The distribution of a linear combination of r independent discrete random variables

Rahim Mahmoudvand [*]
Statistics Group
University of Payame Noor
Toyserkan, Iran

Hossein Hassani [†]
Centre for Optimisation and Its Applications
School of Mathematics, Cardiff University
CF24 4AG, UK
and
Central Bank of the Islamic Republic of Iran
Iran

Abbas Farzaneh
Group of Computer
Islamic Azad University
Toyserkan, Iran

Gareth Howell [‡]
Statistics Group
Cardiff University
UK

Abstract

In this paper we consider a formula to get the exact number of nonnegative integer solution of the equality $a_1x_1 + a_2x_2 + \ldots + a_rx_r = n$ where $a_1, a_2, \ldots, a_r$ and n are fixed integers. Using the obtained formula, we provide a program to list the solutions for every n and $a_1, \ldots, a_r$ by Pascal compiler. We then obtain the distribution of an arbitrary linear combination of discrete random variables based on the proposed algorithm. We also apply the algorithm to obtain the Maximum Likelihood estimation of the parameters of the distribution. The accuracy of the algorithm was illustrated using various examples.

Keywords and phrases : *Nonnegative integer solutions, linear combination, maximum likelihood estimation, discrete random variables.*

[*] *E-mail*: `r_mahmodvand@yahoo.com`

[†] *E-mail*: `HasaniH@cf.ac.uk`

[‡] *E-mail*: `HowellGL@cf.ac.uk`

Journal of Interdisciplinary Mathematics
Vol. ( ), No. , pp. 1–8

1. Introduction

Appearing in many elementary texts in probability, counting techniques play an important role in computing probabilities in random experiments such as throwing dice, or classical occupancy problems. As a result, they have come to form a major part of the mathematics curriculum in many statistical backgrounds. Example of such literature are seen in Ross (1976), and Rosen *et al.* (2000) and so forth.

An interesting problem of counting methods is the number of ways for placing n identical objects into r distinct cells; this is equivalent to the number of nonnegative integer solutions of the following equation,

$$x_1 + x_2 + \ldots + x_r = n. \tag{1}$$

A generalization of this problem is to find the number of nonnegative integer solutions of

$$a_1x_1 + a_2x_2 + \ldots + a_rx_r = n, \tag{2}$$

where, $a_1, \ldots, a_r$ and n are integer.

Equation (2) is well-known as a Linear Diophantine Equation, for which the problem of finding bounds on the number of nonnegative solutions is well studied. Mahmoudvand *et al.* [4] reviewed these and presented a new simple method for finding the number of nonnegative integer solutions of (2) and providing a list of them. The following formula for the solutions of (2) is the subject of their note:

$$\begin{aligned} & s(a_1, \ldots, a_r; n) \\ & := \sum_{w_1=0}^{[n/a_1]} \sum_{w_2=0}^{[(n-a_1w_1)/a_2]} \cdots \sum_{w_{r-1}=0}^{[(n-a_1w_1-\ldots-a_{r-2}w_{r-2})/a_{r-1}]} I(a_r; w_1, \ldots, w_{r-1}), \end{aligned} \tag{3}$$

where $I(a_r; w_1, \ldots, w_{r-1}) = \begin{cases} 1 & \text{if } a_r | n - a_1w_1 - \ldots - a_{r-1}w_{r-1} \\ 0 & \text{otherwise.} \end{cases}$

Proof. Let us first consider $a_i = 1$ for $i = 2, \ldots, r$ in (2). In this case, we must find the number of nonnegative integer solutions for

$$a_1x_1 + x_2 + \ldots + x_r = n. \tag{4}$$

For solving (4), we can give the possible values of x_1 and reform (4) to form (1). Therefore,

$$\sum_{w_1=0}^{[n/a_1]} \binom{n - a_1w_1 + r - 2}{r - 2} \tag{5}$$

is the number of nonnegative integer solutions for equation (4), where $[u]$ is the integer part of u and $r > 2$ is a positive integer. If $r = 2$ we must use $\sum_{w_1=0}^{[n/a_1]} I(a_2, w_1)$ as the number of nonnegative integer solutions, where

$$I(a_2, w_1) = \begin{cases} 1 & a_2 | n - a_1 w_1 \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Now, let $a_i = 1$ for $i = 3, \dots, r$. In this case, we must find the number of nonnegative integer solutions for

