#A15 INTEGERS 9 (2009), 167-175

GENERALIZED LEVINSON–DURBIN SEQUENCES AND BINOMIAL COEFFICIENTS

Paul Shaman
Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104-6340, USA
shaman@wharton.upenn.edu

Received: 7/3/08, Accepted: 2/21/09

Abstract

The Levinson–Durbin recursion is used to construct the coefficients which define the minimum mean square error predictor of a new observation for a discrete time, second-order stationary stochastic process. As the sample size varies, the coefficients determine what is called a Levinson–Durbin sequence. A generalized Levinson–Durbin sequence is also defined, and we note that binomial coefficients constitute a special case of such a sequence. Generalized Levinson–Durbin sequences obey formulas which generalize relations satisfied by binomial coefficients. Some of these results are extended to vector stationary processes.

1. Introduction

The Levinson–Durbin recursion is well-known and widely used in time series analysis. Suppose $\{y_t\}$ is a discrete time, second-order stationary stochastic process. Given $y_1, \ldots, y_n$, for any $n \geq 1$, the recursion obtains simply the coefficients $\alpha_{j,\,n}$ of the best linear predictor of y_{n+1}, $\hat{y}_{n+1} = -\alpha_{1,\,n} y_n - \cdots - \alpha_{n,\,n} y_1$, best in the sense of minimizing expected square error. Levinson [9] devised the recursion in connection with Norbert Wiener's work on prediction. His paper was reprinted as an appendix in Wiener's monograph on time series [17]. Wiener's work had originally been presented in February 1942 in a classified U.S. government report (see [7], p. 159, for some historical details).

Levinson presented his recursion for a stationary process with known structure. His result was rediscovered in the context of statistical estimation by Bartlett [1, pp. 264-265], Daniels [5, p. 183] and Durbin [6].

What is not well-known, and has not been described until recently, is a connection between binomial coefficients and sequences defined by Levinson's recursion. These sequences obey many formulas which are generalizations of those satisfied by binomial coefficients. The purpose of this paper is to describe this connection and present some formulas satisfied by the sequences.

In Section 2 we define the Levinson–Durbin and generalized Levinson–Durbin sequences, we note that binomial coefficients constitute a generalized Levinson–Durbin sequence, and we describe properties of the sequences. Formulas satisfied by the sequences are presented in Section 3, examples are given in Section 4, Section 5 gives an extension for vector-valued processes, and concluding remarks are in Section 6.

2. Definitions and Properties

Definition 1. The sequence $\alpha_{j,\,n},\ j=1,\ldots,n,\ n=1,2,\ldots,$ is a *Levinson–Durbin sequence* if the elements, all real-valued, satisfy

$$\alpha_{j,\,n} = \alpha_{j,\,n-1} + \alpha_{n,\,n}\alpha_{n-j,\,n-1}, \qquad j=1,\ldots,n-1,\ n=2,3,\ldots, \tag{1}$$

and

$$|\alpha_{n,\,n}| < 1, \qquad n=1,2,\ldots. \tag{2}$$

If (1) holds and the $\alpha_{n,\,n}$ are not subject to (2), the $\alpha_{j,\,n}$ are said to form a *generalized Levinson–Durbin sequence.*

The definition shows that Levinson–Durbin and generalized Levinson–Durbin sequences are completely determined by just the $\alpha_{n,\,n}$ values. If we choose $\alpha_{n,\,n}=1$ for all n, then (1) generates the binomial coefficients. In Levinson's formulation the values $-\alpha_{n,\,n}, n=1,2,\ldots,$ are taken to be the partial autocorrelations of the process to be predicted. Furthermore, if the values $-\alpha_{n,\,n},\ n=1,2,\ldots,$ are arbitrarily chosen subject to (2), they constitute the partial autocorrelation sequence of some second-order stationary process. If Γ_n denotes the $n\times n$ covariance matrix of $(y_1,\ldots,y_n)'$, then $\alpha_{j,\,k},\ j=1,\ldots,k,\ k=1,\ldots,n-1,$ can be used to form the lower triangular matrix in a Cholesky factorization of the inverse of Γ_n (see, e. g., [2], p. 491).

