

## Basic sequences and unbiased estimation in quasi power series distributions

Faqir Muhammad\* and M. Kazim Khan<sup>†‡</sup>

### Abstract

By using results from function space theory we give a characterization of when lacunary quasi power series sequences are basic in  $C[0, 1]$ . The paper discusses the links with unbiased estimable functions and the subspaces generated by the density of the lacunary quasi power series distributions. The paper also provides the rates of convergence of all the moments of the classic odds ratio estimator. This extends some known results in Bleimann, Butzer and Hahn's approximation operator.

**Keywords:** UMVU estimators, unbiased estimators, lacunary sequences, Schauder basis.

*Mathematics Subject Classification (2010):* Primary: 62F10, 62F12; Secondary: 62B05, 62B10 62E20.

*Received :* 03.04.2017 *Accepted :* 23.05.2017 *Doi :* 10.15672/HJMS.2017.481

### 1. Introduction

Consider  $X_1, X_2, \dots, X_n$  be a simple random sample from a power series distribution

$$(1.1) \quad \mathbb{P}_\theta(X_1 = k) = a_k \theta^k / G(\theta), \quad k \in \mathcal{A} \subseteq \mathbb{N},$$

where  $a_k > 0$  is a given sequence of real numbers,  $G(\theta) = \sum_{k \in \mathcal{A}} a_k \theta^k$  is their generating function, converging over  $\theta \in \Theta = (0, R)$ , where  $R$  being the radius of convergence of the power series, and  $\mathcal{A}$  is the support set not dependent on  $\theta$ . This class of distributions was introduced by Noack [14], see also [8], [11], [13]. The existence of the UMVU estimator of  $\theta$  depends on the nature of the set  $\mathcal{A}$ , of which the following examples being typical.

---

\*NAMAL College, Mianwali, Pakistan

Email : aioufsd@yahoo.com

<sup>†</sup>Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA

Email : kazim@math.kent.edu

<sup>‡</sup>Corresponding Author.

- (i) (Left truncation) There is a nonnegative integer  $N$  so that  $a_k > 0$  for all  $k \geq N$  and  $a_k = 0$  for  $k < N$ . In this case the UMVU estimator of  $\theta$  exists for all sample sizes. A classic example is the zero-truncated Poisson, for which  $a_0 = 0$  and  $a_k = \frac{1}{k!}$  for  $k = 1, 2, \dots$ . In 1953, R. L. Plackett [18] proposed an unbiased estimator of  $\theta$  in this case. Its efficiency does not go down below 0.95. In 1958 Tate and Goen [19] constructed the UMVU estimator of  $\theta$ . We should remark that computing the variance of the UMVU estimator has remained a challenge, and as far as we know, no known result exists for the variance of the UMVU estimators. Computationally it is known that the Tate and Goen's estimator has variance that is strictly larger than the Cramér-Rao lower bound. This can also be verified directly by a general result of Wijsman [20], [21].

Another well known example of this type is the logarithmic series distribution, for which  $a_0 = 0$  and  $a_k = \frac{1}{k}$  for  $k = 1, 2, \dots$ . Patil [15], and Patil and Joshi [17], showed that the UMVU estimator of  $\theta$  is  $A(n, k - 1)/A(n, k)$ , where

$$A(n, k) = \frac{n!}{k!} |S_k^n|, \quad k = n, n + 1, n + 2, \dots,$$

and  $S_k^n$  is the Stirling number of the first kind. Again its variance is not known.

- (ii) (Right truncation) There is a nonnegative integer  $N$  so that  $a_k = 0$  for all  $k > N$  and  $a_k > 0$  for  $k = 0, 1, \dots, N$ . In this case no UMVU estimator of  $\theta$  exists for any sample size. In this variety the problem of estimating the odds ratio is a typical example. Consider the case of  $a_0 = a_1 = 1$  and  $a_k = 0$  for all  $k \geq 2$ . In this case  $S_n = X_1 + \dots + X_n \sim B(n, p(\theta))$ , where  $p(\theta) = \theta/(1 + \theta)$ , and  $\theta$  being the odds ratio. Now no unbiased estimator of  $\theta$  exists for any sample size. It is easy to show that the usual biased estimator,  $\delta(S_n) = S_n/(n - S_n + 1)$ , is asymptotically normal. However, the computation of the mean squared error is quite nontrivial. Directly the asymptotic mean squared error was computed by Khan [9], [10]. An indirect proof was obtained by Della Vecchia [4] by linking it to an initial value problem. Both results are in the context of approximation theory, in particular, the Bleimann, Butzer and Hahn approximation operator, [3]. Here we extend their result that covers all the asymptotic moments. This is one of the few examples for which obtaining asymptotic moments seems to have tractable form.
- (iii) (Left-right truncation) There are nonnegative integers  $N < M$  so that  $a_k = 0$  for all  $k < N$  and  $k > M$  and  $a_k > 0$  for  $k \in [N, M]$ . Again no UMVU estimator of  $\theta$  exists for any sample size. Some typical examples involve the two-tailed truncated Poisson and the reverse Bessel polynomials case, for which  $N = 1$  and  $M \geq 2$  and  $a_j = (2M - j - 1)! / \{(j - 1)! (M - j)! 2^{M-j}\}$ , for  $j = N, 2, \dots, M$ . Bessel polynomials were introduced by Krall and Fink [12],

$$y_N(\theta) = \sum_{k=0}^N \frac{(N + k)!}{(N - k)! k! 2^k} \theta^k,$$

and are linked with the  $t$ -distribution [1]. The reverse Bessel polynomials are obtained by taking  $G_N(\theta) = \theta^N y_N(1/\theta)$ . This gives

$$G_N(\theta) = \sum_{j=1}^N a_j \theta^j, \quad a_j = \frac{(2N - j - 1)!}{(j - 1)! (N - j)! 2^{N-j}}, \text{ for } 1 \leq j \leq N.$$

Reverse Bessel polynomials find uses in filter design [2]. In this situation the maximum likelihood estimation technique seems to be the main approach for estimating  $\theta$ .