$$a_1 x_1 + a_2 x_2 + x_3 + \dots + x_r = n. \tag{7}$$

For solving (7), we can give the possible values of x_1, x_2 and reform (7) to form (1). Therefore,

$$\sum_{w_1=0}^{[n/a_1]} \sum_{w_2=0}^{[(n-a_1w_1)/a_2]} \binom{n - a_1 w_1 - a_2 w_2 + r - 3}{r - 3} \tag{8}$$

is the number of nonnegative integer solutions for this equation, where $r > 3$ is a positive integer. However, if $r = 3$ we use $\sum_{w_1=0}^{[n/a_1]} \sum_{w_1=0}^{[(n-a_1w_1)/a_2]} I(a_3, w_1, w_2)$ as the number of nonnegative integer solutions, where

$$I(a_3, w_1, w_2) = \begin{cases} 1 & a_3 | n - a_1 w_1 - a_2 w_2 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Continuing the procedure, we can get the following formula for the number of nonnegative integer solutions of (2):

$$\begin{aligned} &s(a_1, \dots, a_r; n) \\ &:= \sum_{w_1=0}^{[n/a_1]} \sum_{w_2=0}^{[(n-a_1w_1)/a_2]} \cdots \sum_{w_{r-1}=0}^{[(n-a_1w_1-\dots-a_{r-2}w_{r-2})/a_{r-1}]} I(a_r; w_1, \dots, w_{r-1}) \end{aligned} \tag{10}$$

where

$$I(a_r; w_1, \dots, w_{r-1}) = \begin{cases} 1 & a_r | n - a_1 w_1 - \dots - a_{r-1} w_{r-1} \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Note also that if $a_i = 1$ for all i, then $s(a_1, \dots, a_r; n)$ is equal to $\binom{n+r-1}{r-1}$, since

$$s(a_1, \dots, a_r; n) = \sum_{w_1=0}^{n} \sum_{w_2=0}^{n-w_1} \cdots \sum_{w_{r-1}=0}^{n-w_1-\dots-w_{r-2}} 1$$

$$= \sum_{w_1=0}^{n} \sum_{w_2=0}^{n-w_1} \cdots \sum_{w_{r-2}=0}^{n-w_1-\ldots-w_{r-3}} \binom{n-w_1-\ldots-w_{r-2}+1}{1}$$

$$= \sum_{w_1=0}^{n} \sum_{w_2=0}^{n-w_1} \cdots \sum_{w_{r-2}=0}^{n+1-w_1-\ldots-w_{r-3}-1} \binom{1+w_{r-2}}{1}.$$

Now equality is obtained using the fact that $\sum_{k=0}^{n-m} \binom{m+k}{m} = \binom{n+1}{m+1}$. □

It has been shown that the number solutions of (2) with some constraints placed on the x_i's can be expressed as a function of the number solutions of (2) without any bounds on x_i's (for details, see Eisenbeis *et al.* [1]). As an example, suppose that we desire to determine the number of positive integer solutions of (2); letting $x_i = z_i + 1$ for each i yields

$$a_1z_1 + a_2z_2 + \ldots + a_rz_r = n - a_1 - \ldots - a_r\,, \tag{12}$$

to be solved in nonnegative integers. Therefore using (12) the number of positive integer solutions of (2) is $s\left(n - \sum_{j=1}^{r} a_j, a_1, \ldots, a_r\right)$.

2. An application

2.1 *Distribution*

There are many problems which can be solved using the proposed algorithm. As a useful example, we use the algorithm to obtain the distribution of a linear combination of r independent discrete random variables.

Let $X_1, \ldots, X_r$ be discrete random variables from a discrete distribution. Suppose $a_1, \ldots, a_r$ are arbitrary fixed integer values. One problem that of some interest is in finding the distribution of the random variable $Y = a_1X_1 + \ldots + a_rX_r$. Let the support of variable Y, the set of all possible values that Y can assume, be denoted by S_Y. Moreover, let the support of X also be denoted by $S_X = \{0, 1, \ldots\}$. Using the method proposed by Mahmoudvand *et al.* in [4], it is easy to write:

$$\begin{aligned} f_Y(n) &= P[a_1X_1 + \ldots + a_rX_r = n] \\ &= \sum_{i=1}^{m} P[X_1 = x_{i_1}, \ldots, X_r = x_{i_r}], \quad n \in S_Y \end{aligned} \tag{13}$$

where $m = s(a_1, \ldots, a_r; n)$ and the $(x_{i_1}, \ldots, x_{i_r})$ are vectors of the solutions for equation (2). If the X_i's are independent, then equation (13) is simplified to:

$$f_Y(n) = \sum_{i=1}^{m} \prod_{j=1}^{r} P[X_j = x_{i_j}], \quad n \in S_Y \tag{14}$$