For $\{\alpha_{j,\,n}\}$ a generalized Levinson–Durbin sequence, define the polynomials

$$A_n(z) = \sum_{j=0}^{n} \alpha_{j,\,n} z^{n-j}, \qquad n=1,2,\ldots, \tag{3}$$

with $\alpha_{0,\,n}=1$ for all n. From (1) it follows that

$$A_n(z) = zA_{n-1}(z) + \alpha_{n,\,n} z^{n-1} A_{n-1}(z^{-1}), \qquad n=2,3,\ldots. \tag{4}$$

Proposition 2. *Let $\{\alpha_{j,\,n}\}$ be a Levinson–Durbin sequence. Then for $n=1,2,\ldots$ all the zeros of $A_n(z)$ lie strictly inside the circle $|\,z|=1$.*

This result is well-known. A proof using induction on n follows from (2), (4), and Rouché's theorem.

The following proposition notes that a symmetric generalized Levinson–Durbin sequence has trivial structure if none of the $\alpha_{n,\,n}$ is equal to 1.

Proposition 3. *Suppose $\{\alpha_{j,\,n}\}$ is a generalized Levinson–Durbin sequence for which $\alpha_{j,\,n} = \alpha_{n-j,\,n},\ j=1,\ldots,n-1,\ n=2,3,\ldots.$ If $\alpha_{n,\,n}\neq 1$, then $\alpha_{j,\,n} = \alpha_{1,\,1}/(1-(n-1)\alpha_{1,\,1}),\ j=1,\ldots,n,\ n=2,3,\ldots,$ for any choice of $\alpha_{1,\,1}$ not equal to $1/k, k=1,2,\ldots.$*

Proof. By (1) $\alpha_{j,\,n+1} = \alpha_{j,\,n} + \alpha_{n+1,\,n+1}\alpha_{n+1-j,\,n}$ and $\alpha_{n+1-j,\,n+1} = \alpha_{n+1-j,\,n} + \alpha_{n+1,\,n+1}\alpha_{j,\,n}$, $j = 1, \ldots, n$. If $\alpha_{j,\,n+1} = \alpha_{n+1-j,\,n+1}$, then $(1-\alpha_{n+1,\,n+1})\alpha_{j,\,n} = (1-\alpha_{n+1,\,n+1})\alpha_{n+1-j,\,n}$, $j = 1, \ldots, n$, and if $\alpha_{n+1,\,n+1} \neq 1$, $\alpha_{j,\,n} = \alpha_{n+1-j,\,n}$, $j = 1, \ldots, n$. This implies $\alpha_{1,\,n} = \cdots = \alpha_{n,\,n}$, $n = 2, 3, \ldots$, and the proposition follows by employing (1) for $n = 2, 3, \ldots$ in succession. □

3. Formulas

The results presented in this section hold for all generalized Levinson–Durbin sequences, and they all reduce to familiar binomial formulas when $\alpha_{n,\,n} = 1$, $n = 1, 2, \ldots$. The proofs of Theorems 4 and 5 use induction arguments and are given in [11].

Theorem 4. *If $\{\alpha_{j,\,n}\}$ is a generalized Levinson–Durbin sequence,*

$$\sum_{j=0}^{n} \alpha_{j,\,n} = \prod_{j=1}^{n} (1 + \alpha_{j,\,j}), \qquad n = 1, 2, \ldots, \tag{5}$$

$$\sum_{j=0}^{n} (-1)^j \alpha_{j,\,n} = \prod_{j=1}^{n} (1 + (-1)^j \alpha_{j,\,j}), \qquad n = 1, 2, \ldots. \tag{6}$$

Equation (5) appears in [5, p. 184]. The theorem implies that $\sum_{j=0}^{n} \alpha_{j,\,n} > 0$ and $\sum_{j=0}^{n} (-1)^j \alpha_{j,\,n} > 0$ are necessary conditions for $\{\alpha_{j,\,n}\}$ to be a Levinson–Durbin sequence.