- (iv) (Mid section truncation) There are nonnegative integers  $N < M$  so that  $a_k > 0$  for all  $k \leq N$  and  $k \geq M$  and  $a_k = 0$  for  $k \in (N, M)$ . In this case for sufficiently large sample sizes the UMVU estimator of  $\theta$  may or may not exist.
- (v) (Lattice case with span larger than one). In this case we have a  $d > 1$  and  $a_k > 0$  for all  $k = c_0 + dk$  and  $a_k = 0$  otherwise, where  $c_0$  is a nonnegative integer. A typical example of this type arises when the information received are the first passage stopping times.

Consider a simple random sample of the waiting times until a ruin takes place in the classical gambler's ruin problem. More precisely,  $X_1 = \inf\{k \geq 1 : T_k \geq 1\}$  when  $T_k = (2U_1 - 1) + \dots + (2U_k - 1)$ , and  $U_1, U_2, \dots \stackrel{iid}{\sim} B(1, p)$  form a Bernoulli process. In this case the support set of  $X_1$  is the set of odd positive integers. Using the reflection principle the distribution of  $S_n = X_1 + X_2 + \dots + X_n$  is

$$\mathbb{P}(S_n = j) = (p/q)^{n/2} \frac{n}{j} \binom{j}{\frac{j-n}{2}} (pq)^{j/2}, \quad j = n, n+2, n+4, \dots,$$

when  $p > \frac{1}{2}$ , and  $q = 1 - p$ . Hence when  $A(n, j) > 0$ , we get  $A(n, j-1) = 0$  almost surely, making estimation of odd power monomials of  $\theta = \sqrt{pq}$ , by an unbiased estimator problematic.

The existence of unbiased estimators of  $\theta$  has some history. It can be settled easily in the above situations of the support set  $A$ . When the support set  $A$  is not of the above types the situation becomes more interesting and the existence of minimum variance unbiased estimators has been studied by Patil [15] and Patil and Joshi [16] using additive number theory.

We show that, in a bit more general context of quasi-power series distributions, the unbiased estimation problem is related to the basis problem in  $C[0, 1]$ , the space of continuous functions with supremum norm. The main results are collected in the next section and all the proofs are collected in the last section.

## 2. The Main Results

It is generally difficult to find the variance of estimators when they do not achieve the Cramér-Rao lower bound, as is typically the case in truncated forms of power series distributions. One exception being the odds ratio for which the limiting form of the second moment can be computed. The following theorem extends this to moments of all orders.

**2.1. Theorem.** *Let  $S_n = X_1 + \dots + X_n$ , where  $X_i \stackrel{iid}{\sim} B(1, p)$  and  $p = \frac{\theta}{1+\theta}$  making  $\frac{p}{1-p} = \theta$  being the odd's ratio. Denote the variance by  $\sigma^2 = p(1-p)$  and the Fisher information for  $\theta$ ,  $I(\theta) = \frac{1}{\theta(1+\theta)^2}$ . For any positive integer  $m$ , we have*

$$n^m \mathbb{E} \left( \frac{S_n}{n - S_n + 1} - \theta \right)^{2m} \rightarrow \left( \frac{1}{I(\theta)} \right)^m \mathbb{E}(Z^{2m}), \quad Z \sim N(0, 1).$$

When the support set has increasingly larger gaps the situation becomes more complicated. One interesting case that consider in the following is the lacunary power series distributions. For this we need to introduce some terminology.

**2.2. Definition.** A sequence of positive integers  $n_1 < m_1 < n_2 < m_2 < \dots$  is called lacunary if

$$\inf_k \frac{m_k}{n_k} > 1, \quad \text{and} \quad \inf_k \frac{n_{k+1}}{m_k} > 1.$$

When  $q = \min\{\inf_k \frac{m_k}{n_k}, \inf_k \frac{n_{k+1}}{m_k}\} > 1$ , the sequence is called  $q$ -lacunary.

When the support set  $A = \{m_0, m_1, \dots\}$  of a power series distribution consists of nonnegative integers forming a lacunary sequence it will be more convenient to transfer this structure over to the exponents of  $\theta$  by considering,

$$\mathbb{P}(Y = k) = b_k \theta^{m_k} / G(\theta), \quad k = 0, 1, 2, \dots,$$

where  $b_k = a_{m_k} > 0$ , for all  $k = 0, 1, 2, \dots$ , and  $G(\theta) = \sum_k b_k \theta^{m_k}$ . This family of distributions for  $\theta \in (0, R)$ , belongs to the one parameter exponential family since  $\mathbb{P}(Y = k) = b_k e^{m_k \ln \theta} / G(\theta)$ , with a complete sufficient statistic,  $\delta(Y) := m_Y$ . The moment generating function of  $\delta(Y)$  being  $G(e^t \theta) / G(\theta)$  for all  $t$  in a neighborhood of zero and

$$\mathbb{E}_\theta(\delta(Y)) = \frac{\theta G'(\theta)}{G(\theta)}, \quad \text{Var}_\theta(\delta(Y)) = \theta \frac{G(\theta)\{\theta G''(\theta) + G'(\theta)\} - \theta(G'(\theta))^2}{G^2(\theta)}.$$

Recall that in a Banach space  $X$  a sequence  $\{e_k\}$  is called a **basis** if for each  $x \in X$  we have a unique decomposition  $x = \sum_k \delta_k e_k$ , where the series converges in the norm of the Banach space. A sequence  $\{e_k\}$  is called **basic** if it forms a basis for the subspace generated by  $\{e_k\}$ .