Let us study an example. Consider the independent random variables $X_1, \ldots, X_5$, sampled from a Poisson distribution with mean 1. For these X_i one may use the formula proposed in [4] to calculate the distribution of the linear combination

$$Y = 3X_1 + 7X_2 + 5X_3 + 4X_4 + 2X_5 \,.$$

Using formula (14) one therefore has

$$f_Y(n) = \sum_{i=1}^{m} \prod_{j=1}^{5} \frac{e^{-1}}{x_{i_j}!} = \sum_{i=1}^{m} \frac{e^{-5}}{\prod_{j=1}^{5} x_{i_j}!}, \quad n = 0, 2, 3, 4, \ldots, \tag{15}$$

where m and the x_{i_j}'s are defined above.

Table 1 shows the results of numerical calculations of such probabilities for various n in this problem, based on (14). To evaluate the importance of these calculations we provide the normal approximation of these probabilities also.

Table 1

Probability distribution of Y by exact and Normal Approximation

n	0	1	2	3	4	5	6	7	8	9	10
Exact	0.007	0.000	0.007	0.007	0.010	0.013	0.011	0.024	0.017	0.026	0.026
Normal approximation	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.003	0.004	0.004	0.005

As it seen from Table 1 there are meaningful differences between the exact and approximated probabilities. As an illustration of this fact, we plot the *Cumulative Distribution Function* (CDF) of Y with Exact and Normal Approximation in Figure 1.

2.2 *Maximum Likelihood estimation*

We may also apply formula (3) to obtain the Maximum Likelihood estimation of the parameters of the distribution. Consider the linear combination Y from above and Let $X_1, \ldots, X_r$ be a random independent

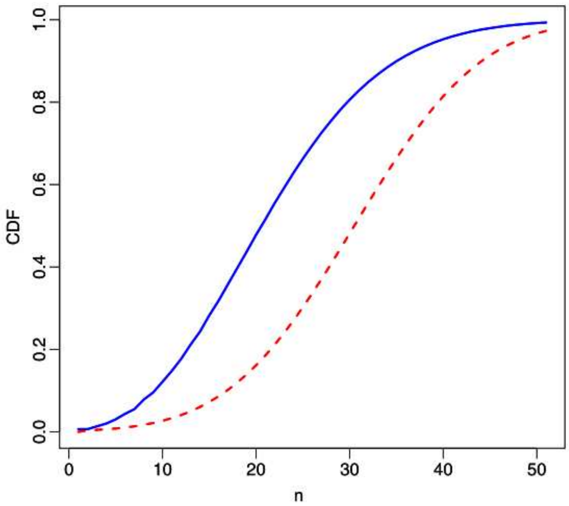

Figure 1

Cumulative Distribution Function of Y with the exact (solid line) and normal approximation (dash line) methods

sample from Poisson distribution with mean λ. Adopt the new convention $Y = n$; we wish to obtain a *Maximum Likelihood Estimation* (MLE) of λ. The likelihood function is as follows:

$$L(\lambda, n) = \begin{cases} e^{-r\lambda} & n = 0 \\ e^{-r\lambda}\lambda^{d} & n = 1 \\ \sum_{i=1}^{m}\prod_{j=1}^{r}\frac{e^{-\lambda}\lambda^{x_{i_j}}}{x_{i_j}!} & \text{otherwise,} \end{cases}$$

where d is the number of coefficients a_i equal to 1. The MLE of λ is thus:

$$\hat{\lambda}_{MLE} = \begin{cases} 0 & n = 0 \\ \frac{d}{r} & n = 1 \\ \lambda_0 & \text{otherwise,} \end{cases}$$

where λ_0 is a solution of the following equality:

$$\sum_{i=1}^{m}\frac{\lambda^{x_{i.}-1}[x_{i.} - r\lambda]}{\prod_{j=1}^{r} x_{i_j}} = 0, \tag{16}$$

and $x_{i.} = \sum_{j=1}^{r} x_{i_j}$.