Theorem 5. *If $\{\alpha_{j,\,n}\}$ is a generalized Levinson–Durbin sequence,*

$$\sum_{j=1}^{n} j\,\alpha_{j,\,n} = \sum_{l=1}^{n} \prod_{j=1}^{l-1} (1 + \alpha_{j,\,j}) l \alpha_{l,\,l} \prod_{k=l+1}^{n} (1 - \alpha_{k,\,k}), \quad n = 1, 2, \ldots, \tag{7}$$

$$\begin{aligned} \sum_{j=1}^{n} (-1)^{j-1} j\,\alpha_{j,\,n} = \sum_{l=1}^{n} \prod_{j=1}^{l-1} \quad & (1 + (-1)^j \alpha_{j,\,j})(-1)^{l-1}\, l\, \alpha_{l,\,l} \\ & \times \prod_{k=l+1}^{n} (1 - (-1)^k \alpha_{k,\,k}), \quad n = 1, 2, \ldots, \end{aligned} \tag{8}$$

where $\prod_1^0(\cdot) = \prod_{n+1}^{n}(\cdot) = 1$.

The next theorem follows from repeated application of (1).

Theorem 6. *If $\{\alpha_{j,\,n}\}$ is a generalized Levinson–Durbin sequence, and* $\alpha_{0,\,n} = 1$,

$n \geq 0$, *then*

$$\alpha_{j,\,n} = \sum_{k=j}^{n} \alpha_{k,\,k}\,\alpha_{k-j,\,k-1}, \qquad n > j \geq 1. \tag{9}$$

We now turn to generalizations of binomial coefficient summations such as $\binom{n}{0}+\binom{n}{2}+\binom{n}{4}+\cdots$ and $\binom{n}{0}+\binom{n}{3}+\binom{n}{6}+\cdots$. Define

$$S_{i,\,n}^{(k)} = \alpha_{i-1,\,n} + \alpha_{i+k-1,\,n} + \alpha_{i+2k-1,\,n} + \cdots, \qquad i = 1,\ldots,k,\ k = 2,3,\ldots,\ n = 1,2,\ldots,$$

where $\alpha_{0,\,n} = 1$, $n \geq 0$, and $\alpha_{j,\,n} = 0$ if $j > n$.

Theorem 7. *If $\{\alpha_{j,\,n}\}$ is a generalized Levinson–Durbin sequence, then*

$$S_{i,\,n+1}^{(k)} = S_{i,\,n}^{(k)} + \alpha_{n+1,\,n+1}\; S_{l_i,\,n}^{(k)}, \qquad l_i = \operatorname{mod}(n+2+k-i, k)+1, \quad i = 1,\ldots,k,\ k = 2,3,\ldots,\ n = 1,2,\ldots. \tag{10}$$

The proof follows immediately from (1). Some patterns may be deduced from the theorem. For example, if k is odd, there is a single value of i for which (10) holds for each n, $i = \operatorname{mod}(1+(n-1)(k+1)/2, k)+1$. However, if k is even, there are two values of i for which (10) holds for each odd value of n, $i = \operatorname{mod}(1+(n-1)(k+2)/4+jk/2,\, k)+1$, $j = 0,1$, and there is no value of i for which (10) holds for an even value of n. Tables 1 and 2 illustrate the results implied by Theorem 7 for the case of the binomial coefficients.

		i		
n	1	2	3	4
3	1	3	3	1
4	2	4	6	4
5	6	6	10	10
6	16	12	16	20
7	36	28	28	36
8	72	64	56	64
9	136	136	120	120
10	256	272	256	240

Table 1: Values of $S_{i,\,n}^{(4)}$ for binomial coefficients.

The next theorem gives closed form expressions for the sums $S_{i,\,n}^{(2)}$, $i = 1,2$.

