THEORY OF CANONICAL MOMENTS AND ITS APPLICATIONS IN POLYNOMIAL REGRESSION - PART I*

by

Tai Shing Lau
Purdue University

Technical Report #83-23

Department of Statistics
Purdue University

August 1983

*This research was supported by NSF Grant No. MCS-8200631 and MCS 81-0167-A1

CHAPTER I

INTRODUCTION

1.1 Optimal Design for Polynomial Regression

Consider the linear model

$$y(x) = \beta_0 + \beta_1 x + \ldots + \beta_m x^m + \varepsilon = \beta^T f(x) + \varepsilon \qquad (1.1.1)$$

where $x \in [a,b]$, $\beta_0, \beta_1, \ldots, \beta_m$ are unknown parameters, and ε is a random variable with mean 0 and variance σ^2 which is independent of x and σ^2 may be known or unknown.

Suppose the experimenter wants to estimate β. Given N uncorrelated observations at $x_1, x_2, \ldots, x_N$, it is known that if

$$X = \begin{pmatrix} 1 & x_1 & \ldots & x_1^m \\ \vdots & \vdots & & \vdots \\ 1 & x_N & \ldots & x_N^m \end{pmatrix}$$

and $X^T X$ is nonsingular then the least squares estimator $\hat{\beta} = (X^T X)^{-1} X^T Y$ is the minimum variance unbiased estimator of β and the covariance matrix of $\hat{\beta}$ is given by

$$E(\hat{\beta}-\beta)(\hat{\beta}-\beta)^T = \sigma^2 (X^T X)^{-1}.$$

Suppose the experimenter agrees to use the least squares estimator. The problem now is how to choose $x_1, \ldots, x_N$ such that the matrix $(X^T X)^{-1}$ is "minimized" in some sense. A design or choice of the N points found in this way is called an exact design. Due to the fine structure of these designs, the problem is hard to solve. Let us write

$$M(\xi) = \frac{1}{N} X^T X = \int_a^b f(x)f(x)^T d\xi(x)$$

where $d\xi(x)$ denote the proportion of observations taken at the point x. Kiefer and Wolfowitz (1959) extended the class of exact ξ to the set of all probability measures on [a,b]. So the problem becomes choosing a probability measure ξ so that the information matrix $M(\xi)$ is "maximized". The resulting measure (design) is the so-called approximate design. It was shown by Kiefer (1961) that from the approximate design an exact design can be obtained which is optimal to within order $\frac{1}{N}$. In this thesis, we will restrict our discussion to approximate designs only.

There are many criteria that have been introduced to measure the "size" of the information matrix $M(\xi)$. For example, the Φ_p optimal design is the one that minimizes

$$\left(\frac{1}{m+1} \operatorname{tr} M^{-p}(\xi)\right)^{1/p} = \left(\frac{1}{m+1} \sum_{i=1}^{m+1} \lambda_i^{-p}\right)^{1/p} \qquad (1.1.2)$$

where λ_i are the eigenvalues of $M(\xi)$.

The case p = 1 can be interpreted as the design that minimizes the trace of the covariance matrix and is called the A-optimal design. If $p \to 0$ ($p \to \infty$, resp.), (1.1.2) becomes $|M^{-1}(\xi)|$, the determinant of $M^{-1}(\xi)$ (the maximum eigenvalue of $M^{-1}(\xi)$, resp.). So the Φ_0(Φ_∞ resp.) optimal design can be interpreted as the design that minimizes the determinant (the maximum eigenvalue) of the covariance matrix. They are referred to as D-optimal (E-optimal) designs in the literature.

Kiefer (1961) proposed using D_s-optimal designs to estimate s out of m+1 parameters. The criterion is to find the design that minimizes the generalized variance of the estimator of the s parameters that are of interest given the presence of the other parameters. It is also used to detect the presence of the s-highest coefficients. For more details about the criteria discussed above, see Kiefer (1974).

Federov (1972) introduced the linear optimal design which minimizes tr $AM^{-1}(\xi)$ for a given non-negative definite matrix A. For example the I_σ optimal design is the one that minimizes the integrated variance tr $M^{-1}(\xi)M(\sigma)$ for a given measure σ. More examples can be found in Federov (1972).

The experimenter is now faced with many possible criteria. Certainly the criteria chosen should reflect what he wants from the experiment and thus it may be varied from case to case. But there are still some common features that may be required of these designs. For example, it is quite desirable to have the design remain unchanged if we use different scales or locations of the measurements in x. Unfortunately, many commonly used criteria do not satisfy this requirement, for example, A- and E- optimality. It can be shown that the D-optimal design and the I_σ optimal design have this property (Kiefer 1959, Studden 1971). They showed that the above criteria are invariant under the linear transformations of the regression functions $(1,x,\ldots,x^m)$ and hence under scale and translation of x.

This will imply, for example, that in searching for the D-optimal and I_σ-optimal designs we may use orthogonal polynomials as regression functions instead of $1, x, \ldots, x^m$. Most of our discussions will concentrate on those designs that will remain the same under the change of measurements of x.

The above and the other appealing properties of the D-optimal design makes it popular among experimenters. It was pointed out by Guest (1958), Cook and Nachtsheim (1982) that the D-optimal design performs much worse than the uniform design and the I_σ-optimal design in the central part of the regression interval. So it is advisable to use criteria other than the D-optimality if the central part of the interval is considered more important. It seems that the I_σ-optimal design should deserve more attention.

In the last few years, many algorithms have been proposed to obtain optimal designs. (See Wynn (1970), Federov (1972) and Wu (1978) among others). It is still desirable to have analytic solutions rather than a numerical approximation. Studden (1980) (1981) used canonical moments, a sequence of numbers that characterize the design ξ on a compact set, to aid in finding the weighted D_s-optimal design. The results are elegant and provide a unified approach to some of the optimal design problems for the polynomial situation. It can be shown that the determinant of the information matrix $M(\xi)$ can be expressed as products of the canonical moments. The canonical moments are useful because each of them ranges freely over the unit interval. We can use simple calculus to obtain the D-optimal design. The resulting design is now expressed in terms of the canonical moments. It can be shown that the canonical moments are invariant under linear transformations of the $y = ax+b$ on the regression interval, where $a > 0$. This property matches with the same property of the D-optimal design. Usually the design is given by their support points and the weights attached to it. So it is necessary to discuss the relations between support, weights and canonical moments. Actually, a more general question can be asked: given a sequence (infinite or finite) of canonical moments, what can we say about the corresponding measures. Indeed, given one of the following, the others will be uniquely determined: a probability measure ξ on $[a,b]$, a system of orthogonal polynomials on $[a,b]$, a sequence of moments, a sequence of canonical moments, the Stieltjes transform of ξ and the corresponding continued fraction expansion. The limiting behavior of the canonical moments plays an important role in determining the properties of the corresponding measure. It can be seen that the canonical moments are very useful in the optimal design problems. For example, in comparing two regression equations that may have the same intercept, the experimenter wants to estimate the difference of the two regression equations given that they have the same intercept. If the D-optimality criterion is applied, we can find the D-optimal design easily with the help of the canonical moments.

In setting up the model (1.1.1), the experimenter may not know whether the model is true or not. So it is natural to require the design to give a check to the model (1.1.1). Unfortunately, the classical optimal designs do not give us satisfactory answers. Several methods have been proposed, see, for example, Box and Draper (1959), Atwood (1971), Broth (1975), Marcus and Sacks (1976). Stigler (1971) proposed that the design should meet 3 conditions:

a. The design should allow for a check of whether or not the assumed model provides an adequate fit to the true regression function.

b. If it is concluded that the model is adequate, it should be possible to make reasonably efficient inferences concerning that model.

c. The optimal design should not depend on unknown parameters.

Stigler proposed a new criterion that satisfies the conditions mentioned above. Unfortunately he could not give a general solution. Läuter (1974) gave a new criterion that satifies the conditions set by Stigler and claimed the calculation is easier than that of Stigler's approach. Studden (1982) used the method of canonical moments to give a general solution using Stigler's criterion. It should be remarked that the criterion introduced by Stigler is essentially a constrained optimization problem. What Läuter did is to change it back to an optimization problem without constraints.

As pointed out by Kiefer (1976), a criterion function is only an approximation to some vague notion of 'goodness'. To combine different criteria will give a better approximation to the real situation but also will bring in the problem of constrained optimization. In general, the analytic solutions for problems of this type are difficult to find. So it seems reasonable to study how an optimal design ξ^*, with respect to a criterion function Φ, performs under other criteria. Sometimes it might happen that there is another design that is only slightly less efficient in terms of Φ, but which is noticeably superior to ξ^* in terms of other criteria of interest. After the comparison of efficiencies of different criteria, maybe we can choose a "better" design. The result of our comparison seems to indicate that the uniform design is not so good as some might expect.

The trigonometric polynomial regression on the circle has been addressed. There is a sequence of parameters related to the measures on the circle that is analogous to the canonical moments. They are related to canonical moments in a certain way if the measure on the circle is symmetric about the x-axis. The corresponding optimal design problems on the circle are also considered.

1.2 Outline of the Thesis

Chapter II will discuss some basic properties of the canonical moments. The relations between ordinary moments, canonical moments, Hankel determinants, continued fractions, orthogonal polynomials and measures are discussed. Some examples are given to illustrate how to obtain canonical moments from the Hankel determinants, continued fraction expansion or orthogonal polynomials. The counter part of the above on the circle is also discussed. The relations between the distributions and the limiting behavior of the canonical moments are investigated.

In Chapter III, we have a section on some admissibility results. The weighted D-optimal design is found when the weight function $w(x) = [x(1-x)]^{\alpha}|x - \frac{1}{2}|^{\gamma}$ $\alpha \geq 0$, $\gamma \geq 0$ and m is odd. An example is given to illustrate the use of weighted D-optimal design. A new proof of the D-optimal design is given for the trigonometric regression on the circle. The method of canonical moments is used to obtain D-optimal rotatable designs. Some limiting designs are discussed.

Chapter IV gives a new proof of the D_s-optimal design for the D_s-optimal design. The explicit form for the support of the weighted D_s-optimal designs when the $w(x) = x$, $1-x$ and $x(1-x)$ are obtained. Some admissibility results are given. The D_s-optimal design for the s highest even (odd) coefficients is found. Weighted D_s-optimal designs for the weight function $w(x) = (1-x^2)|x|^2$ and $|x|^2$ are given. D_s-optimal designs for trigonometric regression and rotatable design are considered. D_s-optimal designs when $s = 1$ for some special weight functions are found.

The I_σ-optimal design and the optimal extrapolation design are considered in Chapter V. The I_σ-optimal design for some σ is given. The optimal weighted extrapolation designs are given for some weight functions. The continued fraction expansion of the Stieltjes transform of the optimal extrapolation design ($w(x) = 1$) is obtained.

In Chapter VI the comparison of models is discussed and the related D-optimal designs are given explicitly.

In Chapter VII the robust-type D-optimal design and some limiting designs are discussed. The performance of different designs under different criteria are compared.

CHAPTER II

THEORY OF CANONICAL MOMENTS

In this chapter we will introduce the theory of canonical moments, which is closely related to the theories of continued fractions and orthogonal polynomials. We will give a brief introduction to the last two subjects in sections 3 and 4. A detailed account of the theory of continued fractions can be found in Perron (1913) and Wall (1948). For the theory of orthogonal polynomials, see Szegö (1975), Geronimus (1948), (1950), (1961a), (1961b), Freud (1971), Chihara (1978), and Nevai (1979). Compared to the above two subjects, there has not been much explicit work published on the theory of canonical moments, except for the papers of Skibinsky (1967), (1968), (1969). The purpose of this chapter is to give a rather detailed account of canonical moments. Some of the results are implicitly given in the theory of continued fractions or orthogonal polynomials and some of the results are new. They are presented here in a form usable in design theory. A few of the results are not actually applied in later chapters but are included for completeness and possible use in the future. It is possible to read the later chapters first, referring to this chapter whenever needed.

2.1 Canonical Moments

In this section we will give the definition of canonical moments and some of its basic properties. We will start with the ordinary moments of measures on [0,1].

Definition 2.1.1: The moment space $\mathcal{M}_{n+1}$ with respect to $f^T(x) = (1,x\ldots,x^n)$ is defined as

$$\mathcal{M}_{n+1} = \{\mu = \int_0^1 f(x)d\xi(x) \mid \xi \text{ is a finite measure}\}.$$

Denote the set of all nonnegative polynomials (on [0,1]) of degree n by $\mathcal{P}_{n+1}$. Note that $\mathcal{P}_{n+1}$ can be identified as a subset of R_{n+1}. We have the following important characterization theorem of $\mathcal{M}_{n+1}$ in terms of $\mathcal{P}_{n+1}$.

Theorem 2.1.1:

(i) $\mu \in \mathcal{M}_{n+1}$ iff $a^T\mu \geq 0$ for all $a \in \mathcal{P}_{n+1}$.

(ii) $\mu \in \text{Int}\,\mathcal{M}_{n+1}$ (interior of $\mathcal{M}_{n+1}$) iff $a^T\mu > 0$ for all $a \in \mathcal{P}_{n+1}\backslash\{0\}$.

Proof: See Karlin and Studden (1966a).

The following theorem will give the form of elements in $\mathcal{P}_{n+1}$.

Theorem 2.1.2: (Markov and Lukacs) Any polynomial $P_n(x)$ in $\mathcal{P}_{n+1}$ has the form

$$P_n(x) = \begin{cases} (A_k(x))^2 + x(1-x)(B_{k-1}(x))^2 & \text{for } n = 2k, \\ x(C_k(x))^2 + (1-x)(D_k(x))^2 & \text{for } n = 2k+1. \end{cases}$$

where $A_k(x)$, $B_{k-1}(x)$, $C_k(x)$ and $D_k(x)$ are polynomials of the degrees indicated.

Proof: See Krein and Nudel'man (1977).

Let

$$\underline{M}_{2n} = \begin{bmatrix} \mu_0 & \mu_1 & \cdots \mu_n \\ \mu_1 & \mu_2 & \cdots \mu_{n+1} \\ \vdots & \vdots & \vdots \\ \mu_n & \mu_{n+1} & \cdots \mu_{2n} \end{bmatrix} \qquad \overline{M}_{2n} = \begin{bmatrix} \mu_1-\mu_2 & \cdots \mu_n-\mu_{n+1} \\ \mu_2-\mu_3 & \cdots \mu_{n+1}-\mu_{n+2} \\ \vdots & \vdots \\ \mu_n-\mu_{n+1} & \cdots \mu_{2n-1}-\mu_{2n} \end{bmatrix}$$

$$\underline{M}_{2n+1} = \begin{bmatrix} \mu_1 & \mu_2 & \cdots \mu_{n+1} \\ \mu_2 & \mu_3 & \cdots \mu_{n+2} \\ \vdots & \vdots & \vdots \\ \mu_{n+1} & \mu_{n+2} & \cdots \mu_{2n+1} \end{bmatrix} \qquad \overline{M}_{2n+1} = \begin{bmatrix} \mu_0-\mu_1 & \cdots \mu_n-\mu_{n+1} \\ \mu_1-\mu_2 & \cdots \mu_{n+1}-\mu_{n+2} \\ \vdots & \vdots \\ \mu_n-\mu_{n+1} & \cdots \mu_{2n}-\mu_{2n+1} \end{bmatrix}$$

where μ_i is the ith moment of ξ. Note that $\mu^T = (\mu_0, \mu_1, \ldots, \mu_n)$ for $\mu \in \text{Int}\, \mathcal{M}_{n+1}$. Denote the determinants of the above matrices by $\underline{D}_{2n}$, $\underline{D}_{2n+1}$, $\overline{D}_{2n}$ and $\overline{D}_{2n+1}$ respectively. The subscripts indicate the highest moment involved in the matrices or determinants. Now, Theorem 2.1.1 and Theorem 2.1.2 give the following useful theorem.

Theorem 2.1.3:

(i) $\mu \in \mathcal{M}_{n+1}$ iff $\underline{M}_n$ and $\overline{M}_n$ are positive semi-definite.

