MINIMUM VARIANCE UNBIASED ESTIMATION AND CERTAIN PROBLEMS OF ADDITIVE NUMBER THEORY[1]

By G. P. Patil

McGill University

0. Introduction and summary. Let T be a subset of the set I of non-negative integers. Define $f(\theta) = \sum a(x)\theta^x$ where the summation extends over T and $a(x) > 0$, $\theta \geqq 0$ with $\theta \varepsilon \Theta$, the parameter space, such that $f(\theta)$ is finite and differentiable. One has $\Theta = \{\theta : 0 \leqq \theta < R\}$ where R is the radius of convergence of the power series of $f(\theta)$. Then a random variable X with probability function

$$\text{Prob}\{X = x\} = p(x, \theta) = a(x)\theta^x/f(\theta) \qquad x \varepsilon T \tag{1}$$

is said to have the generalized power series distribution (GPSD) with range T and the series function $f(\theta)$. It may be observed that the range T could be a countable subset of the real numbers. The GPSD is thus an exponential-type discrete distribution; whereas the power series distribution (PSD) as defined by Noack [5], is a special case of a GPSD. The author [7], [8], [9], [10], [11] has discussed some problems of statistical inference associated with the GPSD and some of its particular forms. Roy and Mitra [14] and Guttman [2] have studied the problem of the minimum variance unbiased (MVU) estimation for the PSD's.

In this paper, we investigate the problem of the existence of the MVU estimator for the parameter θ of the GPSD in terms of the number theoretic structure of its range T. We further provide the MVU estimators for the probability and distribution functions of the GPSD and consider a few special cases of some practical significance.

1. Notation and terminology.

(i) Let $A = \{0, a_1, a_2, \cdots\}$ be a subset of the set I of the non-negative integers. The Schnirelmann density $d(A)$ of the set A is defined by

$$d(A) = \text{g.l.b. } A(m)/m \tag{2}$$

where $A(m)$ denotes the number of positive integers in the set A which do not exceed m. It is easy to see that $0 \leqq A(m) \leqq m$ and therefore $0 \leqq d(A) \leqq 1$. Note that the Schnirelmann density is defined only for such sets in I which contain zero.

(ii) Let $A^{(i)} = \{a_1^{(i)}, a_2^{(i)}, \cdots\}$, $i = 1, 2, \cdots n$ be arbitrary subsets of I. The sum $A_n = \sum_{i=1}^n A^{(i)}$ of the n given subsets is defined as the set of all integers of the form $\sum_{i=1}^n a^{(i)}$ where $a^{(i)} \varepsilon A^{(i)}$. If $A^{(i)} = A$, $A_n = \sum A$ is denoted

Received August 10, 1962; revised March 7, 1963.

[1] Research carried out while the author was a fellow of the 1962 Summer Research Institute of the Canadian Mathematical Congress.

by $n[A]$. The difference $A^{(1)} - A^{(2)}$ is defined as the set of all integers of the form $a^{(1)} - a^{(2)}$ where $a^{(i)} \varepsilon A^{(i)}$.

(iii) The set A is called a basis of the set I if $n[A] = I$ for some n. In such a case n is called the order of the basis A.

(iv) Set $A = \{a\}$ denotes the singleton, the set of only one member, namely a.

(v) A subset T of I is said to be the tail-set of a if $T = \{x : x \geqq a\}$ where $a \varepsilon I$ and is denoted by S_a. Clearly, $S_0 = I$.

(vi) A subset T of I is said to be the index-set of the function $f(\theta)$ if $f(\theta) = \sum a(x)\theta^x$ with $a(x) \neq 0$, $x \varepsilon T$ and is denoted by

$$T = W[f(\theta)]. \tag{3}$$

Clearly, the range of a GPSD is also the index-set of its series function.

(vii) A real-valued parameter $g(\theta)$ is called MVU estimable if it has minimum variance unbiased estimator based on a random sample of some size n.

2. The MVU estimation of θ. For the PSD's whose parameter space contains zero (one can easily deduce for such PSD's that their range also contains zero) Guttman [2] has given a necessary and sufficient condition for a function of the parameter θ to be MVU estimable. He has obtained the condition in terms of some relationship between the power series expansions involved. Here we shall obtain conditions in terms of the structure of the range T of a GPSD.