Theorem 8. *Suppose $\{\alpha_{j,\,n}\}$ is a generalized Levinson–Durbin sequence and let $i_j = 0$ or 1, $j = 1, \ldots, n$, and $[x]$ denote the integer part of x. Then*

$$S^{(2)}_{1,\,n} = \prod_{j=1}^{[\frac{1}{2}\,n]} (1 + \alpha_{2j,\,2j}) \sum_{\text{(even)}} \prod_{k=1}^{[\frac{1}{2}\,(n+1)]} \alpha^{i_{2k-1}}_{2k-1,\,2k-1}, \qquad n = 1, 2, \ldots, \tag{11}$$

where the summation extends over all products for which $i_1 + i_3 + \cdots + i_{2[(n+1)/2]-1}$ is even; the expression for $S^{(2)}_{2,\,n}$ is the same except that the summation extends over all products for which $i_1 + i_3 + \cdots + i_{2[(n+1)/2]-1}$ is odd.

For the proof see [11].

			i		
n	1	2	3	4	5
4	1	4	6	4	1
5	2	5	10	10	5
6	7	7	15	20	15
7	22	14	22	35	35
8	57	36	36	57	70
9	127	93	72	93	127
10	254	220	165	165	220
11	474	474	385	330	385

Table 2. Values of $S^{(5)}_{i,\,n}$ for binomial coefficients.

4. Examples

Levinson–Durbin sequences arise in the context of prediction and parameter estimation for an autoregressive process of order p, defined as the stationary solution of

$$\sum_{j=0}^{p} \alpha_j\,(y_t - \mu) = \varepsilon_t, \qquad t = 0, \pm 1, \pm 2, \ldots,$$

where $\alpha_0 = 1,\ \mu = \mathrm{E}(y_t),\ \{\varepsilon_t\}$ is a sequence of independent, identically distributed random variables, each with mean 0 and variance σ^2, and the zeros of $A_p(z)$ defined at (3) lie strictly inside $|z| = 1$. For such a process the best linear predictor of y_{n+1}, given $y_1, \ldots, y_n$, yields a Levinson–Durbin sequence with $\alpha_{j,\,p} = \alpha_j,\ j = 1, \ldots, p$, and for all $n > p,\ \alpha_{j,\,n} = \alpha_j,\ j = 1, \ldots, p$, and $\alpha_{j,\,n} = 0,\ j = p+1, \ldots, n$. All of the examples which follow arise in the context of estimation of the autoregressive coefficients $\alpha_1, \ldots, \alpha_p$, given observed data $y_1, \ldots, y_T$ and assuming p is known.

Yule–Walker estimation. To obtain the Yule–Walker estimator one minimizes with
respect to $\alpha_1, \ldots, \alpha_p$ the expression $\mathrm{E}\left(\sum_{j=0}^{p} \alpha_j(y_t - \mu)\right)^2$. Then in the resulting p linear equations one replaces the covariances $\mathrm{E}(y_{t-j} - \mu)(y_{t-k} - \mu)$ by the estimates $(1/T)\sum_{t=1}^{T-|j-k|}(y_t - \bar{y})(y_{t+|j-k|} - \bar{y}),\ j = 1, \ldots, p,\ k = 0, 1, \ldots, p$, where $\bar{y}$ is the sample mean. One can also treat the case where the process mean is a polynomial time trend, $\mu(t) = \sum_{j=0}^{m-1} \beta_j t^j,\ t = 1, \ldots, T$. For this case one replaces $\bar{y}$ by the least squares estimator of $\mu(t)$. The value $m = 0$ is used to designate a known mean value. For each m, as p varies, the Yule–Walker autoregressive coefficient estimators combine to form a Levinson–Durbin sequence. This estimator is rather widely employed. It is appealing because it is easy to calculate, and because by Proposition 2 it produces an estimated model that is causal. However, the Yule–Walker estimator can have substantial bias [15]. The bias can be reduced by applying a data taper (see [4]). The tapered Yule–Walker estimator also yields a Levinson–Durbin sequence.