(ii) $\mu \in \text{Int}\, \mathcal{M}_{n+1}$ iff $\underline{D}_0$, $\overline{D}_0$, $\underline{D}_1$, $\overline{D}_1 \ldots, \underline{D}_n$, $\overline{D}_n$ are positive.

Proof: See Karlin and Studden (1966a).

It is known that given $\mu^0 \in \text{Int}\, \mathcal{M}_n$, there exists infinitely many representations for μ^0. Let $v(\mu^0)$ denote the convex set of all measures representing μ^0. It is known that $\mu_n^+ = \max_{v(\mu^0)} \mu_n$ and $\mu_n^- = \min_{v(\mu^0)} \mu_n$ are finite. Since $\bar{\mu} = (\mu^0, \mu_n^+)$ and $\underline{\mu} = (\mu^0, \mu_n^-)$ are on the boundary of $\mathcal{M}_{n+1}$, μ_n^+ (resp. μ_n^-) is the solution of $\overline{D}_n = 0$ (resp. $\underline{D}_n = 0$) by Theorem 2.1.3, where μ_n is taken as a variable. Thus, we have proved the following result.

Theorem 2.1.4: Given $\mu^0 = (\mu_0, \mu_1, \ldots, \mu_{n-1}) \in \text{Int}\,\mathcal{M}_n$, μ_n^+(resp. μ_n^-) is the solution of $\overline{D}_n = 0$ (resp. $\underline{D}_n = 0$).

We now give the definition of the canonical moments.

Definition 2.1.2. The canonical moment p_n is given by

$$p_n = \frac{\mu_n - \mu_n^-}{\mu_n^+ - \mu_n^-}$$

if $\mu_n^+ - \mu_n^- \neq 0$ $n = 1,2,\ldots$. p_n is undefined if $\mu_n^+ - \mu_n^- = 0$.

So p_n can be interpreted as the ratio of the distance of μ_n from the lower boundary to the distance between the upper and lower boundaries of the corresponding convex set $v(\mu^0)$. It is clear that $\mu \in \mathcal{M}_{n+1}$ is a boundary point iff $p_i = 0$ or 1 for some $i \in \{1,2,\ldots,n\}$. In this case $\mu_{i+1}^+ - \mu_{i+1}^- = 0$ and $p_{i+1}, p_{i+2}, \ldots$ are undefined. On the other hand, $\mu \in \text{Int}\,\mathcal{M}_{n+1}$ iff $0 < p_i < 1$ for $i = 1,2,\ldots,n$.

It is easy to see that $\mu_1^+ = \mu_0$ and $\mu_1^- = 0$ for a given $\mu_0 (\neq 0)$, hence $p_1 = \frac{\mu_1}{\mu_0}$. For given μ_0 and μ_1, we see that $\mu_2^+ = \mu_1$ and $\mu_2^- = \frac{\mu_1^2}{\mu_0}$ by the Cauchy-Schwarz inequality, $p_2 = \frac{\mu_0\mu_2 - \mu_1^2}{\mu_1(\mu_0 - \mu_1)}$. In Section 2, we will give a general expression for p_n in terms of Hankel determinants. It will be seen that all the points in the form $\lambda\mu$, where $\lambda > 0$ and $\mu \in \mathcal{M}_{n+1}$, have the same $p_1, p_2, \ldots, p_n$. If we assume $\mu_0 = 1$, i.e. if we restrict ourselves to probability measures, then we can define a 1-1 mapping between $p_1, p_2, \ldots, p_n$ and $\mu_1 \mu_2, \ldots, \mu_n$ for all n. The explicit expression of μ_n in terms of p_1, $p_2, \ldots, p_n$ will be given in section 5.

Let $y = (b-a)x + a$, where $b > a$, be a linear transformation from $[0,1]$ to $[a,b]$. If ξ is a probability measure on $[0,1]$, there is an induced probability measure ξ' on $[a,b]$. We can define the canonical moments of ξ' according to Definition 2.1.2. Skibinsky (1969) proved that the canonical moments of ξ' are the same as those of ξ.

Theorem 2.1.5: The canonical moments are invariant under the linear transformation $y = (b-a)x+a$, where $b > a$.

Proof: The proof was given by Skibinsky and it is included here for the sake of completeness. The n-th moment of ξ' is given by

$$\mu_n' = \int_a^b t^n d\xi'(t) = \int_0^1 [(b-a)t+a]^n d\xi(t)$$

$$= (b-a)^n \mu_n + \sum_{i=0}^{n-1} \binom{n}{i} (b-a)^i \mu_i a^{n-i},$$

where μ_n is the n-th moment of ξ. It is easily checked that

$$\mu_n' - \mu_n'^- = (b-a)^n (\mu_n - \mu_n^-)$$

$$\mu_n'^+ - \mu_n'^- = (b-a)^n (\mu_n^+ - \mu_n^-).$$

Hence the result follows.

Corollary 2.1.5: If $b < a$ in Theorem 2.1.5, the even canonical moments of ξ, say p_{2n}', will remain unchanged while the odd canonical moments p'_{2n+1} are given by

$$p'_{2n+1} = 1 - p_{2n+1}.$$

Proof: The proof is similar to that of Theorem 2.1.5 except that now μ'^{+}_{2n+1} corresponds to μ^{-}_{2n+1} and μ'^{-}_{2n+1} corresponds to μ^{+}_{2n+1}.

In particular, if a = 1 and b = 0 and ξ is symmetric about $\frac{1}{2}$, then $p_{2n+1} = q_{2n+1}$, i.e. $p_{2n+1} = \frac{1}{2}$. On the other hand, if all odd canonical moments are equal to one half, then the transformation will now give the same measure which shows that ξ is symmetric. Thus we have proved

Theorem 2.1.6: ξ is symmetric about the midpoint of its support iff $p_{2n+1} = \frac{1}{2}$ for all n whenever p_n is defined.

The next theorem shows how the canonical moments change under a special nonlinear transformation of x.

Theorem 2.1.7: Let ξ be a measure on [0,1] and ξ' be a symmetric measure on [-1,1] such that $\xi'[-x,x] = \xi[0,x^2]$. Let μ, μ_n, p_n and p'_n denote their corresponding moments and canonical moments, then we have

$$p'_{2n+1} = \frac{1}{2} \quad \text{and} \quad p'_{2n} = p_n.$$

Proof: By direct verification, we see $\mu'_{2n} = \mu_n$ and $\mu'_{2n+1} = 0$. It follows that $\mu'^{+}_{2n} = \mu^{+}_{n}$ and $\mu'^{-}_{2n} = \mu^{-}_{n}$. Thus we have $p'_{2n} = p_n$. The fact that $p'_{2n+1} = \frac{1}{2}$ follows from Theorem 2.1.6.

The results obtained here have been proven by Wall (1948) among the others. Corollary 2.2.2, whose proof is new, is a crucial step in establishing Theorem 2.2.1.

2.2 The Hankel Determinants and Canonical Moments

In this section, we will show how the Hankel determinants and canonical moments can be expressed in terms of each other. We first try to express $\mu_n - \mu_n^-$ and $\mu_n^+ - \mu_n^-$ in terms of canonical moments.

Lemma 2.2.1:
$$\mu_n - \mu_n^- = \frac{\underline{D}_n}{\underline{D}_{n-2}}, \qquad n \geq 1 \tag{2.2.1}$$

$$\mu_n^+ - \mu_n = \frac{\bar{D}_n}{\bar{D}_{n-2}}, \qquad n \geq 1. \tag{2.2.2}$$

Here we assume $\underline{D}_{-1} = \bar{D}_{-1} = \bar{D}_0 = 1$.

Proof: By Theorem 2.1.4, μ_n^- is the solution of $\underline{D}_n = 0$ where μ_n is taken as a variable. Assume $n = 2k$, we have

$$\underline{D}_{2k} = \begin{vmatrix} \mu_0 & \cdots & \mu_k \\ \vdots & & \vdots \\ \mu_k & \cdots & \mu_{2k} \end{vmatrix} = \begin{vmatrix} \mu_0 & \cdots & \mu_k \\ \vdots & & \vdots \\ \mu_k & \cdots & \mu_{2k}^- \end{vmatrix} + \begin{vmatrix} \mu_0 & \cdots & \mu_{k-1} & 0 \\ \vdots & & \vdots & \vdots \\ \mu_{k-1} & \cdots & \mu_{2k-2} & 0 \\ \mu_k & \cdots & \mu_{2k-1} & \mu_{2k} - \mu_{2k}^- \end{vmatrix}$$

$$= 0 + (\mu_{2k} - \mu_{2k}^-)\underline{D}_{2k-2},$$

and the above result follows immediately. The case $n = 2k+1$ can be proved similarly. The proof for (2.2.2) is the same.

Definition 2.2.1: Let $\Delta^m\mu_n$ denote $\int_0^1 x^n(1-x)^m d\xi$ where m,n are nonnegative integers. Define

$$H_k(\Delta^m\mu_n) = \begin{vmatrix} \Delta^m\mu_n & \cdots & \Delta^m\mu_{n+k-1} \\ \vdots & & \vdots \\ \Delta^m\mu_{n+k-1} & \cdots & \Delta^m\mu_{n+2k-2} \end{vmatrix}$$

for $k \geq 1$. If $k = 0$, we let $H_k(\Delta^m\mu_n) = 1$ for all m and n.

Let

$$H_k^{(m,n)}(x) = \begin{vmatrix} \Delta^m\mu_n & \cdots & \Delta^m\mu_{n+k} \\ \vdots & & \\ \Delta^m\mu_{n+k-1} & \cdots & \Delta^m\mu_{n+2k-1} \\ 1 & & x^k \end{vmatrix}, \text{ if } H_k(\Delta^m\mu_n) \neq 0,$$

for $k = 1,2,\ldots$ Define $H_k^{(m,n)}(x) = 1$ for all nonnegative integers m,n.

Remark 2.2.1: If $m = 0$, $H_k(\Delta^0\mu_n) = H_k(\mu_n)$ is the so-called Hankel determinant and $H_k^{(0,n)}(x)$ is the so-called Hankel polynomial. With regard to the notation in section 2.1, we have

$$H_{k+1}(\mu_0) = \underline{D}_{2k}, \quad H_{k+1}(\mu_1) = \underline{D}_{2k+1},$$
$$H_{k+1}(\Delta\mu_0) = \bar{D}_{2k+1}, \; H_{k+1}(\Delta\mu_1) = \bar{D}_{2k+2}.$$

The following lemma is given in Henrici (1973).

Lemma 2.2.2: For all nonnegative integers n and m, all $k \geq 1$,

$$x\; xH_{k+1}(\Delta^m\mu_n)H_k^{(m,n+1)}(x) = H_k(\Delta^m\mu_{n+1})H_{k+1}^{(m,n)}(x)$$

$$+ H_{k+1}(\Delta^m \mu_{n+1}) H_k^{(m,n)}(x), \tag{2.2.3}$$

$$H_{k+1}(\Delta^m \mu_{n+1}) H_{k+1}^{(m,n)}(x) = H_{k+2}(\Delta^m \mu_n) H_k^{(m,n+1)}(x)$$

$$+ H_{k+1}(\Delta^m \mu_n) H_{k+1}^{(m,n+1)}(x). \tag{2.2.4}$$

Proof: The case $m = 0$ would be reduced to that of Henrici (1973). For $m > 0$, the proof is the same and it is omitted.

We have the following interesting and useful corollary.

Corollary 2.2.2: $\underline{D}_n \bar{D}_n = \underline{D}_{n-1} \bar{D}_{n+1} + \bar{D}_{n-1} \underline{D}_{n+1}$, where $n \geq 0$.

Proof: Assume $n = 2k$, put $x = 1$ in (2.2.3) and let $m = 0$, we obtain

$$\begin{vmatrix} \mu_0 & \cdots & \mu_k \\ \vdots & & \vdots \\ \mu_k & \cdots & \mu_{2k} \end{vmatrix} \begin{vmatrix} \mu_1 & \cdots & \mu_{k+1} \\ \vdots & & \vdots \\ \mu_k & \cdots & \mu_{2k} \\ 1 & & 1 \end{vmatrix} = \begin{vmatrix} \mu_1 & \cdots & \mu_k \\ \vdots & & \vdots \\ \mu_k & \cdots & \mu_{2k-1} \end{vmatrix} \begin{vmatrix} \mu_0 & \cdots & \mu_{k+1} \\ \vdots & & \vdots \\ \mu_k & \cdots & \mu_{2k+1} \\ 1 & & 1 \end{vmatrix}$$

$$+ \begin{vmatrix} \mu_1 & \cdots & \mu_{k+1} \\ \vdots & & \vdots \\ \mu_{k+1} & \cdot & \mu_{2k+1} \end{vmatrix} \begin{vmatrix} \mu_0 & \cdots & \mu_k \\ \vdots & & \vdots \\ k-1 & \cdots & 2k-1 \\ 1 & & 1 \end{vmatrix}$$

By using column operations, we see the result immediately. The case when n is odd can be proved similarly, and the proof is, therefore, omitted.

Lemma 2.2.3: $\mu_n^+ - \mu_n^- = \dfrac{\underline{D}_{n-1} \bar{D}_{n-1}}{\underline{D}_{n-2} \bar{D}_{n-2}}$, $n \geq 1$.

Proof: By Lemma 2.2.1, we see that

$$\mu_n^+ - \mu_n^- = \frac{\bar{D}_n}{\bar{D}_{n-2}} + \frac{\underline{D}_n}{\underline{D}_{n-2}}$$

$$= \frac{\underline{D}_{n-2}\bar{D}_n + \underline{D}_n\bar{D}_{n-2}}{\bar{D}_{n-2}\,\underline{D}_{n-2}}$$

$$= \frac{\underline{D}_{n-1}\,\bar{D}_{n-1}}{\underline{D}_{n-2}\,\bar{D}_{n-2}}.$$

We use Corollary 2.2.2 in the last step.

Theorem 2.2.1: For $p \geq 1$, we have $p_n = \dfrac{\underline{D}_n\,\bar{D}_{n-2}}{\underline{D}_{n-1}\,\bar{D}_{n-1}}$ and

$$q_n = 1-p_n = \frac{\underline{D}_{n-2}\,\bar{D}_n}{\underline{D}_{n-1}\,\bar{D}_{n-1}} .$$

Proof: Use Lemma 2.2.1 and Lemma 2.2.3.

So we can express the canonical moments in terms of Hankel determinants. Now we want to do the converse.

Lemma 2.2.4: $\mu_n^+ - \mu_n^- = \prod_{i=1}^{n-1} p_i q_i$ for $n \geq 2$.

Proof: By Lemma 2.2.3, we have

$$\mu_n^+ - \mu_n^- = \frac{\underline{D}_{n-1}\,\bar{D}_{n-1}}{\underline{D}_{n-2}\,\bar{D}_{n-2}}$$

$$= p_{n-1}q_{n-1}\frac{\underline{D}_{n-2}\,\bar{D}_{n-2}}{\underline{D}_{n-3}\,\bar{D}_{n-3}}$$

$$= p_{n-1}q_{n-1}\,(\mu_{n-1}^+ - \mu_{n-1}^-).$$

Hence the result follows immediately.

Remark 2.2.3: The quantity $\mu_n^+ - \mu_n^-$ is called the range of μ_n. The expression above is useful in many extremal problems.

Let us introduce the following notations:

$$\zeta_1 = p_1 \quad \text{and} \quad \zeta_n = q_{n-1}p_n \text{ for } n \geq 2,$$

$$\gamma_1 = q_1 \quad \text{and} \quad \gamma_n = p_{n-1}q_n \text{ for } n \geq 2.$$

Lemma 2.2.5:

$$\zeta_n = \frac{\underline{D}_{n-3}\,\underline{D}_n}{\underline{D}_{n-2}\,\underline{D}_{n-1}}, \qquad n \geq 2,$$

$$\gamma_n = \frac{\bar{D}_{n-3}\,\bar{D}_n}{\bar{D}_{n-2}\,\bar{D}_{n-1}}, \qquad n \geq 2.$$

Proof: Use Theorem 2.2.1.