THEOREM 1. *A necessary and sufficient condition for the parameter θ of the* GPSD *given by* (1) *to be* MVU *estimable is that for some n,*

$$n[T] = S_{\min(n[T])}. \tag{4}$$

PROOF. We need the following lemma.

LEMMA. *A necessary and sufficient condition for the parameter θ of the* GPSD *given by* (1) *to be* MVU *estimable on the basis of a single observation is that $T = S_{\min(T)}$ or equivalently $T + \{1\} \subseteq T$.*

PROOF OF THE LEMMA. Since x is complete sufficient for θ, it follows that an unbiased estimator $h(x)$, if any, of θ is the MVU estimator for θ. Conversely, if θ is not MVU estimable on the basis of a single observation, θ does not have any unbiased estimator either.

To prove the necessity, suppose $h(x)$ denotes an unbiased estimator for θ. One has therefore for all $\theta \varepsilon \Theta$ the identity $\sum h(x)a(x)\theta^x = \sum' a(x-1)\theta^x$ where $\sum$ extends over $x \varepsilon T$ and $\sum'$ extends over $x \varepsilon T + \{1\}$. It follows that $h(x) = a(x-1)/a(x)$ for $x \varepsilon T + \{1\}$ and $=0$ otherwise and further $T + \{1\} \subseteq T$, because, otherwise, it contradicts the fact that $a(x) > 0$ for $x \varepsilon T$. The sufficiency follows if one considers $h(x)$ as defined above to be an estimator for θ.

To prove the theorem, one has only to see that $z = \sum_{i=1}^n x_i$, where $x_1, x_2 \cdots x_n$ is a random sample of size n drawn from the GPSD given by (1), follows a GPSD with range $n[T]$ and the series function

$$f_n(\theta) = [f(\theta)]^n = \sum b(z, n)\theta^z \tag{7}$$

where the summation extends over $n[T]$ and $b(z, n)$ is the coefficient of θ^z in the expansion of $f_n(\theta)$. Clearly, $b(z, n) > 0$ for $z \varepsilon n[T]$. Also z is complete sufficient for θ. Because of the lemma, the theorem now follows.

COROLLARY 1. *θ is not* MVU *estimable if T is finite.*

This follows from the finiteness of $n[T]$ inherited from T. To cite a few applications one may note that θ is MVU estimable for every sample size n for complete or left-truncated Poisson and logarithmic series distributions. It is interesting to observe the statements of Tate and Goen [16], Patil [6], [7], [12] and Roy and Mitra [14] to this effect. One may note further that the right-truncated Poisson and logarithmic series distributions do not have unbiased estimators for θ. Tate and Goen [16] have observed this result for the right-truncated Poisson distribution.

COROLLARY 2. *When it exists, the* MVU *estimator of θ is given by*

$$
\begin{aligned}
h(z, n) &= b(z-1, n)/b(z, n) && z \varepsilon n[T] + \{1\}. \\
&= 0 && \textit{otherwise}
\end{aligned}
\tag{8}
$$

where symbols carry usual sense.

3. Some comments on the MVU estimation of θ. Theorem 1 brings out that the MVU estimability of θ depends only on the structure of the range T of the GPSD and it is curious to note that it has nothing to do with the specific form of the GPSD as determined by the coefficients $a(x)$. The natural and fruitful questions to ask then are:

(i) What should be the structure of T so that Equation (4) holds for some n? The statistical basis for this question is to know if a criterion on T can be devised to examine the MVU estimability of θ for a given GPSD.

(ii) Given T, how can one find the value of n (if it exists) so that Equation (4) holds? The statistical basis for this question is to find a sample-size which guarantees the MVU estimator for θ of a GPSD in case it is shown to be MVU estimable.

(iii) If there are several solutions of the Equation (4) for a fixed T, how can one find the smallest solution? The statistical basis for this question is to know the smallest sample size which guarantees the MVU estimator of θ for a given GPSD.

Curiously enough these three questions that are pivotal for our purposes have been of great interest to the specialists of additive number theory. Question (i) has been solved by Schnirelmann [15], [4] for such sets which contain zero whereas questions (ii) and (iii) do not seem to have been solved in any generality.