Least squares estimation. The least squares estimator is obtained by minimizing with respect to $\alpha_1, \ldots, \alpha_p$ the expression

$$\sum_{t=p+1}^{T} \left(y_t - \mu(t) + \alpha_1(y_{t-1} - \mu(t)) + \cdots + \alpha_p(y_{t-p} - \mu(t))\right)^2,$$

and by substituting for the mean $\mu(t)$ its least squares estimator. For each value of m and for each order p there is a unique autoregressive model for which the coefficient estimator is unbiased to order $1/T$. This model is the fixed point of a contraction mapping determined by the asymptotic bias expression for the least squares estimator. For details of the bias expression and the fixed point models see [15, 13, 10, 14]. Although the least squares estimator does not itself determine a Levinson–Durbin sequence, for each value of m the fixed point models combine as the autoregressive order p varies to form a Levinson–Durbin sequence [11]. For fixed m the sequence is defined, for $p = 1, 2, \ldots$, by

$$\alpha^m_{p,\,p} = \begin{cases} \frac{m}{p+m+1} & \text{for } p \text{ odd,} \\ \frac{m+1}{p+m+1} & \text{for } p \text{ even.} \end{cases}$$

Note that as $m \to \infty$ the coefficients of the fixed point models tend to the binomial coefficients.

Burg and Kay estimators. Estimators of the autoregressive coefficients proposed by Burg (see [3], p. 147) and Kay [8] are determined by defining the sequence $\{\alpha_{n,\,n}\}$ and employing (1). Thus these estimators form Levinson–Durbin sequences.

5. Levinson–Durbin–Whittle Sequences

Whittle [16] generalized the Levinson–Durbin recursion for a vector stationary process $\{\mathbf{y}_t\}$. Unlike a scalar Gaussian stationary process, a vector-valued Gaussian stationary process is not in general time-reversible. This creates complications, and thus Whittle's generalization of the Levinson–Durbin recursion requires the specification of two sequences, one for each direction of time. Assume that the process $\{\mathbf{y}_t\}$ has r components.

Definition 9. The $r \times r$ matrices $\mathbf{A}_{j,\,n}$ and $\bar{\mathbf{A}}_{j,\,n}$, $j = 1, \ldots, n$, $n = 1, 2, \ldots,$ form a *Levinson–Durbin–Whittle sequence* if they satisfy

$$\begin{aligned} \mathbf{A}_{j,\,n} &= \mathbf{A}_{j,\,n-1} + \mathbf{A}_{n,\,n}\bar{\mathbf{A}}_{n-j,\,n-1}, \\ \bar{\mathbf{A}}_{j,\,n} &= \bar{\mathbf{A}}_{j,\,n-1} + \bar{\mathbf{A}}_{n,\,n}\mathbf{A}_{n-j,\,n-1}, \qquad j = 1, \ldots, n-1,\ n = 2, 3, \ldots, \end{aligned} \tag{12}$$

and

$$\text{the eigenvalues of } \mathbf{A}_{n,\,n} \text{ and of } \bar{\mathbf{A}}_{n,\,n} \text{ are less than 1 in magnitude, } n = 1, 2, \ldots. \tag{13}$$

If (12) holds and (13) does not, the matrices $\mathbf{A}_{j,\,n}$ and $\bar{\mathbf{A}}_{j,\,n}$ form a *generalized Levinson–Durbin–Whittle sequence.*

If $\mathbf{A}_{n,\,n} = \bar{\mathbf{A}}_{n,\,n} = \mathbf{I}$, $n = 1, 2, \ldots,$ then (12) generates the binomial sequence

$$\mathbf{A}_{j,\,n} = \bar{\mathbf{A}}_{j,\,n} = \binom{n}{j}\mathbf{I}, \qquad j = 1, \ldots, n-1,\ n = 2, 3, \ldots,$$

where $\mathbf{I}$ is the $r \times r$ identity matrix.

Let $\mathbf{A}_{0,\,n} = \bar{\mathbf{A}}_{0,\,n} = \mathbf{I}$, and define the polynomials

$$\mathbf{A}_n(z) = \sum_{j=0}^{n} \mathbf{A}_{j,\,n}\, z^{n-j}, \qquad \bar{\mathbf{A}}_n(z) = \sum_{j=0}^{n} \bar{\mathbf{A}}_{j,\,n}\, z^{j-n}, \qquad n = 1, 2, \ldots.$$

Whittle [16] proved that the zeros of $|\,\mathbf{A}_n(z)|$ $(|\,\bar{\mathbf{A}}_n(z)|)$ all lie inside (outside) the unit circle $|\,z| = 1$.