Lemma 2.2.6: $\mu_n - \mu_n^- = \prod_{i=1}^{n} \zeta_i, \quad n \geq 1$

$$\mu_n^+ - \mu_n^- = \prod_{i=1}^{n} \gamma_i, \quad n \geq 2.$$

Proof: Use Lemma 2.2.4 and the definition of p_n.

Theorem 2.2.2: $\underline{D}_{2n} = \prod_{i=1}^{n} (\zeta_{2i-1}\zeta_{2i})^{n-i+1},$

$$\underline{D}_{2n+1} = \prod_{i=0}^{n} (\zeta_{2i}\zeta_{2i+1})^{n-i+1}, \quad \zeta_0 = 1$$

$$\bar{D}_{2n} = \prod_{i=1}^{n} (\gamma_{2i-1}\gamma_{2i})^{n-i+1},$$

$$\bar{D}_{2n+1} = \prod_{i=0}^{n} (\gamma_{2i}\gamma_{2i+1})^{n-i+1}, \quad \gamma_0 = 1.$$

Proof: We prove the first one only. By Lemma 2.2.5, we have

$$\frac{\underline{D}_{2n}}{\underline{D}_{2n-2}} = \zeta_{2n} \frac{\underline{D}_{2n-1}}{\underline{D}_{2n-3}}$$

$$= \zeta_{2n}\zeta_{2n-1} \frac{\underline{D}_{2n-2}}{\underline{D}_{2n-4}}$$

$$= \prod_{i=1}^{n} (\zeta_{2i-1}\zeta_{2i}).$$

Thus the result follows. The proof of the others are similar. The following example illustrates how to obtain canonical moments from the Hankel determinants.

Example 2.2.1. Let ξ be the measure which put mass $\frac{1}{n}$ on each of the n equally spaced points in [0,1] i.e. 0, $\frac{1}{n-1}$,..., $\frac{n-2}{n-1}$, 1. It is known that (Muir 1933)

$$\underline{D}_{2k} = (\frac{1}{n-1})^{k(k+1)}(n^2-1)^k(n^2-2^2)^{k-1}\dots(n^2-k^2)\ \frac{(1!2!\dots k!)^4}{1!2!\dots(2k+1)!}$$

where $k \leq n-1$. We want to show by induction that $p_{2k} = \frac{k}{2k+1}\,\frac{n+k}{n-1}$ for $k = 1,\dots,n-1$. It is obvious that $p_{2i+1} = \frac{1}{2}$ for i = 0, 1, n-1 since ξ is symmetric. By Theorem 2.2.2

$$\frac{\underline{D}_{2j}}{\underline{D}_{2j-2}} = (\frac{1}{n-1})^{2j}(n^2-1)\dots(n^2-j^2)\ \frac{(j!)^4}{(2j+1)!(2j)!} = \prod_{i=1}^{j} \zeta_{2i-1}\zeta_{2i}.$$

Let j = 1, we see

$$\frac{\underline{D}_2}{\underline{D}_0} = (\frac{1}{n-1})^2(n^2-1)\ \frac{1}{3!2!} = p_1q_1p_2.$$

Hence $p_2 = \frac{1}{3}\frac{n+1}{n-1}$. Now let us assume $p_{2k-2} = \frac{k-1}{2k-1}\frac{n+k-1}{n-1}$.

Consider the ratio

$$\frac{\underline{D}_{2k}}{\underline{D}_{2k-2}} \Bigg/ \frac{\underline{D}_{2k-2}}{\underline{D}_{2k-4}} = \left(\frac{1}{n-1}\right)^2 (n^2-k^2)\,\frac{k^2}{4(2k-1)(2k+1)} = \zeta_{2k-1}\zeta_{2k}.$$

We see that $p_{2k} = \frac{k}{2k+1}\frac{n+k}{n-1}$.

2.3 Continued Fractions and Stieltjes Transform

In this section we will introduce the theory of continued fractions as a tool for investigating the properties of canonical moments. For the complete theory see Perron (1913) and Wall (1948).

Let us denote

$$\cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \ddots}}}$$

by $\frac{a_1}{b_1} + \frac{a_2}{b_2} + \frac{a_3}{b_3} + \cdots$ and its nth truncation (convergent)

$$\cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \ddots + \cfrac{a_n}{b_n}}}$$

by $\frac{a_1}{b_1} + \frac{a_2}{b_2} + \cdots + \frac{a_n}{b_n}$ or $\frac{A_n}{B_n}$. It is known that

$$A_n = b_n A_{n-1} + a_n A_{n-2} \quad \text{and}$$
$$B_n = b_n B_{n-1} + a_n B_{n-2},$$

with $A_{-1} = B_{-1} = 0$, $A_0 = 0$, $B_0 = 1$.

From the above expression, we can see that

$$A_n = a_1 \begin{vmatrix} b_2 & -1 & & \\ a_3 & b_3 & -1 & \\ & & \ddots & -1 \\ & & a_n & b_n \end{vmatrix}$$

and

$$B_n = \begin{vmatrix} b_1 & -1 & & \\ a_2 & b_2 & -1 & \\ & & \ddots & -1 \\ & & a_n & b_n \end{vmatrix}.$$

For the sake of convenience, we write $A_n = a_1 K\begin{pmatrix} a_3 & \dots & a_n \\ b_2 \; b_3 & \dots & b_n \end{pmatrix}$

and $B_n = K\begin{pmatrix} a_2 & \dots & a_n \\ b_1 & \dots & b_n \end{pmatrix}$.

The following theorem says that the ratio of the numerators (denominators) of two consecutive convergents can be expressed in terms of canonical moments.

Theorem 2.3.1:

$$\frac{A_n}{A_{n-1}} = b_n + \frac{a_n}{b_{n-1}} + \frac{a_{n-1}}{b_{n-2}} + \dots + \frac{a_3}{b_2},$$

$$\frac{B_n}{B_{n-1}} = b_n + \frac{a_n}{b_{n-1}} + \frac{a_{n-1}}{b_{n-2}} + \dots + \frac{a_2}{b_1}.$$

Proof: See Perron (1913).

Let

$$A_{v,\lambda} = K\begin{pmatrix} a_{\lambda+1} \; a_{\lambda+2} & \dots & a_{\lambda+v} \\ b_\lambda \; b_{\lambda+1} & & b_{\lambda+v} \end{pmatrix}$$

$$B_{\nu,\lambda} = K\binom{a_{\lambda+2}\ a_{\lambda+3}\ \cdots\ a_{\lambda+\nu}}{b_{\lambda+1}\ b_{\lambda+2}\ \cdots\ b_{\lambda+\nu}},$$

and $A_{\nu,0} = A_\nu$, $B_{\nu,0} = B_\nu$, $A_{-1,\lambda} = 1$, $B_{-1,\lambda} = 0$, $A_{0,\lambda} = b_\lambda$, $B_{0,\lambda} = 1$, $B_{1,\lambda} = b_{\lambda+1}$. We have

Theorem 2.3.2:

$$K\binom{a_1\ a_2\ \cdots\ a_{\nu+\lambda-1}}{b_0\ b_1\ \cdots\ b_{\nu+\lambda-1}} = K\binom{a_1\ \cdots\ a_{\lambda-1}}{b_0\ b_1\ \cdots\ b_{\lambda-1}} K\binom{a_{\lambda+1}\ \cdots\ a_{\lambda+\nu-1}}{b_\lambda\ \cdots\ b_{\lambda+\nu-1}}$$

$$+ a_\lambda K\binom{a_1\ \cdots\ a_{\lambda-2}}{b_0\ b_1\ \cdots\ b_{\lambda-2}} K\binom{a_{\lambda+2}\ \cdots\ a_{\lambda+\nu-1}}{b_{\lambda+1}\ \cdots\ b_{\lambda+\nu-1}}.$$

Proof: For the proof, we can either follow the proof of Perron (1913) or use Sylvester's identity

$$A\binom{1\ 2\ \cdots\ n}{1\ 2\ \cdots\ n} A\binom{1\ \cdots\ \lambda-1,\ \lambda+2\ \cdots\ n}{1\ \cdots\ \lambda-1,\ \lambda+2\ \cdots\ n} = A\binom{1\ \cdots\ \lambda-1\ \lambda+1\ \cdots\ n}{1\ \cdots\ \lambda-1\ \lambda+1\ \cdots\ n} A\binom{1\ \cdots\lambda\ \ \lambda+2\ \cdots\ n}{1\ \cdots\lambda\ \ \lambda+2\ \cdots\ n}$$

$$- A\binom{1\ \cdots\ \lambda-1\ \lambda+1\ \cdots\ n}{1\ \cdots\ \lambda\ \ \lambda+2\ \cdots\ n} A\binom{1\ \cdots\lambda\ \ \lambda+2\ \cdots\ n}{1\ \cdots\lambda-1\ \lambda+1\ \cdots\ n}$$

where $A\binom{1\ 2\ \cdots\ n}{1\ 2\ \cdots\ n}$ denotes the determinant $\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$ and

$A\begin{pmatrix} i_1 & \cdots & i_p \\ k_1 & \cdots & k_p \end{pmatrix}$ denotes the minor $\begin{vmatrix} a_{i_1k_1} & \cdots & a_{i_1k_p} \\ \vdots & & \vdots \\ a_{i_pk_1} & \cdots & a_{i_pk_p} \end{vmatrix}$.

The continued fraction with its nth convergent equals to $\frac{A_{2n}}{B_{2n}}$, for $n = 1,2,\ldots..$ is given by

$$\frac{a_1b_2}{b_1b_2+a_2} - \frac{a_2a_3b_4}{(b_2b_3+a_3)b_4+b_2a_4} - \frac{a_4a_5b_2b_6}{(b_4b_5+a_5)b_6+b_4a_6} - \frac{a_6a_7b_4b_8}{(b_6b_7+a_7)b_8+b_6a_8} - \cdots$$

It is called the even contraction. Correspondingly the odd contraction is given by

$$\frac{a_1}{b_1} - \frac{a_1a_2b_3}{(b_1b_2+a_2)b_3+b_1a_3} - \frac{a_3a_4b_1b_5}{(b_3b_4+a_4)b_5+b_3a_5} - \frac{a_5a_6b_3b_7}{(b_5b_6+a_6)b_7+b_5b_7} - \cdots$$

Now we define

Definition 2.3.1: The Stieltjes transform of a measure ξ is given by

$$G(z) = \int_{-\infty}^{\infty} \frac{d\xi(x)}{z-x}.$$

For the measure on [0,1], we have

Theorem 2.3.3: The Stieltjes transform of a measure ξ on [0,1] has a continued fraction expansion

$$\int_0^1 \frac{d\xi(x)}{z-x} = \frac{\mu_0}{z} - \frac{\zeta_1}{z} - \frac{\zeta_2}{z} - \cdots \qquad (2.3.1)$$

The above expression terminates with the first $\zeta_n = 0$.

Proof: See Wall (1948).

Remark 2.3.1 (see Jones and Thron (1980)): The expansion on the right of (2.3.1) is unique for a given ξ. It is called the corresponding continued fraction. If we take the even contraction, we obtain

$$\frac{1}{z-\zeta_1} - \frac{\zeta_1\zeta_2}{z-\zeta_2-\zeta_3} - \frac{\zeta_3\zeta_4}{z-\zeta_4-\zeta_5} - \cdots$$

which is called the associated continued fraction. Notice that the left hand side can be written

$$\frac{1}{z} + \frac{\mu_1}{z^2} + \frac{\mu_1}{z^3} + \ldots$$

where μ_k denotes the kth moment of ξ. The following example will show how to obtain p_n from the continued fraction.

Example 2.3.1: We want to find the canonical moments of the Jacobi distribution. It is known that the moments of the Jacobi distribution $x^\alpha(1-x)^\beta (\alpha > -1, \beta > -1)$ are given by

$$\mu_n = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} \int_0^1 t^{n+\alpha}(1-t)^\beta dt = \frac{(\alpha+1)_n}{(\alpha+\beta+2)_n},$$

where $(a)_n = a(a+1) \ldots (a+n-1)$. Compare with the hypergeometric function

$${}_2F_1(a, b, c; z) = 1 + \frac{ab}{c} z + \frac{(a)_2(b)_2}{(c)_2} \frac{z^2}{2!} + \ldots ,$$

we find the corresponding power series for the Jacobi distribution is ${}_2F_1\ (\alpha+1,\ 1,\ \alpha+\beta+2\ \ z)$. It is easy to check that

$${}_2F_1(a,b,c;\ z) = {}_2F_1(a,b+1\ ,c+1;\ z) - \frac{a(c-b)}{c(c+1)}\ z\,{}_2F_1(a+1,b+1,c+2;\ z).$$

Rearranging terms in the preceding expression, we have

$$\frac{{}_2F_1(a,b+1,c+1;\ z)}{{}_2F_1(a,b,c;\ z)} = \frac{1}{1 - \frac{a(c-b)}{c(c+1)}\ z\ \frac{{}_2F_1(a+1,b+1,c+2;z)}{{}_2F_1(a,b+1,c+1;\ z)}} . \qquad (2.3.2)$$

Noticing that $\frac{{}_2F_1(a+1,b+1,c+2;\ z)}{{}_2F_1(a,b+1,c+1;\ z)} = \frac{{}_2F_1(b+1,a+1,c+2;\ z)}{{}_2F_1(b+1,a,c+2;\ z)}$, by using (2.3.2) again, we have

$$\frac{{}_2F_1(b+1,a+1,c+2,\ z)}{{}_2F_1(b+1,a,c+1,z)} = \cfrac{1}{1 - \frac{(b+1)(c+1-b)}{(c+1)(c+2)}\, z\, \frac{{}_2F_1(b+2,a+1,c+3;\ z)}{{}_2F_1(b+1,a+1,c+2;\ z)}}.$$

Applying the same trick successively, we can expand $\frac{{}_2F_1(a,b+1,c+1,z)}{{}_2F_1(a,b,c;\ z)}$ in continued fraction,

$$\frac{{}_2F_1(a,b+1,c+1;\ z)}{{}_2F_1(a,b,c;\ z)} = \frac{1}{1} - \frac{\frac{a(c-b)z}{c(c+1)}}{1} - \frac{\frac{(b+1)(c+1-a)z}{(c+1)(c+2)}}{1} - \cdots - \frac{\frac{(b+n)(c-a+n)z}{(c+2n-1)(c+2n)}}{1} -$$

$\frac{\frac{(a+n)(c-b+n)z}{(c+2n)(c+2n+1)}}{1} - \cdots$. Let $a = \alpha+1$, $b = 0$, $c = \alpha+\beta+1$ and replace z by $\frac{1}{z}$.

We have

$$ {}_2F_1(\alpha+1,1,\alpha+\beta+2;\ \tfrac{1}{z}) = \frac{1}{1} - \frac{\frac{\alpha+1}{\alpha+\beta+2}\frac{1}{z}}{1} - \frac{\frac{\beta+1}{\alpha+\beta+2}\frac{1}{\alpha+\beta+3}\frac{1}{z}}{1} - \cdots$$

$$\frac{\frac{\beta+n}{\alpha+\beta+2n}\frac{n}{\alpha+\beta+2n+1}\frac{1}{z}}{1} - \frac{\frac{\alpha+\beta+n+1}{\alpha+\beta+\ 2n+1}\frac{\alpha+n+1}{\alpha+\beta+2n+2}\frac{1}{z}}{1} - \cdots$$

It is easy to check that $\frac{1}{z}\,{}_2F_1(\alpha+1,\ 1,\ \alpha+\beta+2,\ \frac{1}{z})$ is the Stieltjes transform of the Jacobi measure, i.e.

$$\frac{1}{z}\,{}_2F_1(\alpha+1,\ 1,\ \alpha+\beta+2;\ \tfrac{1}{z}) = \frac{1}{z} - \frac{\frac{\alpha+1}{\alpha+\beta+2}}{1} - \frac{\frac{\beta+1}{\alpha+\beta+2}\frac{1}{\alpha+\beta+3}}{z} - \cdots - \frac{\frac{\beta+n}{\alpha+\beta+2n}\frac{n}{\alpha+\beta+2n+1}}{z}$$

$$- \frac{\frac{\alpha+\beta+n+1}{\alpha+\beta+2n+1}\frac{\alpha+n+1}{\alpha+\beta+2n+2}}{1} - \cdots$$

It is easy to see

$$p_1 = \frac{\alpha+1}{\alpha+\beta+2} \quad \text{and} \quad p_2 = \frac{1}{\alpha+\beta+3}.$$

For n = 2k,

$$\zeta_{2k} = \frac{k(\beta+k)}{(\alpha+\beta+2k)(\alpha+\beta+2k+1)}.$$

Assuming that $q_{2k-1} = \frac{\beta+k}{\alpha+\beta+2k}$, we see

$$p_{2k} = \frac{k}{\alpha+\beta+2k+1} .$$

For $n = 2k+1$

$$p_{2k+1} = \frac{\alpha+\beta+k+1}{\alpha+\beta+2k+1} \frac{\alpha+k+1}{\alpha+\beta+2k+2}.$$

Since $q_{2k} = \frac{\alpha+\beta+k+1}{\alpha+\beta+2k+1}$, we see immediately

$$p_{2k+1} = \frac{\alpha+k+1}{\alpha+\beta+2k+2}.$$

This result was obtained by Skibinsky (1969) by another method.

