J. Korean Math. Soc.
Vol. 16, No. 2, 1980

A REPRESENTATION OF SUM OF DEPENDENT BERNOULLI RANDOM VARIABLES AS SUM OF INDEPENDENT RANDOM VARIABLES

BY C. J. PARK AND HAN YOUNG CHUNG

1. Introduction and summary

Consider a finite or infinite sequence of exchangeable Bernoulli random variables, $X_1, X_2, \cdots$. Let $S_n=\sum_{i=1}^{n} X_i$, $n\geq 1$, then the exchangeability of $(X_1, X_2, \cdots, X_n)$ implies that the probability distribution of S_m and the distribution of $X_{i_1}+X_{i_2}+\cdots+X_{i_m}$, $1\leq i_1<i_2<\cdots<i_m$, for $m=1,2,\cdots,\leq n$ are identical. The distribution of the sum of exchangeable Bernoulli random variables explored by De Finetti [1930] plays an important role in the Bayesian inference. In this paper we consider a sum of exchangeable Bernoulli random variables which has the following property: A sum $S_n=\sum_{i=1}^{n} X_i$ of random variables $X_1, X_2, \cdots$, is said to have an independent sum representation (*ISR*) property if there exists a sequence of independent Bernoulli random variables $Y_1, Y_2, \cdots, Y_n$ such that the distribution of S_n is the same as the distribution of $Y_1+Y_2+\cdots+Y_n$.

Thus when the sum of the exchangeable random variables has the *ISR* property the limiting distribution of the sum may be readily obtained. The sum of exchangeable Bernoulli random variables considered in the classical occupancy problems, see Feller [1972], Harris and Park [1971], and Johnson and Kotz [1977], has the *ISR* property. The sum of exchangeable Bernoulli random variables considered in the *urn* model problems, see Park [1976], [1977] and Johnson and Kotz [1977], also has the the *ISR* property

In this paper an attempt is made to characterize the sum of exchangeable Bernoulli random variables that has the *ISR* property. We have utilized the difference differential equation of the factorial moment generating function of the sum of random variables.

2. Main results

Let $X_1, X_2, \cdots$ be a sequence of exchangeable Bernoulli random variables.

Received July 26, 1979

Let $P_{n,k}=P_r[X_1+X_2+\cdots+X_n=k]$, then it can be shown that $P_{n,k}$ satisfied the following backward recurrence relation.

$$(1)\qquad P_{n-1,k}=\frac{k+1}{n}P_{n,k+1}+\frac{n-k}{n}P_{n,k}\ ,\ n\geq 2\,.$$

Let $\phi_n(t)$ be the factorial moment generating $(f.m.g.)$ function defined by

$$\phi_n(t)=\sum_{k=0}^{n}(1+t)^k P_{n,k}\ ,$$

then $\phi_n(t)$ satisfies the following difference differential equation

$$(2)\qquad \phi_{n-1}(t)=\phi_n(t)-\frac{t}{n}\phi_n'(t),\ \ n\geq 2.$$

THEOREM 1. *If the sum of exchangeable random variables* $X_1, X_2, \cdots, X_n$, *for some* n, *has the ISR property, then the sum of* $X_1, X_2, \cdots, X_m$, *for* $m<n$, *has the ISR property.*

Proof. If the sum $S_n=\sum_{i=1}^{n}X_i$ has the *ISR* property, then the $f.m.g.$ function $\phi_n(t)$ of S_n can be written as

$$\phi_n(t)=\prod_{j=1}^{n}(1+p_j),\ \ 0<p_j\leq 1,$$

and we will prove that the sum S_{n-1} has the *ISR* property by showing that the $f.m.g.$ function $\phi_{n-1}(t)$ of S_{n-1} can be written in the form

$$(3)\qquad \phi_{n-1}(t)=\prod_{j=1}^{n-1}(1+\alpha_j t),\ \text{ where } 0<\alpha_j\leq 1.$$

In equation (2), let $\phi_{n-1}(t)=0$, then we have

$$(4)\qquad \phi_n'(t)/\phi_n(t)=\frac{d}{dt}\ln\phi_n(t)=\frac{n}{t}.$$

For $t>0$, clearly $\phi_n(t)=\prod_{j=1}^{n}(1+p_j t)>0$, thus

$$\frac{d}{dt}\ln\phi_n(t)=\sum_{j=1}^{n}\frac{p_j}{1+p_j t}\ <\ \sum_{j=1}^{n}\frac{1}{t}=\frac{n}{t}.$$