When an unbiased estimator of  $h(\theta)$  exists for  $\theta \in [0, R)$  (assume  $R > 1$  after rescaling if needed) then  $h(\theta)G(\theta) = \sum_k \delta_k b_k \theta^{m_k}$  for all  $\theta \in [0, 1]$  and the series will converge uniformly in  $[0, 1]$  with radius of convergence at least  $R$ . Hence, the function  $g(t) := h(t)G(t)$  will be in the closed linear span of  $\{t^{m_k}\}$ . Conversely consider any member,  $g$ , of the closed linear span of  $\{t^{m_k}\}$  with a power series  $g(t) := \sum_k \alpha_k t^{m_k}$ . If this  $g$  is such that its radius of convergence is at least  $R$  then the function  $h(\theta) := g(\theta) / G(\theta)$  will have an unbiased estimator, with  $\delta_k = \alpha_k / b_k$ . Therefore it is interesting to consider the question of when  $e_k = t^{m_k}$ ,  $k = 0, 1, 2, \dots$  will be a basic sequence in  $C[0, 1]$ . The next two theorems characterizing when lacunary power series distributions form a Schauder basis for its closed linear span in  $C[0, 1]$ .

**2.3. Definition.** Let  $\{e_k\}$  be a sequence in a Banach space and let  $L\{e_k\} = L\{e_1, e_2, \dots\}$  be the linear cover of  $\{e_1, e_2, \dots\}$  and let  $\bar{L}\{e_k\}$  be its closure. Also let  $L^{(k)} = L(e_1, \dots, e_{k-1}, e_{k+1}, \dots)$ . We say the sequence  $\{e_k\}$  is minimal if  $\rho_k = \text{distance of } e_k \text{ and } L^{(k)}$  is strictly positive for each  $k = 1, 2, \dots$ . The sequence is called  $\rho$ -minimal if  $\rho_k > \rho > 0$  for all  $k = 1, 2, \dots$ .

Note that when  $\{e_k\}$  is  $\rho$ -minimal and  $x = \sum_{k=1}^\infty a_k e_k$ , we have  $|a_k| \leq \|x\| / \rho$  for all  $k = 1, 2, \dots$ .

Banach himself knew that every infinite dimensional Banach space contains a closed linear subspace with a basic sequence, i.e. the sequence forms a Schauder basis for its closed linear span. Per Enflo [5] showed that there do exist separable Banach spaces for which the whole space has no Schauder basis. The next theorem helps explain how the linear space of unbiased estimable functions in  $C[0, 1]$  for a lacunary power series distribution look like. We present the result in a bit more generality of quasi lacunary sequences, by relaxing the one-parameter exponential family structure of power series distributions, to highlight only the needed features.

Let  $\{b_k, k = 0, 1, 2, \dots\}$  be a sequence of numbers with  $b_k > 0$  and introduce  $G(\theta) = \sum_{k=0}^\infty b_k \psi_k(\theta)$ . We assume that the series converges uniformly in  $\theta \in [0, 1]$  and the collection  $\{\psi_k(\theta), k = 0, 1, \dots\}$  is assumed to have the following properties.

- i:  $\psi_0(\theta)$  is a positive constant and  $\psi_k(\theta) \in [0, 1]$ , with  $\psi_k(\theta) > 0$  for  $\theta > 0$ .
- ii:  $\psi_k(\theta)$  is increasing on  $[0, 1]$  for every  $k$ , and  $\psi_k(\theta)$  is decreasing to zero in  $k$  for every fixed  $\theta \in [0, 1]$ . Let each function of the sequence  $\{\psi_k(t), k \geq 1\}$  be differentiable, positive on  $(0, 1]$ , increasing and convex on  $[0, 1]$ .
- iii:  $\psi_k(\theta)$  are uniformly bounded in  $k$  and  $\theta \in [0, 1]$ .

For such a sequence if there exist  $n_1 < m_1 < n_2 < m_2 < \dots$  such that

$$t^{m_k} \leq \psi_k(t) \leq t^{n_k}, \quad k = 1, 2, \dots, t \in [0, 1],$$

then the collection  $\{\psi_k(t), k \geq 1\}$  will be called a quasi-power sequence.

**2.4. Theorem.** *There exists a function  $q(\rho) > 1$  for  $\rho \in [0, 1]$  such that if  $\{\varphi_k(\theta), k \geq 1\}$  is a  $\rho$ -minimal quasi-power sequence corresponding to a  $q$ -lacunary sequence  $\{(n_k, m_k), k \geq 1\}$  then  $\{\varphi_k(\theta), k \geq 1\}$  is basic in  $C[0, 1]$ .*

In particular, the above theorem gives that for a sample of size one every unbiased estimable continuous function for this family comes from  $\frac{1}{G(\theta)} \cdot \overline{\text{span}}\{\varphi_k(\theta), k \geq 1\}$  with a unique unbiased estimator which becomes the UMVU automatically. We may conclude more.