Particularly, if $\alpha = \beta = 0$, we have the uniform measure on [0,1] and the corresponding canonical moments are given by $p_{2k-1} = \frac{1}{2}$ and $p_{2k} = \frac{k}{2k+1}$. For the arc-sine distribution, $\alpha = \beta = -\frac{1}{2}$, hence $p_k = \frac{1}{2}$ for all k.

2.4. Canonical Moments and Orthogonal Polynomials

Recall that we defined a sequence of polynomials $H_k^{(m,n)}(x)$, $k = 0,1,2,\ldots$, if $H_k(\Delta^m\mu_n) \neq 0$, in section 2. It is obvious that $H_k^{(m,n)}(x)$, $k = 0,1,2,\ldots$ are orthogonal with respect to $x^n(1-x)^m d$, i.e.

$$\int_0^1 H_k^{(m,n)}(x)\ x^j x^n(1-x)^m d\xi = 0,\ j = 0,1,\ldots,k-1.$$

Rearranging the terms in (2.2.3) and (2.2.4) we have

$$\frac{H_{k+1}^{(m,n)}(x)}{H_{k+1}(\Delta^m\mu_n)} = x\,\frac{H_k^{(m,n+1)}(x)}{H_k(\Delta^m\mu_{n+1})} - \frac{H_k(\Delta^m\mu_n)H_{k+1}(\Delta^m\mu_{n+1})}{H_k(\Delta^m\mu_{n+1})H_{k+1}(\Delta^m\mu_n)}\,\frac{H_k^{(m,n)}(x)}{H_k(\Delta^m\mu_n)}, \quad (2.4.1)$$

$$\frac{H_{k+1}^{(m,n+1)}(x)}{H_{k+1}(\Delta^m\mu_{n+1})} = \frac{H_{k+1}^{(m,n)}(x)}{H_{k+1}(\Delta^m\mu_n)} - \frac{H_{k+2}(\Delta^m\mu_n)H_k(\Delta^m\mu_{n+1})}{H_{k+1}(\Delta^m\mu_{n+1})H_{k+1}(\Delta^m\mu_n)}\,\frac{H_{k+1}^{(m,n+1)}(x)}{H_{k+1}(\Delta^m\mu_{n+1})}$$

(2.4.2)

Let $q_{k+1}^{(m,n)} = \dfrac{H_k(\Delta^m\mu_n)H_{k+1}(\Delta^m\mu_{n+1})}{H_k(\Delta^m\mu_{n+1})H_{k+1}(\Delta^m\mu_n)}$, $\quad k \geq 0$

$$e_{k+1}^{(m,n)} = \frac{H_{k+2}(\Delta^m\mu_n)H_k(\Delta^m\mu_{n+1})}{H_{k+1}(\Delta^m\mu_{n+1})H_{k+1}(\Delta^m\mu_n)}, \qquad k \geq 0, \quad e_0^{(m,n)} = 0$$

$$p_k^{(m,n)}(x) = \frac{H_k^{(m,n)}(x)}{H_k(\Delta^m\mu_n)}, \; k \geq 1 \text{ and } p_0^{(m,n)}(x) = 1, \; p_{-1}^{(m,n)}(x) = 0.$$

We can easily obtain

Theorem 2.4.1: For $k \geq 0$,

$$p_{k+1}^{(m,n)}(x) = (x-e_k^{(m,n)}-q_{k+1}^{(m,n)})p_k^{(m,n)}(x) - q_k^{(m,n)}e_k^{(m,n)}p_{k-1}^{(m,n)}(x) \qquad (2.4.3)$$

$$p_{k+1}^{(m,n)}(x) = (x-q_{k+1}^{(m,n-1)}-e_{k+1}^{(m,n-1)})p_k^{(m,n)}(x)-e_k^{(m,n-1)} q_{k+1}^{(m,n-1)}p_{k-1}^{(m,n)}(x) \qquad (2.4.4)$$

Proof: Brezinski (1980) has proved for the case $m = 0$. The proof for $m \neq 0$ is similar and so is omitted.

By comparing the coefficients in (2.4.3) and (2.4.4), we have

Corollary 2.4.1:
$$e_k^{(m,n)} + q_{k+1}^{(m,n)} = q_{k+1}^{(m,n-1)} + e_{k+1}^{(m,n-1)} \qquad (2.4.5)$$

$$q_k^{(m,n)}e_k^{(m,n)} = e_k^{(m,n-1)}q_{k+1}^{(m,n-1)} \qquad (2.4.6)$$

where $k \geq 1$.

Remark 2.4.1: (2.4.5) and (2.4.6) give the q-d algorithm in numerical analysis if $m = 0$.

Let

$$P_k(x) = p_k^{(0,0)}(x),$$

$$Q_k(x) = p_k^{(1,1)}(x),$$

$$R_k(x) = p_k^{(0,1)}(x),$$

$$S_k(x) = p_k^{(0,0)}(x),$$

i.e. $P_k(x)$, $Q_k(x)$, $R_k(x)$ and $S_k(x)$ are orthogonal with respect to $d\xi$, $x(1-x)d\xi$, $xd\xi$, and $(1-x)d\xi$ respectively. We have

Theorem 2.4.2: Let $P_{-1}(x) = Q_{-1}(x) = R_{-1}(x) = S_{-1}(x) = 0$, $P_0(x) = Q_0(x) = R_0(x) = S_0(x) = 1$. Then

(i) $P_{k+1}(x) = (x-\zeta_{2k}-\zeta_{2k+1})P_k(x)-\zeta_{2k-1}\zeta_{2k}P_{k-1}(x)$, $k \geq 1$, $P_1(x) = x-\zeta_1$,

(ii) $Q_{k+1}(x) = (x-\gamma_{2k+2}-\gamma_{2k+3})Q_k(x)-\gamma_{2k+1}\gamma_{2k+2}Q_{k-1}(x)$, $k \geq 0$,

(iii) $R_{k+1}(x) = (x-\zeta_{2k+1}-\zeta_{2k+2})R_k(x)-\zeta_{2k}\zeta_{2k+1}R_{k-1}(x)$, $k \geq 0$.

(iv) $S_{k+1}(x) = (x-\gamma_{2k+1}-\gamma_{2k+2})S_k(x)-\gamma_{2k}\gamma_{2k+1} S_{k-1}(x)$, $k \geq 1$,

$S_1(x) = x-\gamma_2$.

Proof: By Lemma 2.2.5, we see, for $k \geq 0$,

$$q_{k+1}^{(0,0)} = \zeta_{2k+1}, \quad q_{k+1}^{(1,0)} = \gamma_{2k+2},$$

$$e_{k+1}^{(0,0)} = \zeta_{2k+2}, \quad e_{k+1}^{(1,0)} = \gamma_{2k+3}.$$

So (i) and (iv) follow by substitution in (2.4.3). (ii) and (iii) follow by using (2.4.3), (2.4.5), and (2.4.6). The expressions for $P_1(x)$ and $S_1(x)$ can be found by a direct expansion of the determinant.

The following theorem, which has been proved by Chihara (1978), gives the recurrence relationship for the monic orthogonal polynomial (i.e. the orthogonal polynomials with leading coefficient equal to one) from that of the general orthogonal polynomial.

Theorem 2.4.3: If the orthogonal polynomial system $\{Q_k(x)\}$ satisfies

$$A_k Q_{k+1}(x) = (B_k x - C_k)Q_k(x) - D_k Q_{k-1}(x), \quad k = 0,1,2,\ldots \tag{2.4.7}$$

where we define $Q_{-1}(x) \equiv 0$, $Q_0(x)$ = constant $\neq 0$, then the monic polynomials

$$\bar{Q}_k(x) = \frac{A_0\, A_1\, \ldots\, A_{k-1}}{B_0\, B_1\, \ldots\, B_{k-1}}\, Q_k(x)$$

satisfy

$$\bar{Q}_{k+1}(x) = (x - \frac{C_k}{B_k})\bar{Q}_k(x) - \frac{A_{k-1}D_k}{B_{k-1}B_k}\,\bar{Q}_{k-1}(x).$$

Proof: See Chihara (1978).

Suppose we transform the interval of orthogonality by $y = ax+b$, then the "new" monic orthogonal polynomial $\hat{p}_k(ax+b)$ is related to the "old" orthogonal polynomial $Q_k(x)$ by the following formula:

$$Q_k(x) = \frac{a^{-n}B_0\, \ldots\, B_{k-1}}{A_0\, A_1\, \ldots\, A_{k-1}}\, \hat{p}_k(ax+b).$$

We have the following theorem

Theorem 2.4.4: If $Q_k(x)$ satisfies the recursion formula (2.4.7) then $\hat{P}_k(ax+b)$ satisfies

$$\hat{P}_{k+1}(y) = (y-a\frac{C_k}{B_k}-b)\hat{P}_k(y) - a^2\frac{A_{k-1}D_k}{B_{k-1}B_k}\hat{P}_{k-1}(y)$$

Proof: By induction.

Corollary 2.4.4: If we transform its interval of orthogonality [0,1] to the interval [-1,1], then the monic orthogonal polynomial in Theorem 2.4.1 becomes

$$\hat{P}_{k+1}(x) = (x+1-2\zeta_{2k}-2\zeta_{2k+1})\hat{P}_k(x)-4\zeta_{2k-1}\zeta_{2k}\hat{P}_{k-1}(x).$$

Now with $\hat{P}_1(x) = x+1-2\zeta_1$, $\hat{P}_{k+1}(x)$ is the orthogonal w.r.t. $d\xi'$ where ξ' is the derived measure on [-1,1] induced by the linear transformation on taking [0,1] to [-1,1].

Proof: Put a = 2 and b = -1 in Theorem 2.4.4 and the result follows immediately.

Remark 2.4.4: It can be proved similarly that the following results hold

$$\hat{Q}_{k+1}(x) = (x+1-2\gamma_{2k+2}-2\gamma_{2k+3})\hat{Q}_k(x)-4\gamma_{2k+1}\gamma_{2k+2}\hat{Q}_{k-1}(x), \quad k \geq 0.$$

$$\hat{R}_{k+1}(x) = (x+1-2\zeta_{2k+1}-2\zeta_{2k+2})\hat{R}_k(x)-4\zeta_{2k}\zeta_{2k+1}\hat{R}_{k-1}(x), \quad k \geq 0$$

$$\hat{S}_{k+1}(x) = (x+1-2\gamma_{2k+1}-2\gamma_{2k+2})\hat{S}_k(x)-4\gamma_{2k}\gamma_{2k+1}\hat{S}_{k-1}(x), \quad k \geq 1$$

$$\hat{S}_1(x) = x+1-2\gamma_2$$

where $\hat{Q}_k(x)$, $\hat{R}_k(x)$ and $\hat{S}_k(x)$ is orthogonal with respect to $(1-x^2)d\xi'$ $(1+x)d\xi'$ and $(1-x)d\xi'$ respectively.

Remark 2.4.5: Freud (1971) used the following form of recursion formula

$$\frac{\gamma_k}{\gamma_{k+1}} \hat{P}_{k+1}(x) = (x-\alpha_k)\hat{P}_k(x) - \frac{\gamma_{k-1}}{\gamma_k} \hat{P}_{k-1}(x), \quad x \in [-1,1],$$

where $k = 0, 1, 2, \ldots$, $\hat{P}_{-1}(x) \equiv 0$, $\gamma_{-1} = 0$.

By Theorem 2.4.3, the monic orthogonal polynomial

$$\hat{P}_{k+1}(x) = (x-\alpha_k)\hat{P}_k(x) - \left(\frac{\gamma_{k-1}}{\gamma_k}\right)^2 P_{k-1}(x)$$

Comparing with what we get from Corollary 2.4.4, we see that

$$\alpha_k = -1+2\zeta_{2k}+2\zeta_{2k+1}, \quad k \geq 1, \quad \alpha_0 = -1+2\zeta_1$$

$$\gamma_k = \frac{1}{2^k\sqrt{\zeta_1\zeta_2\cdots\zeta_{2k}}}, \quad k \geq 1.$$

Remark 2.4.6: Karlin and McGregor (1959) investigated the properties of the random walk X_n, $n = 0,1,2,\ldots$, by studying its corresponding orthogonal polynomial system which satisfies

$$Q_0(x) = 1$$

$$xQ_0(x) = \bar{r}_0 Q_0(x) + \bar{p}_0 Q_1(x)$$

$$xQ_k(x) = \bar{q}_k Q_{k-1}(x) + \bar{r}_k Q_k(x) + \bar{p}_k Q_{k+1}(x), \; k \geq 1, \quad x \in [-1,1]$$

where

$\bar{q}_k = Pr\{X_{n+1} = k-1 | X_n = k\}$,

$\bar{r}_k = Pr\{X_{n+1} = k | X_n = k\}$,

$\bar{p}_k = Pr\{X_{n+1} = k+1 | X_n = k\}$,

$Q_k(x)$ is orthogonal with respect to some measure $\psi(x)$.

It is equivalent to

$$Q_0(x) = 1,$$

$$\bar{p}_0 Q_1(x) = (x-\bar{r}_0)Q_0(x),$$

$$\bar{p}_n Q_{n+1}(x) = (x-\bar{r}_n)Q_n(x) - \bar{q}_n Q_{n-1}(x).$$

By Theorem 2.4.3, the monic orthogonal polynomials $\hat{Q}_n(x)$ satisfy the recurrence formula,

$$\hat{Q}_0(x) = 1$$

$$\hat{Q}_1(x) = (x-\bar{r}_0)\hat{Q}_0(x).$$

$$\hat{Q}_{n+1}(x) = (x-\bar{r}_n)\hat{Q}_n(x)-\bar{p}_{n-1}\bar{q}_n\hat{Q}_{n-1}(x).$$

By Corollary 2.4.4, that will imply

$$\bar{r}_n = -1+2\zeta_{2n} + 2\zeta_{2n+1},$$

$$\bar{p}_{n-1}\bar{q}_n = \zeta_{2n-1}\zeta_{2n}.$$

In particular, if we choose $\bar{r}_n = 0$ and $\bar{p}_0 = 1$, then we see $p_{2n-1} = \frac{1}{2}$ and $p_{2n} = \bar{q}_n$, where p_i is the canonical moment of the measure $\psi(x)$.