4. A characterization of the MVU estimability of θ in terms of the range T. We have the following:

THEOREM 2. *A necessary and sufficient condition for the parameter θ of the* GPSD *given by* (1) *to be* MVU *estimable is that the set $T - \{\min(T)\}$ be a basis of I.*

PROOF. Let $Y = X - \min(T)$ where X follows the GPSD given by (1). It can be verified that Y follows a GPSD with range $T - \{\min(T)\}$ and the series function $f(\theta)/\theta^{\min(T)}$. Therefore, based on a random sample $x_1, x_2, \cdots, x_n$ of size n drawn from the GPSD given by (1), or equivalently, based on a random sample $y_1, y_2, \cdots, y_n$ drawn from the GPSD of Y where $y_i = x_i - \min(T)$, it follows from Theorem 1 that θ is MVU estimable if and only if $n[T - \{\min(T)\}] = S_0 = I$ because $\min(n[T - \{\min(T)\}]) = 0$. The statement of the theorem follows from the definition of the basis as given in (iii) of Section 1.

COROLLARY 3. *If θ is* MVU *estimable for a sample of size n, it is* MVU *estimable for every sample size exceeding n.*

PROOF. The corollary follows from Theorem 2 and an elementary fact that a basis of order n is also a basis of order $n + 1$.

THEOREM 3. *The necessary and sufficient condition for the parameter θ of the* GPSD *given by* (1) *to be* MVU *estimable is that the Schnirelmann density of the set $n[T - \{\min(T)\}]$ be unity for some n.*

PROOF. This theorem follows directly from Theorem 2 and a theorem of the additive number theory which states that a necessary and sufficient condition for a set A of non-negative integers to be identical with I is that $d(A) = 1$, where d stands for the Schnirelmann density as defined in (i) of Section 1.

COROLLARY 4. *A sufficient condition for the parameter θ to be* MVU *estimable is that the Schnirelmann density of the set $T - \{\min(T)\}$ be positive.*

PROOF. The corollary follows by applying Schnirelmann's theorem to Theorem 2 where Schnirelmann's theorem states that every set of positive density is a basis of I.

That the sufficient condition of the corollary is not necessary can be seen from the following example. Consider a GPSD with range $S = \{x^2 : x \varepsilon I\}$. One can easily verify that $d(S) = 0$. On the other hand, Lagrange's theorem, claiming every non-negative integer to be the sum of squares of at most four non-negative integers, insures that for $n = 4$, $n[S - \{\min(S)\}] = I$ and therefore from Theorem 2, θ is MVU estimable.

5. The MVU estimation of an arbitrary function of θ. Let $g(\theta)$ be a given function of θ which is such that $g(\theta) \cdot f_n(\theta)$ admits a power series expansion in θ where $f_n(\theta)$ is defined by (7). Proceeding on the same lines as in Section 2, one has the following:

THEOREM 4. *The necessary and sufficient condition for $g(\theta)$ to be* MVU *estimable on the basis of a random sample of size n from the* GPSD *given by* (1) *is that $W[g(\theta) \cdot f_n(\theta)] \subseteq W[f_n(\theta)]$ where W is defined by* (3). *Also, whenever it exists, the* MVU *estimator for $g(\theta)$ is given by $h(z, n) = c(z, n)/b(z, n)$ for $z \varepsilon W[g(\theta) \cdot f_n(\theta)]$ and $=0$, otherwise, where $c(z, n)$ is the coefficient of θ^z in the expansion of $g(\theta) \cdot f_n(\theta)$ and z and $b(z, n)$ are defined as in Section* 2.

One may observe that the variance of $h(z, n)$ is given by $V\{h(z, n)\} = E(\{h(z, n)\}^2) - \{g(\theta)\}^2$. Writing $g_2(\theta) = \{g(\theta)\}^2$, the MVU estimator of $g_2(\theta)$

to be denoted by $h_2(z, n)$ can be written down on the usual lines whenever it exists.

It is clear that, because z is complete sufficient for θ, the MVU estimator of the variance of the MVU estimator of $g(\theta)$ is given by $\{h(z, n)\}^2 - h_2(z, n)$ which, when $g(\theta) = \theta$, reduces to $h(z, n)[h(z, n) - h(z - 1, n)]$, where $h(z, n)$ is defined by (8). The advantage of this form, first obtained by Roy and Mitra [14], is obvious.