**2.5. Theorem.** *There exists a  $\rho \in (0, 1]$  and a  $q > 1$  such that every quasi-power sequence  $\{\psi_k(\theta), k \geq 0\}$  in  $C[0, 1]$  is  $\rho$ -minimal.*

Since the space unbiased estimable functions,  $\mathcal{U}_k$  from a random sample of size  $k$ , must consist of analytic functions over subintervals of  $[0, b)$ , it must be that  $\mathcal{U}_k \subseteq \mathcal{U}_{k+1}$  generate proper subspaces of  $C[0, 1]$  for any finite sample size. Furthermore, for instance, for  $k = 3$  note that

$$\begin{aligned} \sum_{i \geq 1} \sum_{j > i} \sum_{\ell > j} \frac{1}{m_i + m_j + m_\ell} &= \sum_{i \geq 1} \frac{1}{m_i} \sum_{j > i} \frac{1}{1 + \frac{m_j}{m_i}} \sum_{\ell > j} \frac{1}{1 + \frac{m_\ell}{m_i + m_j}} \\ &\leq \sum_{i \geq 1} \frac{1}{m_i} \sum_{j > i} \frac{1}{q^{j-i}} \sum_{\ell > j} \frac{1}{1 + \frac{m_\ell}{2m_j}} \\ &\leq \sum_{i \geq 1} \frac{1}{m_i} \sum_{j > i} \frac{1}{q^{j-i}} \sum_{\ell > j} \frac{2}{q^{\ell-j}} \\ &< \infty. \end{aligned}$$

This idea works for higher values of  $k$  analogously. Hence, by Muntz' theorem the distribution of the complete sufficient statistic with sample size  $k \geq 2$  will not be total in the sense that the linear combinations will not be dense in  $C[0, 1]$ .

### 3. The Proofs

For the case of odds ratio, recall the notation  $a_0 = a_1 = 1$  and  $a_k = 0$  for all  $k \geq 2$ . In this case  $S_n \sim B(n, \theta/(1+\theta))$ . The estimator of the odds ratio is  $\delta(S_n) = S_n/(n - S_n + 1)$ .

*Proof.* (Theorem 2.1) Note that, if  $Z_n = (S_n - np)/(\sigma\sqrt{n})$ , then we may write

$$\sqrt{n} \left( \frac{S_n}{n - S_n + 1} - \theta \right) = \frac{1}{1 + \frac{1}{n} - \frac{S_n}{n}} \left\{ \sqrt{\theta} Z_n - \frac{\theta}{\sqrt{n}} \right\}$$

Let  $C = \{ \frac{S_n}{n} - p \leq \frac{1-p}{2} \}$ . Hence, breaking the expression inside the expectation over  $C$  and  $C^c$ , we have

$$\begin{aligned} n^m \left( \frac{S_n}{n - S_n + 1} - \theta \right)^{2m} &= \frac{1}{(1 + \frac{1}{n} - \frac{S_n}{n})^{2m}} \left\{ \sqrt{\theta} Z_n - \frac{\theta}{\sqrt{n}} \right\}^{2m} \chi_C \\ &\quad + \frac{1}{(1 + \frac{1}{n} - \frac{S_n}{n})^{2m}} \left\{ \sqrt{\theta} Z_n - \frac{\theta}{\sqrt{n}} \right\}^{2m} \chi_{C^c}. \end{aligned}$$

When  $C$  occurs, the first of the two expressions is bounded by  $(\frac{1}{1-p})^{2m}$  and the expectation of the rest converges to  $\mathbb{E}(\sqrt{\theta} Z)^{2m}$  since the probability of the event  $C$  increases

to one. The interchange of limit and expectation is justified by the extended Lebesgue dominated convergence theorem. Therefore the expectation of the whole expression with event  $C$  goes to  $\frac{\theta^m}{(1-p)^{2m}}\mathbb{E}(Z^{2m})$ . Now

$$\frac{\theta^m}{(1-p)^{2m}} = \theta^m(1+\theta)^{2m} = \left(\frac{1}{I(\theta)}\right)^m.$$

Next consider the  $C^c$  term. The first of its two expressions is bounded above by  $n^{2m}$ . The expectation of  $\left\{\sqrt{\theta}Z_n - \frac{\theta}{\sqrt{n}}\right\}^{4m}$  converges to  $\mathbb{E}(\sqrt{\theta}Z)^{4m} < \infty$  and the probability of the event  $C^c$  is bounded above, by Chernoff's large deviation result,

$$\mathbb{P}\left(\frac{S_n}{n} - p > \frac{1-p}{2}\right) \leq e^{-n\gamma(p)}.$$

Since,  $n^{2m}e^{-n\gamma(p)/2} \rightarrow 0$ , by Chebyshev's inequality the expectation of the whole expression with  $C^c$  goes to zero. ■

For the proof of the next two theorems we need to recall a few auxiliary results from function space theory. A sequence  $\{x_k\}$  in a Banach space is called non-closing if  $\|x_k - x_j\| \geq \delta > 0$  for any  $k \neq j$  and some  $\delta$ . Also we recall the following result of Gurariy and Matsaev [6] for lacunary power series.

**3.1. Proposition.** *The following statements are equivalent for a sequence  $0 < n_1 < n_2 < \dots$ :*

- (1) *The collection  $\{t^{n_k}, k = 0, 1, \dots\}$  is non-closing in  $C$ ,*
- (2) *The collection  $\{t^{n_k}, k = 0, 1, \dots\}$  is uniformly minimal in  $C$ ,*
- (3) *The collection  $\{t^{n_k}, k = 0, 1, \dots\}$  is basic in  $C$ ,*
- (4) *The collection  $\{n_k, k = 0, 1, \dots\}$  is a lacunary sequence.*

The following proposition for quasi power series is proved in Gurariy and Lusky [7] in more generality, p. 82-83. For our special case that we need, we provide a shorter argument.