It can be checked directly that the denominators of the n-th convergent of the associated continued fraction are exactly the orthogonal polynomials $P_k(x)$. So given the recursion formula we can write down the continued fraction expansion of the corresponding measure. We have immediately:

<u>Theorem 2.4.5</u>. If $\int_0^1 \frac{d\xi(x)}{z-x} = \frac{1}{x-\zeta_1} - \frac{\zeta_1\zeta_2}{x-\zeta_2-\zeta_3} - \frac{\zeta_3\zeta_4}{x-\zeta_4-\zeta_5} - \ldots,$ then

(i) $\int_0^1 \frac{x(1-x)}{z-x}\, d\xi = \frac{c_1-c_2}{z-\gamma_2-\gamma_3} - \frac{\gamma_3\gamma_4}{z-\gamma_4-\gamma_5} - \ldots$

(ii) $\int_0^1 \frac{x}{z-x}\, d\xi = \frac{\zeta_1}{z-\zeta_1-\zeta_2} - \frac{\zeta_2\zeta_3}{z-\zeta_3-\zeta_4} - \ldots$

(iii) $\int_0^1 \frac{(1-x)}{z-x}\, d\xi = \frac{\gamma_1}{z-\gamma_2} - \frac{\gamma_2\gamma_3}{z-\gamma_3-\gamma_4} - \ldots$

If we transform the interval from [0,1] to [a,b], by Theorem 2.4.4, we see

<u>Theorem 2.4.6</u>. $\int_a^b \frac{d\xi'(x)}{z-x} = \frac{1}{z-a-\zeta_1(b-a)} - \frac{\zeta_1\zeta_2(b-a)^2}{z-a-(\zeta_2+\zeta_3)(b-a) -} \ldots$

where ξ' is the probability measure on [a,b] induced by ξ and

$$\int_0^1 \frac{d\xi(x)}{z-x} = \frac{1}{z-\zeta_1} - \frac{\zeta_1\zeta_2}{z-\zeta_2-\zeta_3-} \ldots$$

Let ξ be a probability measure on [0,1] with canonical moments $p_1, p_2, \ldots$ and $\tilde{\xi}$ be the probability measure on [-1,1] with canonical moments $\frac{1}{2}, p_1, \frac{1}{2}, p_2, \ldots$ The orthogonal polynomial $\tilde{p}_n(x)$ corresponding to $\tilde{\xi}$ is easily seen to be

$$K\begin{pmatrix} & -\zeta_1 & \cdots & \zeta_{n-1} & \\ x & & x \ldots x & & x \end{pmatrix}.$$

The following will show that $\tilde{p}_n(x)$ is closely related to the orthogonal polynomial related to ξ on [0,1].

<u>Lemma 2.4.7</u>: $K\begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{2n-1} \\ x & 1 & x & 1 \cdots x & 1 \end{pmatrix} = K\begin{pmatrix} a_1 & a_2 & \cdots & a_{2n-1} \\ 1 & x & 1 \quad 1 & x \end{pmatrix}.$

<u>Proof</u>: By induction.

<u>Theorem 2.4.7</u>: $K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2n-1} \\ x \; x & \cdots & x \quad x \end{pmatrix} = K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2n-1} \\ x^2 \; 1 & \cdots & x^2 \quad 1 \end{pmatrix} = P_n(x^2),$

$$K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2n} \\ x \; x & \cdots & x \quad x \end{pmatrix} = x\, K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2n} \\ 1 \; x^2 & \cdots & x^2 \quad 1 \end{pmatrix} = x\, R_n(x^2),$$

where $P_n(x)$ and $R_n(x)$ are defined in Theorem 2.4.2.

<u>Proof</u>: It is clear that the above statements hold for the case $n = 1$. Suppose it is true for $n \le k$. For $n = k+1$, we have

$$K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k+1} \\ x \; x & \cdots & x \quad x \end{pmatrix} = x\, K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k} \\ x \; x & \cdots & x \quad x \end{pmatrix} - \zeta_{2k+1} K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k+1} \\ x \; x & \cdots & x \quad x \end{pmatrix}$$

$$= x^2 K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k} \\ 1 \; x^2 & \cdots & x^2 \quad 1 \end{pmatrix} - \zeta_{2k+1} K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k-1} \\ x^2 \; 1 & \cdots & x^2 \quad 1 \end{pmatrix}$$

$$= K\begin{pmatrix} -\zeta_1 & & \cdots -\zeta_{2k-1} & -\zeta_{2k} & -\zeta_{2k+1} & \\ 1 & x^2 & 1 & x^2 & 1 & x^2 \end{pmatrix}$$

$$= K\begin{pmatrix} -\zeta_1 & -\zeta_2 & \cdots & -\zeta_{2k+1} \\ x^2 \; 1 & x^2 & \cdots & x^2 \quad 1 \end{pmatrix}$$

Based on the above results, we see

$$K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k+2} \\ x \; x & \cdots & x \quad x \end{pmatrix} = xK\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k+1} \\ x \; x & \cdots & x \quad x \end{pmatrix} - \zeta_{2k+2} K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k} \\ x \; x & \cdots & x \quad x \end{pmatrix}$$

$$= xK\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k+1} \\ x^2 \; 1 & \cdots & x^2 \quad 1 \end{pmatrix} - \zeta_{2k+2}\, xK\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k} \\ 1 \; x^2 & \cdots & x^2 \quad 1 \end{pmatrix}$$

$$= x\, K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k+1} \\ 1 \quad x^2 & \cdots \quad 1 & x^2 \end{pmatrix} - \zeta_{2k+2}\, x\, K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k} \\ 1 \quad x^2 & \cdots & x^2 \quad 1 \end{pmatrix}$$

$$= x\, K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k+1} & -\zeta_{2k+2} \\ 1 \quad x^2 & \cdots & 1 \quad x^2 & 1 \end{pmatrix}.$$

The proof is completed.

The following examples show how to find the canonical moments from the recursion formula of the orthogonal polynomials.

Example 2.4.1: Let $d\xi(x) = (1-x^2)^\alpha |x|^\gamma$ on $[-1,1]$. The corresponding orthogonal polynomials satisfy (see Chihara (1978))

$$\hat{p}_n(x) = x\, \hat{p}_{n-1}(x) - \lambda_n \hat{p}_{n-2}(x)$$

where

$$\lambda_{2m} = \frac{(2m+\gamma-1)(2m+2\alpha+\gamma+1)}{(4m+2\alpha+\gamma-3)(4m+2\alpha+\gamma-1)}$$

and

$$\lambda_{2m+1} = \frac{4m(m+\alpha)}{(4m+2\alpha+\gamma-1)(4m+2\alpha+\gamma+1)}.$$

From Corollary 2.4.4, we see $\lambda_2 = p_2 = \frac{\gamma+1}{2\alpha+\gamma+3}$, and

$$\lambda_3 = \frac{4(1+\alpha)}{(2\alpha+\gamma+3)(2\alpha+\gamma+5)} = \frac{2\alpha+2}{2\alpha+\gamma+3}\,\frac{2}{2\alpha+\gamma+5}$$

$$= q_2 p_4$$

which implies $p_4 = \frac{2}{2\alpha+\gamma+5}$. By induction, we can show that

$$p_{2i} = \frac{\gamma+i}{2\alpha+\gamma+2i+1}, \quad i \text{ odd},$$

$$p_{2i} = \frac{i}{2\alpha+\gamma+2i+1}, \quad i \text{ even}.$$

All the odd canonical moments are equal to $\frac{1}{2}$ since $d\xi(x)$ is symmetric.

Example 2.4.2: It has been proved by Skibinsky (1969) that the canonical moments of the Binomial distribution B(N,p) are given by

$$p_{2i-1} = p \qquad \text{and} \qquad p_{2i} = \frac{i}{N}, \qquad i = 1,2,\ldots,N.$$

The following is a new proof via the comparison of the coefficients of the orthogonal polynomials. It is known that the orthogonal polynomials corresponding to the binomial distribution are the so-called Krawtchouk polynomials (Szegö (1975)) which can be written as

$$p^n(n!)^{-1}M_n(x;\ -N,\ -1/q), \qquad n = 0,1,\ldots,N,$$

where

$$M_n(x;\ \beta,c) = (\beta)_n\ {}_2F_1(-n,\ -x;\ \beta;\ 1 - \tfrac{1}{c}),$$
$$(\beta)_n = \beta(\beta+1)\ldots(\beta+n-1).$$

So $P_n(x)$ is proportional to ${}_2F_1(-n,-x,-N,\frac{1}{p})$, which is equal to

$$1 - \frac{nx}{N}\frac{1}{p} + \frac{(-n)_2(-x)_2}{(-N)_2}\frac{1}{p^2}\frac{1}{2!} + \ldots + \frac{(-n)_{n-2}(-x)_{n-2}}{(-N)_{n-2}}\frac{1}{p^{n-2}}\frac{1}{(n-2)!}$$
$$+ \frac{(-n)_{n-1}(-x)_{n-1}}{(-N)_{n-1}}\frac{1}{p^{n-1}}\frac{1}{(n-1)!} + \frac{(-n)_n(-x)_n}{(-N)_n}\frac{1}{p^n}\frac{1}{n!}.$$

For example, $P_1(x) = x-Np$. So, by Theorem 2.4.4, we see $p_1 = p$. In general, suppose $P_{n+1}(x) = (x-\alpha_n)P_n(x)-\beta_n P_{n-1}(x)$. Then we see

$$\text{coef(coefficient) of } x^n \text{ in } P_{n+1}(x) = \text{coef of } x^{n-1} \text{ in } P_n(x)-\alpha_n,$$

$$\text{coef of } x^{n-1} \text{ in } P_{n+1}(x) = [\text{coef of } x^{n-2} \text{ in } P_n(x)]-\alpha_n[\text{coef of } x^{n-1} \text{ in } P_n(x)-\beta_n].$$

It can be shown that the coef of x^{n-1} in $P_n(x)$ is $-\frac{n(n-1)}{2} - p(N-n+1)n$, and the coef of x^{n-2} in $P_n(x)$ is

$$\frac{n(n-1)(n-2)(3n-1)}{24} + p(N-n+1)\ \frac{n(n-1)(n-2)}{2} + \frac{n}{2}\ (n-1)(N-n+2)(N-n+1)p^2.$$

So we see that

$$\alpha_n = nq + (N-n)p$$
$$= N[q\ \frac{n}{N} + \frac{N-n}{N}\ p],$$
$$\beta_n = (N-n+1)pqn$$
$$= N^2[\frac{N-n+1}{N}\ pq\ \frac{n}{N}].$$

Hence, we can conclude

$$p_{2i-1} = p \quad \text{and } p_{2i} = \frac{i}{N} \quad \text{for } i = 1,\dots,N.$$

We also get the simple recusion formula of the Krawtchouk polynomial

$$p_{n+1}(x) = (X-N[q\ \frac{n}{N} + \frac{N-n}{N}\ p])p_n(x)-N^2[\frac{N-n+1}{N}\ pq\ \frac{n}{N}]p_{n-1}(x)$$
$$= (X-N+np)p_n(x)-(N-n+1)npq\ p_{n-1}(x).$$

The form given here is much simpler than that of Lesky (1962) and Karlin and McGregor (1961).

The following theorem shows that the L_2 norm of an orthogonal polynomial can be expressed in terms of the canonical moments.

Theorem 2.4.8:

(i) $$\int_0^1 p_k^2(x)d\xi(x) = \zeta_1\zeta_2 \cdots \zeta_{2k}$$

(ii) $$\int_0^1 x(1-x)Q_k^2(x)d\xi(x) = \gamma_3\gamma_4 \cdots \gamma_{2k+2}$$

(iii) $$\int_0^1 x\ R_k^2(x)d\xi(x) = \zeta_2 \cdots \zeta_{2k+1}$$

(iv) $\int_0^1 (1-x)S_k^2(x)d\xi(x) = \gamma_2 \cdots \gamma_{2k+1}.$

Proof: Note that

$$\int_0^1 p_k^2(x)d\xi(x) = \int_0^1 x^k p_k(x)d\xi(x)$$

$$= \frac{H_{k+1}(\mu_0)}{H_k(\mu_0)} = \zeta_1 \cdots \zeta_{2k}.$$

So (i) is proved. The proof of the others are similar.

Remark 2.4.6: Let $\hat{P}_n((b-a)x+a)$, $x \in [0,1]$, be the orthogonal polynomial on [a,b] obtained from $P_n(x)$ which is the given orthogonal polynomial on [0,1]. By Remark 2.4.2, we see that

$$\hat{P}_n((b-a)x+a) = (b-a)^n P_n(x) \qquad x \in [0,1].$$

So

$$\int_0^1 (\hat{P}_n((b-a)x+a))^2 d\xi(x) = (b-a)^{2n} \zeta_1\zeta_2 \cdots \zeta_{2n}.$$

In particular, if $-a = b = 1$, then $\int_0^1 (\hat{P}_n(2x-1))^2 d\xi(x) = 2^{2n} \zeta_1 \cdots \zeta_{2n}.$

The following theorem is the confluent form of the Christoffel-Dorboux Identity.

Theorem 2.4.9:

$$1 + \sum_{k=1}^{n} \frac{p_k^2(x)}{\zeta_1 \cdots \zeta_{2k}} = \frac{p_{n+1}'(x)p_n(x) - p_n'(x)p_{n+1}(x)}{\zeta_1 \cdots \zeta_{2n}}.$$

Proof: See Szegö (1975) or Chihara (1978).

The value of the orthogonal polynomial at certain points can be expressed in terms of canonical moments as the following theorem shows.

Theorem 2.4.10:
$$K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_n \\ 1 \; 1 & \cdots & 1 \; 1 \end{pmatrix} = q_1 \cdots q_n$$

$$K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2n} \\ 0 \; 1 & \cdots & 1 \; 0 \end{pmatrix} = 0$$

$$K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2n+1} \\ 0 \; 1 & \cdots 0 & 1 \end{pmatrix} = (-1)^{n+1}\zeta_1\zeta_3\cdots\zeta_{2n+1}$$

$$K\begin{pmatrix} -\frac{1}{2} & -\frac{1}{2}p_2 & -\frac{1}{2}q_2, \ldots, & -\frac{1}{2}q_{2n} \\ \frac{1}{2} \quad 1 & \frac{1}{2} & 1 \ldots \quad \frac{1}{2} & 1 \end{pmatrix}$$

$$= (-1)^{n-1}\left(\frac{1}{4}\right)^{n-1} p_2 q_2 p_4 \cdots p_{2n}.$$

Proof: The first three can be obtained from direct calculation and the last one can be obtained by using the recursion relation

$$p_{n+1}(x) = (x - \tfrac{1}{2})p_n(x) - \zeta_{2n-1}\zeta_{2n}\, {}_{n-1}(x).$$

By Theorem 2.3.1 we can determine the canonical moments from the ratio of orthogonal polynomials.

Theorem 2.4.11: Given

$$\int_0^1 \frac{d\xi(x)}{z-x} = \frac{1}{z} - \frac{\zeta_1}{1} - \frac{\zeta_2}{z} - \frac{\zeta_3}{1} - \cdots$$

We have

(i) $$\frac{R_n(z)}{P_n(z)} = \frac{K\begin{pmatrix}-\zeta_1 & \cdots & -\zeta_{2n} \\ z\ 1 & \cdots\ 1 & z\end{pmatrix}}{K\begin{pmatrix}-\zeta_1 & \cdots & -\zeta_{2n-1} \\ z\ 1 & \cdots\ z & 1\end{pmatrix}} = z - \frac{\zeta_{2n}}{1} - \frac{\zeta_{2n-1}}{z} - \cdots - \frac{\zeta_1}{z},$$

(ii) $$\frac{P_n(z)}{R_{n-1}(z)} = \frac{K\begin{pmatrix}-\zeta_1 & \cdots & -\zeta_{2n-1} \\ z\ 1 & \cdots\ 1 & z\end{pmatrix}}{K\begin{pmatrix}-\zeta_1 & \cdots & -\zeta_{2n-2} \\ z\ 1 & \cdots\ 1 & z\end{pmatrix}} = 1 - \frac{\zeta_{2n-1}}{z} - \frac{\zeta_{2n-2}}{1} - \cdots$$

(iii) $$\frac{P_{n-1}(z)}{P_n(z)} = \frac{1}{z-\zeta_{2k-2}-\zeta_{2k-1}} - \frac{\zeta_{2k-3}\zeta_{2k-2}}{z-\zeta_{2k-4}-\zeta_{2k-3}} - \cdots$$

2.5. Canonical Moments and Ordinary Moments

In this section we show how to express the ordinary moments in terms of canonical moments. The results we obtained here were also found by Perron (1913) and Wall (1948). The proof given here is new and more elementary than that given by Perron (1913) and Wall (1948). The results we obtain here are more general than that of Skibinsky (1967).