This means that $\phi_{n-1}(t)=0$ cannot have positive roots. Clearly, zero is not a root of $\phi_{n-1}(t)=0$. Thus all the real roots of $\phi_{n-1}(t)=0$ must be negative. On the other hand, for $t<0$, $\phi_n'(t)/\phi_n(t)$ has simple poles at every zero of $\phi_n(t)$ (including multiple zeroes) and the equation (4) implies that $\phi_{n-1}(t)$ must have real zeros which are less than the largest zero of $\phi_n(t)$. This proves that $\phi_{n-1}(t)$ can be written in the form given by (3). Hence it follows that the sum S_{n-1} has the *ISR* property, and by induction S_m, for $m\leq n$, must have the *ISR* property.

EXAMPLE. Consider an *urn* which contains b black and r balls. We draw m balls at random one at a time without replacement and let $X_i=1(0)$ if the ith ball is a black (red) ball. Then the sum S_m of random variable $X_1, X_2, \cdots, X_m$ has the *ISR* property for $m\leq r+b$ because $\phi_{r+b}(t)^b=(1+t)^b$ which say sthat the sum $X_1+X_2+\cdots+X_{b+r}$ has *ISR* property. That is, using Theorem 1, the *f. m. g.* function $\phi_m(t)$ of the sum $X_1+X_2+\cdots+X_m$ given by

$$\phi_m(t)=\sum_{k=0}^{\min(b,m)}\left[\binom{b}{k}\binom{m}{k}\Big/\binom{r+b}{k}\right]t^k$$

can be written

$$\phi_m(t)=\prod_{j=0}^{\min(b,m)}(1+\alpha_j t),\quad 0<\alpha_j\leq 1.$$

The equation (2) can be written as

$$\phi_{n-1}(t)=\left(1-\frac{t}{n}D\right)\phi_n(t),$$

where $D\phi_n(t)=\phi_n'(t)$. From Theorem 1 it follows that if

$$\phi_n(t)=\prod_{j=1}^{n}(1+p_j t),\quad 0<p_j\leq 1,$$

then the operator $\left(1-\frac{t}{n}D\right)$ preserves the *ISR* property in the sense that $\phi_{n-1}(t)$ can be written $\phi_{n-1}(t)=\prod_{j=1}^{n-1}(1+\alpha_j t),\ 0<\alpha_j\leq 1.$

In the above exmple, let $P_{n,k}=P_r[X_1+X_2+\cdots+X_n=k]$, and let $\phi_n(t)=\sum_{k=0}^{n}(1+t)^k P_{n,k}$, then the following forward difference differential equation for $\phi_n(t)$ can be readily obtained

$$\phi_{n+1}(t)=\left(1+\frac{b}{r+b-n}t\right)\phi_n(t)-\frac{t(t+1)}{r+b-n}\phi_n'(t) \tag{5}$$

Note that, using the equation (5), the proof of the *ISR* property cannot be easily provided. However, the following theorem asserts that the equations (2) and (5) are equivalent.

THEOREM 2. *Let $X_1, X_2, \cdots, X_n$ be a sequence of exchangeable random variables and let $\phi_n(t)$ denote the f. m. g. function of the sum $(X_1+X_2+\cdots+X_n)$, then there exist constants A_n and B_n which are independent of t such that the forward difference differential equation is given by*

$$\phi_{n+1}(t)=(1+A_n t)\phi_n(t)-B_n t(t+1)\phi_n'(t),\quad n\geq 2. \tag{6}$$

Proof. It can be easily shown that using (2) and (5), A_n and B_n can be determined by the following system of equations:

$$nA_n - n\phi_n'(0) B_n = \phi_n'(0)$$
$$(n-1)\phi_n'(0)A_n - (n-1)(\phi_n'(0)+\phi_n''(0))B_n = \phi_n''(0)$$

This system of equations has a unique solution A_n and B_n, hence Theorem 2 follows.

REMARKS. 1. The hypergeometric example has the forward difference differential equation of the form of equation (6) with $A_n = b/(r+b-n)$ and $B_n = 1/(r+b-n)$.