**3.2. Proposition.** *The quasi-power sequence  $\{\psi_k, k \geq 1\}$  is non-closing in  $C$  if and only if the corresponding sequence  $n_1 < m_1 < n_2 < m_2 < \dots$  is lacunary.*

*Proof.* Given lacunarity of  $\{n_k, m_k, k \geq 1\}$  we have,

$$\|\psi_k - \psi_j\| \geq \max_{t \in [0,1]} \|t^{m_k} - t^{n_j}\|.$$

The last term is at least  $\delta > 0$  due to the above proposition of Gurariy and Matsaev. Therefore,  $\{\psi_k, k \geq 1\}$  is non-closing in  $C$ . Conversely, when  $\{\psi_k, k \geq 1\}$  is non-closing in  $C$  we want to prove the lacunarity of the corresponding sequence  $\{n_k, m_k, k \geq 1\}$ . For this we note that for any  $k \geq 1$  we have

$$0 < \delta \leq \|\psi_k - \psi_{k+1}\| \leq \max_{t \in [0,1]} \|t^{n_k} - t^{m_{k+1}}\|.$$

This implies that  $\max_{t \in [0,1]} \|t^j - t^k\| \geq \delta$ , where  $j, k \in \{n_k, m_k, k \geq 1\}$  and  $j \neq k$ . This implies that the system  $t^{n_1}, t^{m_1}, t^{n_2}, t^{m_2}, \dots$ , is also non-closing in  $C$ . Again by the above proposition of Gurariy and Matsaev  $\{n_k, m_k, k \geq 1\}$  is lacunary. ■

Let  $t_1 \in (0, 1]$  be a point of maxima of  $|f(t)|$  where  $\|f\| = 1$ . For any given  $\delta > 0$  let  $t_0 = t_1 - \delta > 0$ . Then, by the Lagrange mean value theorem,  $|f(t_0)| \geq 1 - M'\delta$  where  $M' = \|f'\|_C$ .

We will need the following version of this result whose proof is not difficult and therefore we omit it.

**3.3. Lemma.** Let  $f \in C[a, b]$  and  $g \in C^1[a, b]$ . Then we have

$$\max_{t \in [a, b]} |f(t) + g(t)| \geq \frac{\max_{t \in [a, b]} f(t) - \min_{t \in [a, b]} f(t) - \max_{t \in [a, b]} |g'(t)|(b-a)}{2}.$$

The proofs of the theorems use the modified Banach-Grinblum criterion which we recall. Let  $P, Q$  be two subspaces of a Banach space. The *inclination* of  $P$  to  $Q$ , denoted as  $\widehat{(P, Q)}$  is defined as

$$\widehat{(P, Q)} = \inf_{x \in P, \|x\|=1} \text{distance}(x, Q) = \inf_{x \in P, y \in Q, x \neq y} \frac{\|x + y\|}{\|x\|}.$$

We have  $0 \leq \widehat{(P, Q)} \leq 1$ . When  $\widehat{(P, Q)} = 1$  we say that  $P$  is **orthogonal** to  $Q$ . Let  $\{e_k, k \geq 1\}$  be a complete (normalized) sequence in a Banach space  $X$ , and let  $L_{n,m} = L(e_n, e_{n+1}, \dots, e_m)$  represent the linear span of  $\{e_n, \dots, e_m\}$ . The Banach-Grinblum criterion says that  $\{e_k, k \geq 1\}$  is a basis in  $X$  if and only if  $(L_{1,n}, \widehat{L_{n+1,m}}) \geq \delta > 0$  for every choice of  $n < m$  and  $\delta$  does not depend on the choice of  $n, m$ . When  $\delta = 1$  can be used then the basis is orthogonal or monotone. The modified Banach-Grinblum criterion says that  $\{e_k\}$  is basic if and only if for all  $x \in L_{1,n}$  and  $y \in L_{n+1,m}$  we have  $\|x + y\| \geq \beta > 0$  where  $\beta$  does not depend on  $n, m$ .

*Proof.* (Theorem 2.4) Let  $L_{i,j}$  be the linear span of  $\varphi_i, \dots, \varphi_j$ , and use the notation

$$x_k(t) = \sum_{j=1}^k \alpha_j \varphi_j(t), \quad y_k(t) = \sum_{j=k+1}^m \alpha_j \varphi_j(t),$$

with  $\|x_k\| = \max_{t \in [0,1]} |x_k(t)| = 1$  and  $\|y_k\| = 1$ . By the  $\rho$ -minimality,  $|\alpha_j| \leq \frac{1}{\rho}$ , for  $j = 1, 2, \dots$ . Therefore, for any  $t < 1$  we have

$$\begin{aligned} |x'_k(t)| &= \left| \sum_{j=1}^k \alpha_j \varphi'_j(t) \right| \leq \frac{1}{\rho} \sum_{j=1}^k \varphi'_j(1) \\ &\leq \frac{1}{\rho} \sum_{j=1}^k m_j \leq \frac{m_k}{\rho} \frac{q}{q-1}. \end{aligned}$$

Denote the last bound by  $M_k$ ,  $k = 1, 2, \dots$ . Let  $t_1$  be the right point of maxima of the function  $x_k(t)$  on  $[0, 1]$ . Then by the mean value theorem for a small given  $\delta > 0$  we have a point  $t_0$  such that  $|x_k(t_1)| = 1$  and  $t_0 = t_1 - \delta \leq 1 - \delta$ . Take  $\delta = \delta_k = 1/(2M_k) < t_1$ , for large  $q = q(\rho)$ . Then note that  $|x_k(t_0)| \geq 1 - \delta_k M_k = \frac{1}{2}$ . Now consider the function  $y_k(t) \in L_{k+1,m}$ . We will prove that  $|y_k(t_0)| \leq \frac{1}{4}$ . Once again,

$$|y_k(t_0)| \leq \frac{1}{\rho} \sum_{j=k+1}^m |\varphi_j(t_0)| \leq \frac{1}{\rho} \sum_{j=k+1}^m t_0^{n_j}.$$