It is known that any matrix can be diagonalized by two triangular matrices by using Gauss's elimination algorithm. We will apply such a diagonalization method to the Hankel matrix. The following theorem is taken from Gantmacher (1959).

Theorem 2.5.1: Every matrix $A = (a_{ij})_{i,j=1}^{n}$ of rank r in which $D_k = A\begin{pmatrix}1 & 2 & \cdots & k \\ 1 & 2 & \cdots & k\end{pmatrix} \neq 0$, for $k = 1,2,\ldots,r$ can be represented in the following form

$$A = FDL = \begin{bmatrix} 1 & 0 & \dots & 0 \\ f_{21} & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ f_{n1} & f_{n2} & \dots & 1 \end{bmatrix} \begin{bmatrix} D_1 & & & & & \\ & \frac{D_2}{D_1} & & & & \\ & & \ddots & & & \\ & & & \frac{D_r}{D_{r-1}} & & \\ & & & & 0 & \\ & & & & & \ddots \; 0 \end{bmatrix} \begin{bmatrix} 1 & \ell_{12} & \dots & \ell_{1n} \\ 0 & 1 & \dots & \ell_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

where

$$f_{gk} = \frac{A\binom{1\ 2\ \dots\ k-1\ g}{1\ 2\ \dots\ k-1\ k}}{A\binom{1\ 2\ \dots\ k}{1\ 2\ \dots\ k}}, \quad \ell_{kg} = \frac{A\binom{1\ 2\ \dots\ k-1\ k}{1\ 2\ \dots\ k-1\ g}}{A\binom{1\ 2\ \dots\ k}{1\ 2\ \dots\ k}}$$

$$(g = k+1 \dots .n, \quad k = 1,2,\dots,r)$$

and f_{gk} and ℓ_{kg} are arbitrary for $g = k+1,\dots,n$; $k = r+1,\dots,n$.

In our case, the Hankel matrix is symmetric and nonsingular, we see immediately that $F = L^T$. We have the following corollary.

Corollary 2.5.1: The Hankel matrix can be put into the following form

$$\underset{\sim}{M}_{2n} = M\binom{1\ 2\ \dots\ n+1}{1\ 2\ \dots\ n+1} = \begin{bmatrix} \mu_0 & \cdots & \mu_n \\ \vdots & & \vdots \\ \mu_n & \cdots & \mu_{2n} \end{bmatrix} = \begin{bmatrix} 1 & & & 0 \\ k_{01} & 1 & & \\ \vdots & \vdots & \ddots & \\ k_{0n} & k_{1n} & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & \zeta_1\zeta_2 & & \\ & & \ddots & \\ & & & \zeta_1\cdots\zeta_{2n} \end{bmatrix} \begin{bmatrix} 1 & k_{01} & \cdots & k_{0n} \\ & 1 & & k_{1,n} \\ & 0 & \ddots & \\ & & & 1 \end{bmatrix}$$

$$= K^T D\ K$$

where

$$k_{i-1,j-1} = \frac{M\binom{1\ 2\ \dots\ j-1\ i}{1\ 2\ \dots\ j-1\ j}}{M\binom{1\ 2\ \dots\ j}{1\ 2\ \dots\ j}}, \quad j \geq i \geq 1.$$

Proof: Directly from Theorem 2.5.1 and Theorem 2.2.2.

Let $x^T = (x_0, x_1, \dots, x_n)$ and $y^T = (y_0, y_1, \dots, y_n)$, we see $x^T \underline{M}_{2n} y = x^T K^T D K y$. So we have

Theorem 2.5.2: $$\sum_{p,q=0}^{n} \mu_{p+q} x_p y_q = (k_{00}x_0 + \dots + k_{0n}x_n)(k_{00}y_0 + \dots + k_{0n}y_n)$$
$$+ \zeta_1\zeta_2(k_{11}x_1 + \dots + k_{1n}x_n)(k_{11}y_1 + \dots + k_{1n}y_n)$$
$$+ \zeta_1\zeta_2\zeta_3\zeta_4(k_{22}x_2 + \dots + k_{2n}x_n)(k_{22}y_2 + \dots + k_{2n}y_n)$$
$$+ \dots + \zeta_1 \cdots \zeta_{2n} x_n y_n.$$

Particularly, we have

Corollary 2.5.2: Without loss of generality, we assume $p \leq q$. Then $\mu_{p+q} = k_{0p}k_{0q} + \zeta_1\zeta_2 k_{1p}k_{1q} + \zeta_1\zeta_2\zeta_3\zeta_4 k_{2p}k_{2q} + \dots + \zeta_1 \cdots \zeta_{2p} k_{pp}k_{pq}$.

We will give a procedure to compute k_{ij} below.

Let $p_k(x) = p_0^{(k)} + p_1^{(k)}x + \dots + p_k^{(k)}x^k$, $p_k^{(k)} = 1$, be the kth orthogonal polynomial. Let

$$L = \begin{pmatrix} p_0(0) & & & 0 \\ p_0^{(1)} & p_1^{(1)} & & \\ \vdots & \vdots & & \\ p_0^{(n)} & p_1^{(n)} & \dots & p_n^{(n)} \end{pmatrix}$$

We have the following lemma

<u>Lemma 2.5.1</u>: $L^{-1} = K^T$.

<u>Proof</u>: It is obvious that

$$L\,\underline{M}_{2n}\,L^T = D.$$

By comparing with Corollary 2.5.1, we see $L^{-1} = K^T$.

As a consequence, we have $K^T L = I$ i.e.

$$1 = p_0(x),$$

$$x = k_{01}p_0(x) + k_{11}p_1(x),$$

$$x^2 = k_{02}p_0(x) + k_{12}p_1(x) + k_{22}p_2(x).$$

If an orthogonal polynomial system $\{p_n(x)\}$ is given by

$$p_n(x) = (x-\alpha_n)p_{n-1}(x) - \beta_n p_{n-2}(x), \quad n \geq 2$$

$$p_0(x) = 1,\ p_1(x) = x-\alpha_1$$

put it in a matrix form, (Brezinski 1980)

$$\begin{bmatrix} \alpha_1 & 1 & & 0 \\ \beta_2 & \alpha_2 & 1 & \\ & \ddots & \ddots & \ddots\ 1 \\ 0 & & \beta_{n+1} & \alpha_{n+1} \end{bmatrix} \begin{bmatrix} p_0^{(0)} & & & 0 \\ p_0^{(1)} & p_1^{(1)} & & \\ \vdots & & \ddots & \\ p_0^{(n)} & p_1^{(n)} & \cdots & p_n^{(0)} \end{bmatrix} = \begin{bmatrix} p_0^{(0)} & & & 0 \\ p_0^{(1)} & p_1^{(1)} & & \\ \vdots & \vdots & \ddots & \\ p_0^{(n)} & p_1^{(n)} & \cdots & p_n^{(n)} \end{bmatrix} \begin{bmatrix} 0 & 1 & & 0 \\ & & 1 & \\ 0 & & \ddots & 1 \\ & & & 0 \end{bmatrix}.$$

That is

$$\begin{bmatrix} 1 & & & & \\ k_{01} & 1 & & 0 & \\ k_{02} & k_{12} & 1 & & \\ \vdots & & & & \\ k_{0n} & k_{1n} & k_{2n} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 & 1 & & & 0 \\ \beta_2 & \alpha_2 & 1 & & \\ & & \ddots & & \\ & & & \alpha_n & 1 \\ 0 & & & \beta_{n+1} & \alpha_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & & 0 \\ & & 1 & \\ & & & \ddots \\ 0 & & & 1 \\ & & & 0 \end{bmatrix} \begin{bmatrix} 1 & & & 0 \\ k_{01} & 1 & & \\ k_{02} & k_{12} & 1 & \\ \vdots & & & \ddots \\ k_{0n} & k_{1n} & & 1 \end{bmatrix}$$

$$\begin{bmatrix} k_{01} & & & \\ k_{02} & k_{12} & & 0 \\ \vdots & & & \\ k_{0n} & k_{1n} & & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \tag{2.5.1}$$

By Theorem 2.4.2, we have $\alpha_1 = \zeta_1$, $\alpha_n = \zeta_{2n-2} + \zeta_{2n-1}$ and $\beta_n = \zeta_{2n-3}\zeta_{2n-2}$. Substitute these values in (2.5.1), we can express k_{ij} in terms of ζ_k, $k = 1,2,\ldots$ For example

$$k_{01} = \zeta_1$$

$$k_{02} = \zeta_1^2 + \zeta_1\zeta_2 = \zeta_1(\zeta_1 + \zeta_2)$$

$$k_{12} = \zeta_1 + \zeta_2 + \zeta_3$$

By Corollary 2.5.2, we can write down the expression of the ordinary moments in terms of the canonical moments, for example,

$\mu_1 = k_{01} = \zeta_1$

$\mu_2 = k_{02} = \zeta_1(\zeta_1 + \zeta_2)$.

Consider the matrix

$$\underset{\sim}{M}_{2n+1} = \begin{bmatrix} \mu_1 & \cdots & \mu_{n+1} \\ \vdots & & \vdots \\ \mu_{n+1} & \cdots & \mu_{2n+1} \end{bmatrix} \left(= M^1\begin{pmatrix} 1 & 2 \ldots n+1 \\ 1 & 2 \ldots n+1 \end{pmatrix}\right).$$

We can apply Theorem 2.5.1 and obtain

<u>Theorem 2.5.3</u>: $\sum_{p,q=0}^{n} \mu_{p+q+1} x_p y_q = \zeta_1 (k'_{00} x_0 + \ldots + k'_{0n} x_n)(k'_{00} y_0 + \ldots + k'_{0n} y_n)$

$$+ \zeta_1 \zeta_2 \zeta_3 (k'_{11} x_1 + \ldots + k'_{1n} x_n)(k'_{11} y_1 + \ldots + k'_{1n} y_n)$$

$$+ \zeta_1 \zeta_2 \zeta_3 \zeta_4 \zeta_5 (k'_{22} x_2 + \ldots + k'_{2n} x_n)(k'_{22} y_2 + \ldots + k'_{2n} y_n)$$

$$+ \ldots + \zeta_1 \ldots \zeta_{2n+1} x_n y_n.$$

<u>Corollary 2.5.3</u>: Without loss of generality, we assume $p \leq q$.

$$\mu_{p+q+1} = \zeta_1 k'_{0p} k'_{0q} + \zeta_1 \zeta_2 \zeta_3 k'_{1p} k'_{1q} + \zeta_1 \ldots \zeta_5 k'_{2p} k'_{2q} + \ldots + \zeta_1 \ldots \zeta_{2p+1} k'_{pp} k'_{pq}.$$

As in the previous case, k'_{ij} is given by

$$k'_{ij} = \frac{M'\binom{1\ 2\ \ldots\ i-1\ i}{1\ 2\ \ldots\ i-1\ j}}{M'\binom{1\ 2\ \ldots\ i}{1\ 2\ \ldots\ j}}.$$

In finding the expression of μ_n in terms of canonical moments, Skibinsky (1967) defined

$$S_{i,n-i} = \begin{cases} \dfrac{M\binom{1\ \ldots\ m-i\ m-i+1}{1\ \ldots\ m-i\ m+1}}{M\binom{1\ \ldots\ m-i+1}{1\ \ldots\ m-i+1}} & \text{for } n = 2m. \\[2ex] \dfrac{M'\binom{1\ \ldots\ m-i\ m-i+1}{1\ \ldots\ m-i\ m+1}}{M'\binom{1\ \ldots\ m-i+1}{1\ \ldots\ m-i+1}} & \text{for } n = 2m+1. \end{cases}$$

So we see immediately

$$S_{i,n-i} = k'_{m-i+1,m+1}$$

for $n = 2m+1$, or

$$k'_{ij} = S_{j-i,\ i+j+1}$$

If we let q = 0 in Corollary 2.5.3, we see

$$\mu_{p+1} = \zeta_1 k'_{0p} k'_{0q} = \zeta_1 S_{p,p+1}$$

which is the result obtained by Skibinsky (1967). By using the same method in deriving (2.5.1), we find:

$$\begin{bmatrix} 1 & & \\ k'_{01} & 1 & \\ k'_{02} & k'_{12} & 1 \end{bmatrix} \begin{bmatrix} \zeta_1+\zeta_2 & 1 & \\ \zeta_2\zeta_3 & \zeta_3+\zeta_4 & 1 \\ & \zeta_4\zeta_5 & \zeta_5+\zeta_6 \end{bmatrix}$$

$$= \begin{bmatrix} k'_{01} & 1 & & \\ k'_{12} & k'_{12} & & \\ k'_{13} & k'_{13} & k'_{23} & 1 \end{bmatrix}.$$

So we have

$$k'_{01} = S_{1,2} = \zeta_1+\zeta_2$$

$$k'_{02} = S_{2,3} = (\zeta_1+\zeta_2)^2 + \zeta_2\zeta_3$$

$$k'_{12} = S_{1,4} = \zeta_1+\zeta_2+\zeta_3+\zeta_4.$$

2.6. Measures on the Circle

In this section we will introduce some parameters of the measures on the circle and show how they relate to the canonical moments under some conditions. More details of the following can be found in Geronimus (1948).

Let σ be a measure with support on n+1 points on the circle. Define the (trigonometric) moments by

$$c_k = \frac{1}{2\pi}\int_0^{2\pi} e^{-ik\theta} d\sigma(\theta), \; k = 0, \pm1, \pm2, \ldots, \pm n. \tag{2.6.1}$$

The Toeplitz determinant is given by

$$\Delta_k = |c_{i-j}|_0^k = \begin{vmatrix} c_0 & c_1 & \cdots & c_k \\ c_{-1} & c_0 & \cdots & c_{k-1} \\ \vdots & \vdots & & \vdots \\ c_{-k} & c_{-k+1} & \cdots & c_0 \end{vmatrix}, \quad k = 0,1,\ldots,n. \tag{2.6.2}$$

The polynomials

$$\varphi_k(z) = \frac{1}{\Delta_{k-1}} \begin{vmatrix} c_0 & c_1 & \cdots & c_{k-1} & 1 \\ c_{-1} & c_0 & \cdots & c_{k-2} & z \\ \vdots & & & \vdots & \vdots \\ c_{-k+1} & c_{k+2} & \cdots & c_0 & z^{k-1} \\ c_{-k} & c_{k+1} & \cdots & c_{-1} & z^k \end{vmatrix} \quad k = 0,1,\ldots,n, \ \Delta_{-1} = 1, \tag{2.6.3}$$

are orthogonal with respect to σ. More precisely, we have

$$\frac{1}{2\pi}\int_0^{2\pi} \varphi_k(e^{i\theta})e^{-i\ell\theta}d\sigma(\theta) = \begin{cases} 0, & 0 \leq \ell \leq k-1 \\ h_k = \dfrac{\Delta_k}{\Delta_{k-1}}, & \ell = k \end{cases}. \tag{2.6.4}$$

Define the parameters $\{a_k\}$ by

$$(-1)^k a_k = \frac{|c_{i-j+1}|_0^k}{|c_{i-j}|_0^k} = \frac{1}{\Delta_k} \begin{vmatrix} c_1 & c_2 & \cdots & c_{k+1} \\ c_0 & c_1 & \cdots & c_k \\ c_{-1} & c_0 & \cdots & c_{k-1} \\ \vdots & \vdots & & \vdots \\ c_{-k+1} & c_{-k+2} & \cdots & c_1 \end{vmatrix}, \quad k = 0,1,\ldots,n. \tag{2.6.5}$$

Notice that $\Delta_k > 0$ since σ has its support resting on n+1 points.

<u>Theorem 2.6.1</u>: The orthogonal polynomials $\{\varphi_k(z)\}_0^n$ are connected by the relationship

$$\varphi_{k+1}(z) = z\,\varphi_k(z) - \bar{a}_k\varphi_k^*(z), \text{ where } \varphi_k^*(z) = z^k\bar{\varphi}_k(\tfrac{1}{z}) \tag{2.6.6}$$

and the three-term recurrence formula

$$\bar{a}_k\,\varphi_{k+2}(z) = (\bar{a}_k z + \bar{a}_{k+1})\varphi_{k+1}(z) - z\bar{a}_{k+1}(1-|a_k|^2)\varphi_k(z) \tag{2.6.7}$$

$$\varphi_0(z) = 1,\quad \varphi_1(z) = z-\bar{a}_0.$$

Proof: See Geronimus (1948).

