·015432	·134259	·350309						
7	·006121	·068236	·243402	·364482					
8	·002861	·031883	·153645	·311302					
9	·000936	·015417	·088118	·234517	·322018				
10	·000363	·006999	·048207	·159749	·284680				
11	·000140	·003123	·025081	·101699	·222909	·294667			
12	·000054	·001356	·012735	·060301	·162349	·263204			
13	·000021	·000583	·006283	·034325	·110419	·215108	·268595		
14	·000008	·000247	·002988	·018791	·069671	·162343	·245939		
15	·000003	·000104	·001401	·009957	·042310	·114785	·206307	·250237	
16	·000001	·000043	·000645	·005132	·024690	·076853	·160938	·231556	
17	·000000	·000018	·000292	·002582	·013915	·049126	·118127	·198288	·235244

توزيع مجموع حدانيتين مستقلتين وغير متطابقتين

محمد طلعت عبد الناصر وأبو الحسن خضر
قسم الرياضيات بجامعة الملك عبد العزيز، جدة، المملكة العربية السعودية

خلاصة

لقد امكن اثبات نظرية حول توزيع W_n^l مجموع المتغيرات العشوائية المستقلة $\{X_{n,i}\}_{i=1}^n$ حيث $X_{n,i}^l$ هو $b(l,_n^i), l=1, 2 \ldots$. كما امكن ايجاد التوزيع المضبوط لـ $l=1$ وكذلك دستور مولد للاحتمالية لـ $l=2, 3, \ldots\ldots\ldots$