First note that

$$\begin{aligned} r &= \frac{t_0^{n_{j+1}}}{t_0^{n_j}} \leq (1-\delta)^{n_{j+1}-n_j} = (1-\delta)^{n_j \left( \frac{n_{j+1}}{n_j} - 1 \right)} \\ &= \left( (1-\delta)^{\frac{1}{\delta}} \right)^{\delta n_j \left( \frac{n_{j+1}}{n_j} - 1 \right)} \leq (1/e)^{\delta n_{k+1}(q-1)} \\ &= (1/e)^{\frac{\rho(q-1)}{2qm_k} n_{k+1}(q-1)} = (1/e)^{\frac{\rho(q-1)^2}{2q}} \leq \frac{1}{2}, \quad \text{for large } q = q(\rho). \end{aligned}$$

Using the same technique we have

$$t_0^{n_{k+1}} \leq (1-\delta)^{n_{k+1}} \leq \left( (1-\delta)^{1/\delta} \right)^{\delta n_{k+1}} \leq (1/e)^{\rho(q-1)/2}.$$

Therefore we see that

$$|y_k(t_0)| \leq \frac{1}{\rho} \times \frac{t_0^{n_{k+1}}}{1-r} \leq \frac{2}{\rho} \exp\{-\rho(q-1)/2\} \leq \frac{1}{4} \quad \text{for large } q = q(\rho).$$

This shows that  $\|x_k - y_k\| \geq |x_k(t_0) - y_k(t_0)| \geq \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$ . By the modified Banach-Grinblum criterion, we have  $\{\varphi_k(t)\}$  is basic. ■

*Proof.* (Theorem 2.5) Given a finite system  $\{\psi_j, j = 1, 2, \dots, n\}$ , by induction on  $n$ , we will catch both  $q$  and  $\rho$ . Assume that there are  $q, \rho$  such that every  $q$ -lacunary system  $\{\psi_j, j = 1, 2, \dots, k\}$  is  $\rho$ -minimal. Consider the  $q$ -lacunary system  $\{\psi_j, j = 1, 2, \dots, k+1\}$  and we will prove its  $\rho$ -minimality. Recall that  $\|\psi_j\| = 1$  for all  $j$ . For any  $j_0$  in between 1 and  $k+1$ , take

$$\varphi(t) = \sum_{j=1, j \neq j_0}^{k+1} \alpha_j \psi_j \in L(\psi_1, \dots, \psi_{j_0-1}, \psi_{j_0+1}, \dots, \psi_{k+1}).$$

It is sufficient to show that

$$\|\psi_{j_0} + \varphi\| \geq \rho.$$

Otherwise, by the triangular inequality, we will have  $\|\varphi\| \leq 1 + \rho$  and  $\rho$ -minimality of  $\{\psi_j, j = 1, \dots, k+1, j \neq j_0\}$  provides,

$$|\alpha_j| \leq \frac{\|\varphi\|}{\rho} \leq \frac{1+\rho}{\rho}, \quad \psi_j, \quad j = 1, \dots, k+1, \quad j \neq j_0.$$

Write  $\varphi$  as  $\varphi = \varphi_1 + \varphi_2$ , where

$$\varphi_1 = \sum_{j=1}^{j_0-1} \alpha_j \psi_j, \quad \varphi_2 = \sum_{j=j_0+1}^{k+1} \alpha_j \psi_j.$$

We will use the Lemma 3.3 on the interval  $[1-\delta, 1]$  and later we will take the constant  $\delta$  to be  $\delta = \frac{1}{n_{j_0}}$ . Also we will take  $f = \psi_{j_0}$  and  $g = \varphi_1$ . Because each of  $f$  and  $t \in [0, 1]$  we have

$$\varphi_1'(t) \leq \frac{1+\rho}{\rho} \sum_{j=1}^{j_0-1} m_j \leq \frac{q(1+\rho)m_{j_0-1}}{\rho(q-1)}.$$

Also,

$$\begin{aligned} \|\psi_{j_0} + \varphi_1\| &\geq \sup_{t \in [1-\delta, 1]} |\phi_{j_0}(t) + \varphi_1(t)| \\ &\geq \frac{1}{2} \left( 1 - (1-\delta)^{n_{j_0}} - \frac{q(\rho+1)}{\rho(q-1)} m_{j_0-1} \delta \right) \\ &\geq \frac{1}{2} \left( 1 - \frac{1}{e} - \frac{\rho+1}{\rho(q-1)} \right) \\ &\geq \frac{1}{2} \left( \frac{1}{2} - \frac{\rho+1}{\rho(q-1)} \right). \end{aligned}$$

Here we used that  $(1-\delta)^{n_{j_0}} = \left\{ (1-\delta)^{1/\delta} \right\}^1 \leq 1/e$ .

Next, we estimate  $\varphi_2(1-\delta)$ , by choosing a  $\delta'$  so that  $1-\delta < 1-\delta'$ , i.e.,  $\delta' < \delta = \frac{1}{n_{j_0}}$ . In this regard take  $\delta' = \frac{1}{\lambda m_{j_0}}$  where  $\lambda > 1$  will be set later (actually we will take  $\lambda = 25$ ). Note that since  $n_{j_0} < m_{j_0} < \lambda m_{j_0}$  we have  $\delta = \frac{1}{n_{j_0}} > \frac{1}{\lambda m_{j_0}}$  and therefore  $1-\delta < 1-\delta'$ .