Remark 2.6.1: $\varphi_k^*(z)$ satisfies the recurrence formula $\varphi_{k+1}^*(z) = \varphi_k^*(z) - a_k z\varphi_k(z)$. From the above theorem, we can deduce

$$h_{k+1} = \frac{\Delta_{k+1}}{\Delta_k} = c_0 \prod_{i=0}^{k} (1-|a_0|^2). \tag{2.6.8}$$

Using (2.6.4), we obtain

$$\frac{\Delta_{k+1}\Delta_{k-1}}{\Delta_k^2} = 1-|a_k|^2,\quad |a_k| = \frac{\sqrt{\Delta_k^2-\Delta_{k+1}\Delta_{k-1}}}{\Delta_k}. \tag{2.6.9}$$

We observe that $|a_k| \leq 1$. Indeed, we have

Theorem 2.6.2: The necessary and sufficient conditions for the complex numbers c_k, $k = 0,1,\ldots$ to have a representation $c_k = \frac{1}{2\pi}\int_0^{2\pi} e^{-ik\theta}\,d\sigma(\theta)$ where σ is a measure with its support rested on n+1 points in $[0,2\pi]$ is $|a_k| < 1$, $k = 0,1\ldots.n-1$ and $|a_n| = 1$.

Proof: See Geronimus (1948).

Example 2.6.1: For the probability measure concentrated at one point $\theta = \theta_0$, it is clear that $c_0 = 1$ and $c_1 = e^{-i\theta_0}$. By (2.6.5), we obtain $a_0 = e_1 = e^{-i\theta_0}$, i.e. $|a_0| = 1$.

Example 2.6.2: Consider the probability measure $\sigma(\theta = v + \frac{2\pi r}{m}) = \frac{1}{m}$, $r = 0,1,\ldots,m-1$. This is the probability measure putting equal mass on m points which are equidistant on the circle. c_k is given by

$$\frac{1}{m}\sum_{r=0}^{m-1} e^{-ik(v+\frac{2\pi r}{m})}$$

$$= \frac{1}{m} e^{-ikv}\sum_{r=0}^{m-1} (\cos k \frac{2\pi r}{m} - i\sin k \frac{2\pi r}{m})$$

$$= \begin{cases} 0 & \text{if } k \not\equiv 0 \pmod m \\ e & \text{if } k \equiv 0 \pmod m. \end{cases}$$

By (2.6.5), we see $a_0 = a_1 = \ldots = a_{m-2} = 0$ and $a_{m-1} = e^{-ikv}$. Hence we have $|a_{m-1}| = 1$. In particular, if $v = 0$, we have $a_{m-1} = 1$.

Example 2.6.3: Let $d\sigma(\theta) = \frac{1}{2\pi} d\theta$, it can be shown that $c_0 = 1$ and $c_k = 0$ for $k \geq 1$ (see Mardia (1972)). Correspondingly we have

$$a_0 = a_1 = \ldots = 0.$$

Example 2.6.4: Let $d\sigma(\theta) = \frac{1}{2\pi} \frac{1-\rho^2}{1+\rho^2-2\rho \cos \theta} d\theta$. This is also called the wrapped Cauchy distribution (see Mardia (1972)). c_k is given by $\rho^{|k|}$. The parameters are $a_0 = p$, $a_1 = a_2 = a_3 = \ldots = 0$.

Let ξ be a probability measure on the interval $[-1,1]$. Define

$$\sigma(\theta) = \begin{cases} \xi(1) - \xi(\cos \theta), & 0 \leq \theta \leq \pi, \\ \xi(\cos\theta) - \xi(0), & -\pi \leq \theta \leq 0. \end{cases}$$

So $\sigma(\theta)$ is a measure on the circle and $c_0 = \frac{1}{\pi}$. Geometrically, σ is obtained from ξ by splitting equally the mass at x_0 on the x-axis to the two points on the semi-circles that have same x-coordinate as x_0. Now all c_k are real and $c_k = c_{-k}$, so it makes sense to talk about maximizing (minimizing) c_{k+1} given c_0, $c_1,\ldots,c_k$. Thus we can define the canonical moments for the measures on the circle as

$$p_k = \frac{c_k - c_k^-}{c_k^+ - c_k^-},$$

where $c_k^+(c_k^-)$ is the maximum (minimum) of c_k given $c_0, c_1, \ldots, c_{k-1}$. It can be seen from (2.6.5) that $a_k = 1$ gives $c_{k+1} = c_{k+1}^+$ and $a_k = -1$ gives $c_{k+1} = c_{k+1}^-$. Corresponding to Lemma 2.2.1, we have, using the same trick,

$$c_{k+1} - c_{k+1}^- = (a_k+1)\,\frac{\Delta_k}{\Delta_{k-1}}.$$

Similarly, we can show $c_{k+1}^+ c_{k+1} = (1-a_k)\,\frac{\Delta_k}{\Delta_{k-1}}$. Consequently, we have $c_{k+1}^+ - c_{k+1}^- = \frac{\Delta_k}{\Delta_{k-1}}\,2$ and $p_{k+1} = \frac{1+a_k}{2}$.

The p_n defined above is equal to the one we defined in Section 1 for the measure ξ on [-1,1] since

$$\begin{aligned} c_k &= \frac{1}{2\pi}\int_0^{2\pi} e^{-ik\theta} d\sigma(\theta) \\ &= \frac{1}{\pi}\int_{-1}^{1} T_k(x)\, d\xi(x) \\ &= \frac{1}{\pi}\int_{-1}^{1} (2^{k-1}x^k + \ldots)\, d\xi(x) \end{aligned}$$

We can see that two p_n's are the same by using the same line of reasoning in the proof of Theorem 2.1.5.

As a final remark in this section, we notice that if we rotate the measure $d\sigma$ on the circle by an angel θ, $|a_k|$ will remain unchanged.

2.7 Canonical Moments and the Support of ξ

Given a terminated sequence of canonical moments $p_1,\ldots,p_n$, where $p_n = 1$ or 0, the corresponding ξ is uniquely determined. It is desirable to find the support and the weight attached to each point in the support. Since the weight can be found easily if the support is known, we will concentrate on the finding of the support of ξ in this section.

Recall that ξ has a continued fraction expansion

$$\int_0^1 \frac{d\xi(t)}{x-t} = \frac{1}{x-}\,\frac{\xi_1}{1-}\,\frac{\xi_2}{x-\ldots}$$

If $P_n = 1$, say, then the right hand side terminates and has the form

$$\int_0^1 \frac{d\xi(t)}{x-t} = \frac{1}{x-}\,\frac{\zeta_1}{1-}\,\frac{\zeta_2}{x-}\,\ldots-\frac{\zeta_n}{\tau_n}$$

where $\zeta_n = q_{n-1}P_n = q_{n-1}$ and $\tau_n = 1$ or x according to n is odd or even. So the zeros of the denominator of the whole expression, namely

$$K\begin{pmatrix} & -\zeta_1 & \cdots & -\zeta_n \\ x & 1 & & \tau_n \end{pmatrix}$$

are the support points of ξ. Similarly, if $p_n = 0$, the roots of

$K\begin{pmatrix} & -\zeta_1 & \cdots & -\zeta_{n-1} \\ x & 1 & \cdots & \tau_{n-1} \end{pmatrix} = 0$ are the support points of the measure ξ. So we have the following theorem:

Theorem 2.7.1: If $p_n = 1$, the support of ξ is the set of zeros of $K\left(\begin{matrix} -\zeta_1 & \cdots & -\zeta_n \\ x\ 1 & \cdots & \tau_n \end{matrix}\right)$. If $p_n = 0$, the support of ξ is the set of zeros of $K\left(\begin{matrix} -\zeta_1 & \cdots & -\zeta_{n-1} \\ x\ 1 & \cdots & \tau_{n-1} \end{matrix}\right)$.

Sometimes, it is helpful to change the form of the polynomial in such a way that the roots remain unchanged. We have the following lemma.

Lemma 2.7.1: $K\left(\begin{matrix} a_1\ a_2 & \cdots & a_m \\ x\ 1 & \cdots & \tau_m \end{matrix}\right) = \tau_m K\left(\begin{matrix} a_1 & \cdots & a_m \\ 1\ x & \cdots & \tau_{m+1} \end{matrix}\right)$.

Proof: When $m = 1$, the statement obviously holds. Suppose the statement holds for $m \leq k$. Then by expanding in the last row,

$$K\left(\begin{matrix} a_1\ a_2 & \cdots & a_k\ a_{k+1} \\ x\ 1\ x & \cdots & \tau_k\ \tau_{k+1} \end{matrix}\right) = \tau_{k+1} K\left(\begin{matrix} a_1 & \cdots & a_k \\ x\ 1 & \cdots & \tau_k \end{matrix}\right) + a_{k+1} K\left(\begin{matrix} a_1 & \cdots & a_{k-1} \\ x\ 1 & \cdots & \tau_{k-1} \end{matrix}\right).$$

By induction hypothesis, it becomes

$$\tau_{k+1}\tau_k K\left(\begin{matrix} a_1 & \cdots & a_k \\ 1\ x & \cdots & \tau_{k+1} \end{matrix}\right) + a_{k+1}\tau_{k-1} K\left(\begin{matrix} a_1 & \cdots & a_{k-1} \\ 1\ x & \cdots & \tau_k \end{matrix}\right)$$

$$= \tau_{k+1} K\left(\begin{matrix} a_1 & \cdots & a_{k+1} \\ 1\ x & \cdots & \tau_{k+2} \end{matrix}\right).$$

The last line is obtained by noticing that

$$\tau_k = \tau_j \quad \text{if} \quad k-j \equiv 0 \pmod 2.$$

Lemma 2.7.2: $K\left(\begin{matrix} a_1\ a_2 & \cdots & a_m \\ x\ 1 & \cdots & \tau_m \end{matrix}\right) = K\left(\begin{matrix} a_m & \cdots & a_1 \\ \tau_m\ \tau_{m-1} & \cdots & x \end{matrix}\right)$.

Proof: The result holds since we just transpose the corresponding matrix.

<u>Lemma 2.7.3</u>: $K\begin{pmatrix} & -\zeta_1 & -\zeta_2 & \cdots & -\zeta_m \\ x & 1 & x & \cdots & \tau_m \end{pmatrix} = K\begin{pmatrix} & -p_m & -\gamma_m & \cdots & -\gamma_2 \\ x & 1 & & \cdots & \tau_m \end{pmatrix}.$

<u>Proof</u>: It can be easily checked that the statement holds for $m = 1$ and 2. Suppose the statement holds for $m \leq k$, then by expanding in the last row

$$K\begin{pmatrix} & -\zeta_1 & \cdots & -\zeta_{k+1} \\ x & & \cdots & \tau_{k+1} \end{pmatrix} = \tau_{k+1} K\begin{pmatrix} & -\zeta_1 & \cdots & -\zeta_k \\ x & 1 & \cdots & \tau_k \end{pmatrix}$$

$$- \zeta_{k+1} K\begin{pmatrix} & -\zeta_1 & \cdots & -\zeta_{k-1} \\ x & 1 & \cdots & \zeta_{k-1} \end{pmatrix}.$$

By induction hypothesis, it becomes

$$\tau_{k+1} K\begin{pmatrix} & -p_k & -\gamma_k & \cdots & -\gamma_2 \\ x & 1 & & \cdots & \tau_k \end{pmatrix} - \zeta_{k+1} K\begin{pmatrix} & -p_{k-1} & -\gamma_{k-1} & \cdots & -\gamma_2 \\ x & 1 & & \cdots & \tau_{k-1} \end{pmatrix}.$$

By writing $p_k = \gamma_{k+1} + p_k p_{k+1}$ and $p_{k-1} = \gamma_k + p_{k-1} p_k$, we have

$$\tau_{k+1}\left[K\begin{pmatrix} & -\gamma_{k+1} & -\gamma_k & \cdots & -\gamma_2 \\ x & 1 & & \cdots & \tau_k \end{pmatrix} - p_k p_{k+1} K\begin{pmatrix} & -\gamma_{k-1} & \cdots & -\gamma_2 \\ x & 1 & \cdots & \tau_k \end{pmatrix}\right]$$

$$- \zeta_{k+1}\left[K\begin{pmatrix} & -\gamma_k & \cdots & -\gamma_2 \\ x & 1 & \cdots & \tau_{k-1} \end{pmatrix} - p_{k-1} p_k K\begin{pmatrix} & -\gamma_{k-2} & \cdots & -\gamma_2 \\ x & 1 & \cdots & \tau_{k-1} \end{pmatrix}\right]$$

$$= \tau_{k+1} K\begin{pmatrix} & -\gamma_{k+1} & -\gamma_k & \cdots & -\gamma_2 \\ x & 1 & & \cdots & \tau_k \end{pmatrix} - \zeta_{k+1} K\begin{pmatrix} & -\gamma_k & \cdots & -\gamma_2 \\ x & 1 & \cdots & \tau_{k-1} \end{pmatrix}$$

$$- p_k p_{k+1}\left[\tau_{k+1} K\begin{pmatrix} & -\gamma_{k-1} & \cdots & -\gamma_2 \\ x & 1 & \cdots & \tau_k \end{pmatrix} - \gamma_k K\begin{pmatrix} & -\gamma_{k-2} & \cdots & -\gamma_2 \\ x & 1 & \cdots & \tau_{k-1} \end{pmatrix}\right]$$

$$= \tau_{k+1} K\begin{pmatrix} & -\gamma_{k+1} & \cdots & -\gamma_2 \\ x & 1 & \cdots & \tau_k \end{pmatrix} - \zeta_{k+1} \tau_{k+1} K\begin{pmatrix} & -\gamma_k & \cdots & -\gamma_2 \\ 1 & x & \cdots & \tau_k \end{pmatrix}$$

$$- p_k p_{k+1}\left[\tau_{k+1} K\begin{pmatrix} & -\gamma_{k-1} & \cdots & -\gamma_2 \\ x & 1 & \cdots & \tau_k \end{pmatrix} - \tau_{k+1} \gamma_k K\begin{pmatrix} & -\gamma_{k-2} & \cdots & -\gamma_2 \\ 1 & & \cdots & \tau_k \end{pmatrix}\right]$$

$$= \tau_{k+1}\left[K\begin{pmatrix} -\zeta_{k+1} & -\gamma_{k+1} & \cdots & -\gamma_2 \\ 1 & x & \cdots & \tau_k \end{pmatrix} - p_k p_{k+1} K\begin{pmatrix} -\gamma_k & \cdots & -\gamma_2 \\ 1\ x & \cdots & \tau_k \end{pmatrix}\right]$$

$$= K\begin{pmatrix} & -p_{k+1} & \cdots & -\gamma_2 \\ x & 1 & \cdots & \tau_{k+1} \end{pmatrix}.$$

<u>Theorem 2.7.1</u>: (Studden) The support of measures corresponding to $(p_1,\ldots,p_m,0)$ and $(p_m,\ldots,p_1,0)$ are the same.

<u>Proof</u>: By Lemma 2.7.3, we see that

$$K\begin{pmatrix} & -\zeta_1 & \cdots & -\zeta_m \\ x & 1 & \cdots & \tau_m \end{pmatrix} = K\begin{pmatrix} & -p_m - \gamma_m & \cdots & -\gamma_2 \\ x & 1 & \cdots & \tau_m \end{pmatrix}.$$

The result follows immediately.

<u>Theorem 2.7.2</u>: (Studden) The support of measures corresponding to $(p_1,\ldots,p_m,1)$ and $(q_m,\ldots,q_1,1)$ are the same.