Properties of generalized Levinson–Durbin–Whittle sequences are discussed in [12]. The analogues of Theorems 4, 5, and 8 are given there. The analogue of Theorem 6 is

$$\mathbf{A}_{j,\,n} = \sum_{k=j}^{n} \mathbf{A}_{k,\,k}\bar{\mathbf{A}}_{k-j,\,k-1}, \qquad \bar{\mathbf{A}}_{j,\,n} = \sum_{k=j}^{n} \bar{\mathbf{A}}_{k,\,k}\mathbf{A}_{k-j,\,k-1}, \qquad n > j \geq 1.$$

The formulas satisfied by generalized Levinson–Durbin–Whittle sequences simplify for time-reversible vector processes. These processes, however, are very restrictive. For example, all of their components are exactly in phase with each other.

6. Concluding Remarks

In this paper we have examined properties of generalized Levinson–Durbin sequences. The sequences are completely determined by the $\alpha_{n,n}$ values, and they obey formulas which generalize relations satisfied by binomial coefficients. The correspondence between generalized Levinson–Durbin sequences and binomial coefficients, however, is limited by symmetry considerations. The binomial coefficients essentially are the only generalized Levinson–Durbin sequence which satisfies the symmetry condition $\alpha_{j,n} = \alpha_{n-j,n}$. Many binomial coefficient formulas depend upon this symmetry condition, and thus the collection of formulas satisfied by all generalized Levinson–Durbin sequences is less expansive than that for the binomial coefficients.

Finally, we note that the results of this paper have implications for the numerical structure of autoregressive coefficient estimates for an estimator which forms a Levinson–Durbin sequence. Examples of several such estimators are given in Section 4.

References

[1] M. S. Bartlett, *An Introduction to Stochastic Processes*, Cambridge University Press, Cambridge, 1955.

[2] K. N. Berk, Consistent autoregressive spectral estimates, *Ann. Statist.* **2** (1974) 489-502.

[3] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed., Springer-Verlag, New York, 2002.

[4] R. Dahlhaus, Parameter estimation of stationary processes with spectra containing strong peaks, in *Robust and Nonlinear Time Series Analysis*, J. Franke et al., eds., Springer-Verlag, New York, 1984, 50-67.

[5] H. E. Daniels, The approximate distribution of serial correlation coefficients, *Biometrika* **43** (1956) 169-185.

[6] J. Durbin, The fitting of time-series models, *Rev. Inst. Internat. Statist.* **28** (1960) 233-244.

[7] T. Kailath, A view of three decades of linear filtering theory, *IEEE Trans. Information Theory* **20** (1974) 146-181.

[8] S. M. Kay, Recursive maximum likelihood estimation of autoregressive processes, *IEEE Trans. Acoust. Speech Signal Process.* **31** (1983) 56-65.

[9] N. Levinson, The Wiener RMS (root-mean-square) error criterion in filter design and prediction, *J. Math. Phys. Mass. Inst. Tech.* **25** (1947) 261-278.

[10] D. T. Pham, On the asymptotic expansions for the bias and covariance matrix of autoregressive estimators, in *Developments in Time Series Analysis*, T. Subba Rao, ed., Chapman and Hall, London, 1993, 80-100.

[11] P. Shaman, Generalized Levinson–Durbin sequences, binomial coefficients and autoregressive estimation, 2008, submitted for publication.

[12] P. Shaman, Properties of generalized Levinson–Durbin–Whittle sequences, *J. Statist. Plann. Inference* **138** (2008) 2808-2814.

[13] P. Shaman and R. A. Stine, The bias of autoregressive coefficient estimators, *J. Amer. Statist. Assoc.* **83** (1988) 842-848.

[14] R. A. Stine and P. Shaman, A fixed point characterization for bias of autoregressive estimators, *Ann. Statist.* **17** (1989) 1275-1284.

[15] D. Tjøstheim and J. Paulsen, Bias of some commonly-used time series estimates, *Biometrika* **70** (1983) 389-399.

[16] P. Whittle, On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix, *Biometrika* **50** (1963) 129-134.

[17] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series, with Engineering Applications*, MIT Press, Cambridge, MA, 1949.