Since  $\psi_j(t)$  is monotone increasing in  $t$ , we see that

$$\begin{aligned} |\varphi_2(t)| &\leq |\varphi_2(1 - \delta')| \quad 0 \leq t \leq 1 - \delta', \\ &\leq \sum_{j=j_0+1}^{k+1} |\alpha_k| \psi_j(1 - \delta') \leq \frac{1 + \rho}{\rho} \sum_{j=j_0+1}^{k+1} (1 - \delta')^{n_j} \\ &= \frac{1 + \rho}{\rho} (1 - \delta')^{n_{j_0+1}} \sum_{j=j_0+1}^{k+1} \left\{ (1 - \delta')^{1/\delta'} \right\}^{\delta'(n_j - n_{j_0+1})} \\ &\leq \frac{1 + \rho}{\rho} (1 - \delta')^{n_{j_0+1}} \sum_{j=j_0+1}^{k+1} \left( \frac{1}{e} \right)^{\delta' n_{j_0+1} (\frac{n_j}{n_{j_0+1}} - 1)}. \end{aligned}$$

Here we notice that for  $j = j_0 + 2, j_0 + 3, \dots, k + 1$ , we have

$$\begin{aligned} \frac{n_j}{n_{j_0+1}} - 1 &= \frac{n_j}{n_{j-1}} \frac{n_{j-1}}{n_{j-2}} \dots \frac{n_{j_0+2}}{n_{j_0+1}} - 1 \\ &\geq q^{j-j_0-1} - 1, \quad \text{since } j - 1 = j_0 + (j - j_0) - 1. \end{aligned}$$

This holds for  $j = j_0 + 1$  as well. Therefore, we have

$$\begin{aligned} |\varphi_2(t)| &\leq \frac{1 + \rho}{\rho} (1 - \delta')^{n_{j_0+1}} \sum_{j=j_0+1}^{k+1} \left( \frac{1}{e} \right)^{\delta' n_{j_0+1} (q^{j-(j_0+1)} - 1)} \\ &= \frac{1 + \rho}{\rho} (1 - \delta')^{n_{j_0+1}} \sum_{i=0}^{k+1-j_0-1} \left( \frac{1}{e} \right)^{\delta' n_{j_0+1} (q^i - 1)}. \end{aligned}$$

Now note that for any  $q > 1$  and any  $i = 0, 1, 2, \dots$  we have  $i(q - 1) \leq (q^i - 1)$ . Therefore, we have

$$\begin{aligned} |\varphi_2(t)| &\leq \frac{1 + \rho}{\rho} (1 - \delta')^{n_{j_0+1}} \sum_{i=0}^{k+1-j_0-1} \left( \frac{1}{e} \right)^{\delta' n_{j_0+1} i (q-1)} \\ &= \frac{1 + \rho}{\rho} (1 - \delta')^{n_{j_0+1}} \sum_{i=0}^{k+1-j_0-1} \left( \frac{1}{e} \right)^{\frac{n_{j_0+1}}{\lambda m_{j_0}} i (q-1)} \\ &\leq \frac{1 + \rho}{\rho} (1 - \delta')^{n_{j_0+1}} \sum_{i=0}^{k+1-j_0-1} \left( \frac{1}{e} \right)^{\frac{q}{\lambda} i (q-1)} \\ &\leq \frac{1 + \rho}{\rho} (1 - \delta')^{n_{j_0+1}} \sum_{i=0}^{k+1-j_0-1} \left( \frac{1}{e} \right)^{\frac{i q}{\lambda}}, \quad \text{when } q \geq 2, \\ &\leq \frac{1 + \rho}{\rho} (1 - \delta')^{n_{j_0+1}} \frac{1}{1 - e^{-q/\lambda}}. \end{aligned}$$

Now we work on the middle term, by noticing that

$$(1 - \delta')^{n_{j_0+1}} = \left\{ \left( 1 - \frac{1}{\lambda m_{j_0}} \right)^{\lambda m_{j_0}} \right\}^{n_{j_0+1}/(\lambda m_{j_0})} \leq (1/e)^{n_{j_0+1}/(\lambda m_{j_0})} \leq (1/e)^{q/\lambda}.$$

Hence, we see that for all  $0 \leq t \leq 1 - \delta'$ , we have

$$\begin{aligned} |\varphi_2(t)| &\leq \frac{1 + \rho}{\rho} \frac{e^{-q/\lambda}}{1 - e^{-q/\lambda}} \\ &\leq \frac{1.01(1 + \rho)}{\rho} e^{-q/\lambda}, \quad \text{for all } \frac{q}{\lambda} \geq 6. \end{aligned}$$

Therefore by one of the above results,

$$\begin{aligned} |\{\psi_{j_0}(t) + \varphi_1(t)\}'| &\leq \frac{q(\rho+1)}{\rho(q-1)} m_{j_0-1} + m_{j_0} \\ &\leq \frac{q(\rho+1)}{\rho(q-1)} \frac{m_{j_0}}{q} + m_{j_0} = m_{j_0} \frac{\rho q + 1}{\rho(q-1)}. \end{aligned}$$

Let  $t_1$  be the largest point of maxima of the function  $|\psi_{j_0}(t) + \varphi_1(t)|$  with  $t_0 = t_1 - \delta'$  and as earlier we regard that  $t_0 \in (0, 1)$ . Also note that  $0 < t_0 < 1 - \delta'$ . From our earlier estimates and the Lagrange mean value theorem, we have

$$\begin{aligned} |\psi_{j_0}(t_0) + \varphi_1(t_0)| &\geq \frac{1}{2} \left( \frac{1}{2} - \frac{\rho+1}{\rho(q-1)} \right) \cdot -\delta' m_{j_0} \frac{\rho q + 1}{\rho(q-1)} \\ &= \frac{1}{4} - \frac{\rho+1}{2\rho(q-1)} - \frac{\rho q + 1}{\lambda\rho(q-1)}. \end{aligned}$$