It may be of some interest here to record that $g(\theta) = 1/f(\theta)$ is MVU estimable if and only if the range T of a GPSD contains zero, whereas $g(\theta) = 1/\theta$ is not MVU estimable (and hence in the case of a GPSD, as it is here, it does not have an unbiased estimator either) of any GPSD. From this follows that the parameter $\alpha = 1/f(\theta)$ where $f(\theta) = -\log(1 - \theta)$ of the logarithmic series distribution does not have an unbiased estimator. Also one may note the result of Quenouille as stated on page 5 and 34 of [13] as a special case which states that the reciprocal of the parameter of the Poisson distribution does not have an unbiased estimator.

It is easy to see that, on the basis of an observation from the binomial distribution with parameters $p = \theta/(1 + \theta)$ and n, a polynomial in p is MVU estimable if and only if the degree of the polynomial does not exceed n. For the negative binomial distribution, however, $p = 1 - \theta$ and a polynomial of every degree in p is MVU estimable.

6. The MVU estimation of the probability and distribution functions of a GPSD. Let $g(\theta) = p(k, \theta)$, where $k \varepsilon T$ is known and $p(x, \theta)$ is defined by (1). One has $W[g(\theta) \cdot f_n(\theta)] = (n - 1)[T] + \{k\}$, whereas $W[f_n(\theta)] = n(T)$. Clearly the conditions of Theorem 4 are satisfied. Thus the probability function of a GPSD has a MVU estimator for every sample size n and it can be verified that it is given by $h(z, n; k) = c(z, n; k)/b(z, n)$ for $z \varepsilon (n - 1)[T] + \{k\}$ and $=0$, otherwise, where $c(z, n; k) = a(k)b(z - k, n - 1)$.

Further, the MVU estimator of the variance of the MVU estimator under consideration always exists and is given by $h(z, n; k)\{h(z, n; k) - h(z - k, n - 1; k)\}$. The advantage of this form may be obvious.

In order to estimate the distribution function of the GPSD given by (1), let $g(\theta) = F(r, \theta) = \sum p(k, \theta)$ where the summation extends over $T_r = \{x : x \leqq r, x \varepsilon T\}$. Note that T_r is finite. It is easy to see that $F(r, \theta)$ has a MVU estimator for every sample size and it is given by $\sum_{k \varepsilon T_r} h(z, n; k)$ where $h(z, n; k)$ is as defined in this section, because z is complete sufficient for θ and hence the MVU estimator of a sum of a finite number of MVU estimable functions of θ is given by the sum of the individual MVU estimators.

In general, one may note that $g(\theta) = m(\theta)/f(\theta)$ where $W[m(\theta)] \subset T = W[f(\theta)]$ is MVU estimable for every sample size and for every GPSD with range T. Its MVU estimator can be easily written down on the usual lines.

One may record below a few applications of the results of this section to the following GPSD's of some practical importance.

(i) *Poisson distribution:* One has $a(k) = 1/k!$, $k = 0, 1, 2, \cdots$; $b(z, n) =$

$n^z/z!$. Therefore the MVU estimator for $p(k, \theta) = e^{-\theta}\theta^k/k!$ can be obtained as a binomial probability

$$h(z; n; k) = \binom{z}{k}\left(1 - \frac{1}{n}\right)^{z-k}\left(\frac{1}{n}\right)^k \qquad \text{for } z \geqq k$$
$$= 0, \qquad \text{otherwise}$$

(ii) *Zero-truncated Poisson distribution:* One has $a(k) = 1/k!$, $k = 1, 2, \cdots$; $b(z, n) = (n!/z!)\mathcal{S}_z^n$ where $\mathcal{S}_z^n$ is the Stirling number of the second kind with arguments n and z. Therefore the MVU estimator for $p(k, \theta) = e^{-\theta}\theta^k/k![1 - e^{-\theta}]$ can be obtained as

$$h(z, n; k) = \binom{z}{k}\mathcal{S}_{z-k}^{n-1}/n\mathcal{S}_z^n.$$

(iii) *Binomial distribution:* One has

$$a(k) = \binom{m}{k}, \qquad k = 0, 1, 2 \cdots m; \qquad b(z, n) = \binom{mn}{z}.$$

Therefore the MVU estimator for

$$p(k, \theta) = \binom{m}{k}\left(\frac{\theta}{1 + \theta}\right)^k\left(\frac{1}{1 + \theta}\right)^{m-k}$$

can be obtained as a hypergeometric probability

$$h(z, n; k) = \binom{m}{k}\binom{mn - m}{z - k}\Big/\binom{mn}{z}.$$

One may note here that for $k = m = 1$, $h(z, n; 1) = z/n$ which is well-known.