2. Suppose the sum of Bernoulli random variables has the difference differential equation of the form (6) given in Theorem 2. It is not generally true that the sequence of Bernoulli random variables is exchangeable.

3. Consider an *urn* which has b black and r red balls. We draw a ball at random and replace it together with s balls of the same color. Repeat the procedure n times and let $X_i=1$ (0) if the ith ball drawn is a black (red) ball. Then $X_1, X_2, \cdots, X_n$ are exchangeable Bernoulli random variables and they have the *ISR* property if $s=-1$ (Hypergeometric) and $s=0$ (Binomial). For $s \geq 1$, they do not, in general, enjoy the *ISR* property.

Now we consider an occupancy problem. Assum that m balls are randomly distributed into N equi-probable cells. $X_{m,i}=1$ (0) if the i-th cell is empty (occupied). Then for each m, $X_{m,1}, X_{m,2}, \cdots, X_{m,N}$ are exchangeable Bernoulli random variables. Let

$$Q_{m,k} = P_r[X_{m,1}+X_{m,2}+\cdots+X_{m,N}=k] \quad \text{and} \quad \phi_m(t) = \sum_{k=0}^{N-1}(1+t)^k Q_{m,k}$$

then the forward difference differential equation for $\phi_m(t)$ can be written

$$\phi_{m+1}(t) = \phi_m(t) - \frac{t}{N}\phi_m'(t). \tag{7}$$

Notice that $\phi_m(t)$ is a polynomial in t of degree $N-1$. Using the similar argument used in the proof of Theorem 2, the following theorem can be readily proved. See Harris and Park [1971].

THEOREM 3. *If the sum* $S_m = X_{m,1}+X_{m,2}+\cdots+N_{m,N}$ *satisfies the ISR property for some* m, *then* $X_{n,1}+X_{n,2}+\cdots+X_{n,N}$ *for* $n \geq m$ *has the ISR property. Notice that for the occupancy problem we have* $\phi_1(t)=(1+t)^{N-1}$, *which says that* $X_{1,1}+X_{1,2}+\cdots+X_{1,N}$ *has the ISR property, thus for* $n \geq 1$, $X_{n,1}+X_{n,2}+\cdots+X_{n,N}$ *has the ISR property.*

The backward difference differential equation for $\phi_m(t)$ can be be written as

$$(8)\qquad \phi_m(t)=\left(1+\frac{t}{N-1}D+\frac{t^2}{(N-1)(N-2)}D^2+\cdots+\frac{t^{N-1}}{(N-1)!}D^{N-1}\right)\phi_{m+1}(t).$$

From (7) we can say that the operator $\left(1-\frac{t}{N}D\right)$ preserves the *ISR* property, that is, if $\phi_m(t)$ can be written as $\prod_{j=1}^{N-1}(1+p_jt)$, $0<p_j\leq 1$, then $\phi_{m+1}(t)$ can be also written as $\prod_{j=1}^{N-1}(1+\alpha_jt)$, $0<\alpha_j\leq 1$. In the same sense, can it be said that the operator

$$\left(1+\frac{t}{N-1}D+\frac{t^2}{(N-1)(N-2)}D^2+\cdots+\frac{t^{N-1}}{(N-1)!}D^{N-1}\right)$$

preserves the *ISR* property? We leave this problem to the reader.

References

1. De Finetti, B., *Funzione caratteristica di un fenomeno aleatorio,* Atti Accad. Naz. Lincei, Rend. Cl. Fis. Mat. Nat., (6)**4**(1930), 86–133.
2. Feller, W., *Probability Theory and Its Applications,* **11**(1971), 2nd ed., John Wiley & Sons, Inc., New York.
3. Harris, B. and Park, C. J., *A note on the asymptotic normality of the distribution of the number of empty cells in occupancy problems,* Ann. Inst. Stat. Math., **23** (3) (1971), 507–513.
4. Johnson, N. L. and Kotz, S., *Urn Models and Their Application,* John Wiley & Sons, Inc., New York (1977).
5. Park, C. J., *A note on the asymptotic normality of the distribution of the number of empty celis under Bose-Einstein Statistics,* Journal of K. M. S., **14**(1976), 217–221.
6. Park, C. J., *A Note on the Hypergeometric distribution,* IMS Bulletin, **6**(1977), 104.

San Diego State University and Seoul National University