Therefore we have a lower estimate

$$\begin{aligned} \|\psi_{j_0} + \varphi_1 + \varphi_2\| &\geq |\psi_{j_0}(t_0) + \varphi_1(t_0) + \varphi_2(t_0)| \\ &\geq \frac{1}{4} - \frac{\rho+1}{2\rho(q-1)} - \frac{\rho q + 1}{\lambda\rho(q-1)} - \frac{1.01(\rho+1)}{\rho} e^{-q/\lambda}. \end{aligned}$$

Now in the induction we put  $\rho = \frac{1}{8}$  and restrict  $\lambda = 25$  to get, for a sufficiently large  $q$ ,

$$\begin{aligned} \frac{\rho+1}{2\rho(q-1)} &< \frac{1}{24}, & \text{because } \frac{q+1}{2\rho(q+1)} &\rightarrow 0, \text{ as } q \rightarrow \infty, \\ \frac{\rho q + 1}{\lambda\rho(q-1)} &< \frac{1}{24}, & \text{because } \frac{\rho q + 1}{\rho(q-1)} &\rightarrow 1, \\ \frac{1.01(\rho+1)}{\rho} e^{-q/\lambda} &< \frac{1}{24}, & \text{because it goes to zero.} \end{aligned}$$

Hence, we have shown that

$$\|\psi_{j_0} + \varphi\| \geq \frac{1}{8}, \quad \text{for all } j = 2, 3, \dots, k.$$

The same argument, in a bit simplified form, works for  $j = 1$  and  $j = k + 1$ . Thus the system  $\{\psi_j, j = 1, 2, \dots, k + 1\}$  is  $\frac{1}{8}$ -minimal and by induction we obtain that the system  $\{\psi_j, j = 1, 2, \dots\}$  is  $\frac{1}{4}$ -minimal. ■

#### 4. Acknowledgement

This paper benefited from discussions with our colleague and friend, Vladimir Gurariy, before his untimely passing away. We would also like to thank the referees for their careful reading of the manuscript and valuable suggestions.

#### References

- [1] Berg, C. and Vignat, C. *Linearization coefficients of Bessel polynomials and properties of Student  $t$ -distribution*, Constructive Approximation, **27**, 15–32, 2008.
- [2] Bianchi, G. and Sorrentino, R. *Electronic Filter Simulation and Design*, (McGraw Hill, NY, 2007).
- [3] Bleimann, G.; Butzer, P. L. and Hahn, L. *A Bernstein-type operator approximating continuous functions on the semi-axis*, Nederl. Akad. Wetensch. Indag. Math., **42**, 255–262, 1980.
- [4] Della Vecchia, B. *Some properties of a rational operator of Bernstein-type*, In: Progress in Approximation Theory, 177–185, (Academic Press, Boston, 1991).

- [5] Enflo, Per. *A counter example to the approximation problem in Banach spaces*, Acta Mathematica, **130** (1), 309–317, 1973.
- [6] Gurariy, V.I. and Matsaev, V.I. *Lacunary power sequence in the spaces  $C$  and  $L_p$* , Izv. Akad. Nauk SSR Ser. Mat. **30**, 3–14, 1966 (in Russian). Amer. Math. Soc. Trans. **72**, 9–21, 1968 (in English).
- [7] Gurariy, V.I. and Lusky, W. *Geometry of Müntz Spaces and Related Questions*, Lect. Notes Math. **1870**, Springer-Verlag, Berlin, 2005.
- [8] Joshi, S.W. and Park, C.J. *Minimum variance unbiased estimation for truncated power series distribution*, Sankhya: The Indian Journal of Statistics, Series A, **36**(3), 305–314, 1974.
- [9] Khan, R. A. *A note on a Bernstein-type operator of Bleimann, Butzer and Hahn*, J. Approx. Theory, **53**, 295–303, 1988.
- [10] Khan, R. A. *Reverse martingales and approximation operators*, J. Approx. Theory, **80**, 367–377, 1995.
- [11] Khatri, C.G. *On certain properties of power series distributions*, Biometrika, **46**, 486–490, 1959.
- [12] Krall, H. L., and Fink, O., *A new class of orthogonal polynomials: The Bessel polynomials*, Trans. Amer. Math. Soc., **65**, 100–115, 1948.
- [13] Lehmann, E. L. and Casella, George. *Theory of Point Estimation*, 2nd ed. (Springer-Verlag, N.Y. 1998).
- [14] Noack, A. *A class of random variables with discrete distributions*, Ann. Math. Stat. **21**, 127–132, 1950.
- [15] Patil, G. P. *Minimum Variance Unbiased Estimation and Certain Problems of Additive Number Theory*, Ann. Math. Stat. **34**, 1050–1056, 1963.
- [16] Patil, G. P. and Joshi, S. W. *Further Results on Minimum Variance Unbiased Estimation and Additive Number Theory*, Ann. Math. Stat. **41**(2), 567–575, 1970.
- [17] Patil, G. P. and Wani, J. K. *On Certain Structural Properties of the Logarithmic Series Distribution and the First Type Stirling Distribution*, Sankhya: The Indian Journal of Statistics, Series A, **27**(2/4), 271–280, 1965.
- [18] Plackett, R. L. *The truncated Poisson distribution*, Biometrics, **9**, 485–488, 1953.
- [19] Tate, R. F. and Goen, R. L. *Minimum Variance Unbiased Estimation for the Truncated Poisson Distribution*, Ann. Math. Stat., **29**, 755–765, 1958.
- [20] Wijsman, R. A. *On the Attainment of the Cramér-Rao Lower Bound*, Ann. Statist., **1**(3), 538–542, 1973.
- [21] Zacks, S. *Theory of Statistical Inference*, (John Wiley & Sons, N.Y. 1971).