3. Computation using pascal code

Here we provide a program by Pascal compiler to obtain the solutions by using the formula of Mahmoudvand *et al.* (2009). Consider the following procedure composed of $r-1$ nested **for** loops:

count := 0
  for w_1 := 0 **to** [n / a_1]
   for w_2 := 0 **to** [($n-$ w_1a_1)/ a_2]
    for w_3 := 0 **to** [($n-$ w_1a_1- w_2a_2)/ a_3]
     .
     .
     .
     for w_{r-1} := 0 **to** [($n-w_1a_1-\ldots-w_{r-2}a_{r-2}$)/ a_{r-1}]
      if $a_r|n-$ $w_1a_1-\ldots-w_{r-1}a_{r-1}$ **then**
       count:= **count**+ 1

The value of count upon exit from this procedure is thus $s(a_1,\ldots,a_r,n)$. we provide a program that may be applied for every r .

```
type
  dlist =^ node;
   node = record
   befor:dlist;
   fact:integer;
   high:integer;
   index:integer;
   next:dlist;
end;
var
   list,l,first,last:dlist;
   i,x,n,s,ar,k:integer;
{− − − − − − − −}
procedure insertlist(var list:dlist;x:integer);
var
  temp,l:dlist;
begin
   new(temp);
   temp ˆ .befor:=nil;
   temp ˆ .high:=n div x;
   temp ˆ .index:=0;
   temp ˆ .next:=nil;
if list=nil then
   list:=temp
else
begin
   l:=list;
   while(l ˆ .next <> nil) do
    l: =ˆ .next;
    l ˆ .next:=temp;
    temp ˆ .befor:=l;
   end;
end;
{− − − − − − −}
function test1:boolean;
var l:dlist;
begin
   test1:=true;
   l:=list;
while(l <> nil) do
begin
   if l ˆ .index > l ˆ .high then
   test1:=false;
   l:=l ˆ .next;
   end;
end;
{− − − − − − − −}
procedure print;
var l:dlist;
s:integer;
begin
   l:=list;
   s:=0;
while(l <> nil) do
begin
   write(l ˆ .index, '');
   s:=s+l ˆ .fact*l ˆ .index;
   l:=l ˆ.next;
end;
if((n-s) mod ar)=0 then
begin
   write((n-s)div ar);
   writln;
end
else
begin
   write( #13 );
   write(' ');
   write( #13 );
   end;
end;
{− − − − − − −}
begin
   list:=nil;
   i:=1;
   write('enter the number of terms:');
   readln(k);
   write('enter the value of n:');
   readln(n);
for i:=1 to k-1 do
begin
   write('enter a',i,':');    readln(x);
```

```
insertlist(list,x); end;   write('enter
  a',k,':');
  readln(ar)
if list=nil then
  exit;
  l:=list;
  while(l ^ .next <> nil) do
  l:=l ^ .next;
  last:=l;
  first:=list;
  writeln;
while(first ^ .index <= first ^
   .high)do
begin
if test1 then
begin
  print;
  last ^ .index:=last ^ .index+1;
end;
  l:=list ^ .next;
while(l ^ .index <= l ^ .high) and
   (l <> nil) do
  l:=l ^ .next;
if(l <> nil) then
begin
  l:=l ^ .befor;
  l ^ .index:=l ^ .index+1;
if(l ^ .index <= l ^ .high) then
begin
  l:=l ^ .next;
while(l <> nil) do
begin
  l ^ .index=0;
  l:=l ^ .next;
end;
  l:=list;
  s:=0;
while(l <> nil) do
begin
  l ^ .high:=(n-s) div l ^ .fact;
  s:=s+l ^ .fact*l ^ .index;
  l:=l ^ .next;
  end;
  end;
  end;
end;
  readln;
end.
```

References

[1] C. Eisenbeis, O. Temam and H. Wijsho, On efficiently characterizing solutions of linear diophantine equations and its application to data dependence analysis, in *7th International Symposium on Computer and Information Sciences*, Antalya, Turquie, 1992; appeared also as INRIA Research Report No. 1616, Fvrier 1992.

[2] H. K. Rosen, G. J. Michaels, L. J. Gross, W. J. Grossman and R. D. Shier, *Handbook of Discrete and Combinatorial Mathematics*, CRC Press, New York, 2000.

[3] S. M. Ross, *A First Course in Probability*, CRC Press, New York, 1976.

[4] R. Mahmoudvand, H. Hassani, A. Farzaneh and G. Howell (2009), The exact number of nonnegative integer solutions for the equality $a_1x_1 + \ldots + a_rx_r \leq n$, *IAENG Journal of Applied Mathematics*, forthcoming.

Received January, 2009