(iv) *Negative binomial distribution:* One has

$$a(k) = \binom{m + k - 1}{k}, \qquad k = 0, 1, 2, \cdots; \qquad b(z, n) = \binom{mn + z - 1}{z}.$$

Therefore the MVU estimator for

$$p(k, \theta) = \binom{m + k - 1}{k}\theta^k(1 - \theta)^m$$

is obtained as a hypergeometric probability

$$h(z, n; k) = \binom{m + k - 1}{k}\binom{mn - m + z - k - 1}{z - k}\Big/\binom{mn + z - 1}{z}.$$

One may note here that for $k = 0$, $m = 1$, $h(z, n; 0) = (n - 1)/(n + z - 1)$ which has been obtained by Haldane [3] and by Girshick, Mosteller and Savage [1].

(v) *Logarithmic series distribution:* One has $a(k) = 1/k$, $k = 1, 2, \cdots$; $b(z, n) = (n!/z!)|S_z^n|$ where S_z^n is the Stirling number of the first kind with arguments n and z. Therefore, the MVU estimator for $p(k, \theta) = \alpha\theta^k/k$ where $1/\alpha = -\log(1 - \theta)$ is obtained as

$$h(z, n; k) = [z!/nk(z - k)!][|S_{z-k}^{n-1}|/|S_z^n|].$$

The MVU estimators of the variances of the MVU estimators of the above probability functionr can be easily obtained by using the relevant formulae of this section.

Acknowledgment. This author is thankful to his colleague Professor J. Lambek of McGill University for drawing his attention to the reference [4].

REFERENCES

[1] Girschick, M. A., Mosteller, F. and Savage, L. J. (1946). Unbiased estimates for certain binomial sampling problems with applications. *Ann. Math. Statist.* **17** 13–23.

[2] Guttman, Irwin (1958). A note on a series solution of a problem in estimation. *Biometrika* **45** 565–567.

[3] Haldane, J. B. S. (1945). On a method of estimating frequencies. *Biometrika* **33** 222–225.

[4] Khinchin, A. Y. (1952). *Three Pearls of Number Theory.* (Translation of the second, revised Russian edition.) Graylock Press, Rochester, New York.

[5] Noack, A. (1950). A class of random variables with discrete distributions. *Ann. Math. Statist.* **21** 127–132.

[6] Patil, G. P. (1957). Thesis for the Associateship of the Indian Statistical Institute.

[7] Patil, G. P. (1961). Asymptotic bias and efficiency of ratio estimates for generalized power series distributions and certain applications. *Sankhyā* **23** 269–280.

[8] Patil, G. P. (1962). Estimation by two-moments method for generalized power series distributions and certain applications. *Sankhyā* **24** 1–14.

[9] Patil, G. P. (1962). Maximum likelihood estimation for generalized power series distribution and its application to truncated binomial distribution. *Biometrika* **49** 222–232.

[10] Patil, G. P. (1962). On homogeneity and combined estimation for generalized power series distributions and certain applications. *Biometrics* **18** 365–374.

[11] Patil, G. P. (1962). Certain properties of the generalized power series distributions. *Ann. Inst. Statist. Math.* **14** 179–182.

[12] Patil, G. P. (1962). Some methods of estimation for the logarithmic series distribution. *Biometrics* **18** 68–75.

[13] Quenouille, M. H. (1960). Quoted in *The Advanced Theory of Statistics* **2** by M. G. Kendall and Alan Stuart. Griffin, London.

[14] Roy, J. and Mitra, S. K. (1958). Unbiased minimum variance estimation in a class of discrete distributions. *Sankhyā* **18** 371–378.

[15] Schnirelmann, L. (1932). Über additive Eigenschaften von Zahlen. *Math. Ann.* **107** 649–690.

[16] Tate, R. F. and Goen, R. L. (1958). Minimum variance unbiased estimation for the truncated Poisson distribution. *Ann. Math. Statist.* **29** 755–765.