<u>Proof</u>: By Lemma 2.7.3, we see that

$$K\begin{pmatrix} & -\zeta_1 & \cdots & -\zeta_m & -q_m \\ x & 1 & \cdots & \tau_m & \tau_{m+1} \end{pmatrix} = K\begin{pmatrix} & -1 & 0 & -\gamma_m & \cdots & -\gamma_2 \\ x & 1 & x & 1 & \cdots & \tau_{m+1} \end{pmatrix}.$$

By simple row operation, we can write the right hand side as

$$(x-1)K\begin{pmatrix} -\gamma_m & \cdots & -\gamma_2 \\ x\ 1 & \cdots & \tau_{m+1} \end{pmatrix}.$$

On the other hand,

$$K\begin{pmatrix} & -q_m & -\gamma_{m-1} & \cdots & -\gamma_2 & -p_1 \\ x & 1 & & \cdots & \tau_m & \tau_{m+1} \end{pmatrix} = K\begin{pmatrix} & -1 & 0 & -\gamma_2 & \cdots & -\gamma_m \\ x & 1 & x & & \cdots & \tau_{m+1} \end{pmatrix}$$

$$= (x-1)K\begin{pmatrix} & -\gamma_2 & \cdots & -\gamma_m \\ x & 1 & \cdots & \tau_{m+1} \end{pmatrix}.$$

Since $K\left(\begin{matrix} & -\gamma_m & & \cdots & -\gamma_2 \\ x & & 1 & \cdots & \tau_{m+1} \end{matrix}\right) = K\left(\begin{matrix} & -\gamma_2 & \cdots & -\gamma_m \\ x & 1 & & \tau_{m+1} \end{matrix}\right)$,

the result follows.

The following theorem gives the support of measures with terminated canonical moments.

Theorem 2.7.3: (i) The measure corresponding to $(P_1,\ldots,P_{2k-1},0)$ is supported on the zeros of $P_k(x) = 0$

(ii) $(P_1,\ldots,P_{2k-1}, 1)$ is supported by the zeros of $x(x-1)Q_{k-1}(x) = 0$

(iii) $(P_1,\ldots,P_{2k}, 0)$ is supported by the zeros of $x\, R_k(x) = 0$

(iv) $(P_1,\ldots,P_{2k}, 1)$ is supported by the zeros of $(x-1)S_k(x) = 0$.

Proof: (i) The support of the measure corresponding to $(P_1,\ldots,P_{2k-1}, 0)$ is given by the zeros of $K\left(\begin{matrix} & -\zeta_1 & \cdots & & -\zeta_{2k-1} \\ x & 1 & \cdots & x & 1 \end{matrix}\right) = 0$, i.e. $P_k(x) = 0$.

(ii) The support of the measure corresponding to $(P_1,\ldots,P_{2k-1}, 1)$ is given by the zeros of $K\left(\begin{matrix} & -\zeta_1 & \cdots & & -\zeta_{2k} \\ x & 1 & \cdots & x & x \end{matrix}\right) = 0$. By using Lemma 2.7.1 and Lemma 2.7.3, we see that

$$K\left(\begin{matrix} & -\zeta_1 & \cdots & -\zeta_{2k} \\ x & 1 & \cdots & x \end{matrix}\right) = K\left(\begin{matrix} & -1 & 0 & \gamma_{2k-1} & \cdots & -\gamma_2 \\ x & 1 & x & 1 & \cdots & x \end{matrix}\right)$$

$$= (x-1)K\left(\begin{matrix} & -\gamma_{2k-1} & \cdots & -\gamma_2 \\ x & 1 & \cdots & x \end{matrix}\right)$$

$$= (x-1)x\, K\left(\begin{matrix} & -\gamma_2 & \cdots & -\gamma_{2k-1} \\ 1 & x & \cdots & 1 \end{matrix}\right)$$

$$= (x-1)x\, Q_{k-1}(x).$$

(iii) The support of the measure corresponding to $(P_1,\ldots,P_{2k}, 0)$ is given by the zeros of $K\left(\begin{matrix} & -\zeta_1 & \cdots & -\zeta_{2k} \\ x & 1 & \cdots & x \end{matrix}\right)$. By using Lemma 2.7.1,

$$K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k} \\ x & 1 \; \cdots & x \end{pmatrix} = x\, K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k} \\ 1 & x \; \cdots & x \;\; 1 \end{pmatrix}$$

$$= x\, R_k(x).$$

(iv) The support of the measure corresponding to $(P_1,\ldots,P_{2k}, 1)$ is given by the zeros of $K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k+1} \\ x & 1 \; \cdots & x \;\; 1 \end{pmatrix} = 0$. By Lemma 2.7.3

$$K\begin{pmatrix} -\zeta_1 & \cdots & -\zeta_{2k+1} \\ x & 1 \; \cdots & x \;\; 1 \end{pmatrix} = (x-1)K\begin{pmatrix} -1 & 0 & -\gamma_{2k} & \cdots & \gamma_2 \\ x & 1 & x \;\; 1 & \cdots & 1 \end{pmatrix}$$

$$= (x-1)K\begin{pmatrix} -\gamma_2 & -\gamma_3 & \cdots & -\gamma_{2k} \\ x \;\; 1 & & \cdots & 1 \end{pmatrix}$$

$$= (x-1)S_k(x).$$

Definition 2.7.1: The index $I(\xi)$ of a measure is the number of support points of ξ with the convention that the two endpoints counted one half and the interior counted one.

Theorem 2.7.4: Suppose $0 < p_i < 1$, $i = 1,\ldots,n-1$, $p_n = 1$ or 0. The index of ξ is given by $\frac{n}{2}$.

Proof: Direct verification.

Definition 2.7.2: Let μ^0 be an interior point of $\mathcal{M}_{n+1}$. A representation ξ for μ^0 of index $I(\xi) = \frac{n+1}{2}$ is called principal and any representation of index $I(\xi) \leq \frac{n+2}{2}$ is called canonical. A canonical or principal representation is further designated by the term upper if it involves the endpoint 1 and the term lower if it does not involve endpoint 1.

Remark 2.7.1: For any $\mu^0 \in \operatorname{Int} \mathcal{M}_{n+1}$, we can define $(p_1, p_2,\ldots,p_n)$ such that $0 < p_i < 1$ for $1 \leq i \leq n$. So upper principal representation is the one that corresponds to $(\mu_0, p_1, p_2,\ldots,p_n, 1)$ and the lower principal representation is the one that corresponds to $(\mu_0, p_1, p_2,\ldots,p_n, 0)$. Similarly, the upper canonical representation are those correspond to $(\mu_0, p_1, p_2,\ldots,p_n, p_{n+1}, 1)$, here $0 < p_{n+1} \leq 1$. The lower canonical representation are those correpond to $(_0, p_1,\ldots,p_n,p_{n+1}, 0)$, here $0 \leq p_{n+1} < 1$.

2.8 Canonical Moments and Measures

Recall that the Stieltjes transform of a measure on [-1,1] has a continued fraction expansion in terms of the canonical moments. It is desirable to find the explicit form of ξ for a given infinite sequence of canonical moments. It is found that the task is difficult if not impossible except in some special cases. Our next aim is to determine the behavior of ξ given certain limiting properties of the sequence of canonical moments. It is known that the canonical moments are closely related to the coefficients of the three-term recursion formula that defines the orthogonal polynomial system. So we use the results from the theory of orthogonal polynomial extensively in this section. Usually the support of the optimal design can be given in terms of the roots of some orthogonal polynomial. So it is interesting to determine the limiting distribution of those roots. We will give a necessary and sufficient condition for the limiting distribution to be arc-sine.

Given the Stieltjes transform G(z), we can 'recover' the measure by the following

Theorem 2.8.1: (Stieltjes - Perron inversion formula)

$$\frac{1}{2}[\xi(x^+) + \xi(x^-)] - \frac{1}{2}[\xi(y^+) + \xi(y^-)] = -\frac{1}{\pi} \lim_{\eta \to 0^+} \mathrm{Im} \int_{y+i\eta}^{x+i\eta} G(z)dz.$$

Proof: See Perron (1913) or Wall (1948).

Corollary 2.8.1: If $d\xi(x) = f(x)dx$ and $G(z)$ is analytic, in some segment $[x-\varepsilon, x+\varepsilon] \subset [-1,1]$, then $f(x) = -\frac{1}{\pi} \mathrm{Im}\, G(x)$.

Proof: Let $x_0 \in (x-\varepsilon, x]$. By Cauchy's theorem

$$\frac{1}{\pi}\int_{x_0}^{x} G(z)dz + \frac{1}{\pi}\int_{x}^{x+i\eta} G(z)dz + \frac{1}{\pi}\int_{x+i\eta}^{x_0+i\eta} G(z)dz + \frac{1}{\pi}\int_{x_0+i\eta}^{x} G(z)dz = 0.$$

Hence

$$-\frac{1}{\pi} \mathrm{Im} \int_{x_0}^{x} G(z)dz = \frac{1}{\pi} \mathrm{Im} \int_{x}^{x+i\eta} G(z)dz + \frac{1}{\pi} \mathrm{Im} \int_{x+i\eta}^{x_0+i\eta} G(z)dz$$

$$+ \frac{1}{\pi} \mathrm{Im} \int_{x+i\eta}^{x} G(z)dz.$$

Now let $\eta \to 0^+$, the first and the third integral will tend to 0, for $\left|\frac{1}{\pi}\int_{x}^{x+i\eta} G(z)dz\right| \leq \frac{1}{\pi} M\eta$ where $M = \max|G(x)|$ in the rectangle under consideration. Similar conclusion holds for the third integral. So we have

$$-\frac{1}{\pi} \lim_{\eta \to 0^+} \mathrm{Im} \int_{x_0+i\eta}^{x+i\eta} G(z)dz = -\frac{1}{\pi} \mathrm{Im} \int_{x_0}^{x} G(u)du.$$

Taking the derivative, we obtain

$$f(x) = -\frac{1}{\pi} \mathrm{Im}\, G(x).$$

Example 2.8.1: Let ξ be a measure on $[-1,1]$ such that the corresponding Stieltjes transform is given by

$$G(z) = \frac{1}{z} - \frac{\frac{1}{4}}{z} - \frac{\frac{1}{4}}{z} - \cdots$$

It is easy to show that $G(z) = 2(z - \sqrt{z^2-1})$. Here we take the branch so that $G(z) \to 0$ as $z \to \infty$ along the real axis. It is clear that $G(z)$ is analytic on $[-1, 1]$. Thus

$$f(x) = -\frac{1}{\pi} \operatorname{Im} 2(x - \sqrt{x^2-1}) = \frac{2}{\pi}\sqrt{1-x^2}.$$

Example 2.8.2: Let ξ be a measure and its Stieltjes transform is given by

$$G(z) = \frac{1}{z} - \frac{\frac{1}{2}}{z} - \frac{\frac{1}{4}}{z} - \cdots$$

From Example 2.8.1, it can be shown that

$$G(z) = \frac{1}{\sqrt{z^2-1}}.$$

By Corollary 2.8.1, we see

$$f(x) = \frac{1}{\pi}\frac{1}{\sqrt{1-x^2}}.$$

For a measure $\sigma(\theta)$ defined on the circle, we have a transform that corresponds to the Stieltjes transform. The material below can be found in Geronimus (1961a).

Let $F(z) = \frac{1}{2\pi c_0}\int_0^{2\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z}\, d\sigma(\theta)$. $F(z)$ has a continued fraction expansion

$$F(z) = 1 - \frac{2a_0z}{1-a_0z} - \frac{a_1(1-|a_0|^2)z}{a_0+a_1z} - \frac{a_0a_2(1-|a_1|^2)z}{a_1+a_2z} - \frac{a_1a_3(1-|a_2|^2)z}{a_2+a_3z} - \cdots$$

The measure $\sigma(\theta)$ can be recovered from $F(z)$ by

Theorem 2.8.2: $\frac{\sigma(\theta+0)+\sigma(\theta-0)}{2} = \text{const} + c_0 \lim_{r\to 1^-} \int_0^\theta \text{Re}\{F(re^{i\varphi})\}d\varphi.$

Proof: See Geronimus (1961a).

Geronimus also proved the following useful corollaries.

Corollary 2.8.2: $P(\theta) = \sigma'(\theta) = \lim_{r\to 1^-} c_0 \text{ Re } F(re^{i\theta}).$

Corollary 2.8.3: $\sigma'(\theta) = c_0 \lim_{n\to\infty} \frac{h_n}{|\varphi_n^*(e^{i\theta})|^2}.$

Recall that a probability measure on [-1, 1] say ξ, is related to a measure $\sigma(\theta)$ on the circle by

$$\sigma(\theta) = \begin{cases} -\xi(\cos\theta), & 0 \le \theta \le \pi \quad x = \cos\theta, -1 \le x \le 1 \\ \xi(\cos\theta), & \pi \le \theta \le 2\pi. \end{cases}$$

Recall also that the canonical moments can be expressed in terms of the parameters that define the orthogonal polynomials on the circle, namely,

$$p_k = \frac{a_{k-1} + 1}{2} .$$

Example 2.8.3: Let ξ be the measure on [-1, 1] with given p_1, $0 < p_1 < 1$, and $p_2 = p_3 = \ldots = \frac{1}{2}$. That is equivalent to the measure on the circle with $a_0 = 2p_1-1$, $a_1 = a_2 = \ldots = 0$. So

$$\varphi_1^*(z) = 1-a_0 z = \varphi_2^*(z) = \varphi_3^*(z) = \ldots$$

By Corollary 2.8.3

$$P(\theta) = \sigma'(\theta) = \frac{2}{\pi} \frac{1-a_0^2}{1-2a_0\cos\theta+a_0^2}.$$

This is the so-called Wrapped Cauchy Distribution. Converting it back to the [-1, 1] interval, we have

$$d\xi(x) = w(x)dx = \frac{1}{\pi} \frac{1}{\sqrt{1-x^2}} \frac{1-a_0^2}{1-2a_0 x+a_0^2} . \tag{2.8.1}$$

Kiefer and Studden (1976) have shown that the limiting optimal extrapolation design has the following density

$$\frac{1}{\pi} \frac{1}{\sqrt{1-x^2}} \frac{\sqrt{x_0^2-1}}{|x_0-x|} . \tag{2.8.2}$$

Compare with (2.8.1), we see that

$$a_0 = \begin{cases} x_0 - \sqrt{x_0^2 - 1} & \text{if } \; x_0 > 1. \\ x_0 + \sqrt{x_0^2 - 1} & \text{if } \; x_0 < -1. \end{cases}$$

Using the results of Section 4, we can write down the orthogonal polynomials with respect to $d\xi(x)$ as

$$T_n(x) - a_0 T_{n-1}(x), \quad \text{for } \; n \geq 1$$

where $T_n(x)$ is the n-th Tchebycheff polynomial.

<u>Example 2.8.4</u>: Given the canonical moments P_1, P_2, $\frac{1}{2}$, $\frac{1}{2}$, $\frac{1}{2}$,.... or a_0, a_1, $a_2 = a_3 = \ldots = 0$, we have $\varphi_2^*(z) = \varphi_3^*(z) = \ldots = 1-a_0(1-a_1)z-a_1z^2$. So Corollary 2.8.3 gives

$$P(\theta) = \sigma'(\theta) = \frac{1}{\pi}\frac{(1-a_0^2)(1-a_1^2)}{1+a_0^2(1-a_1^2)+a_1^2-2a_0(1-a_1)\cos\theta+2a_0a_1(1-a_1)\cos\theta-2a_1\cos\theta}.$$

The density on [-1, 1] is given by

$$\frac{1}{\pi}\frac{1}{\sqrt{1-x^2}}\frac{(1-a_0^2)(1-a_1^2)}{(1+a_1)^2+a_0^2(1-a_1)^2-2xa_0(1-a_1)^2-4a_1x^2}. \qquad (2.8.3)$$

The orthogonal polynomials are of the form

$$T_n(x)-a_0(1-a_1)T_{n-1}(x)-a_1T_{n-2}(x), \qquad n \geq 2.$$

In particular, if $a_0 = 0$ ($P_1 = \frac{1}{2}$) (2.8.3) will be reduced to

$$\frac{1}{\pi}\frac{1}{\sqrt{1-x^2}}\frac{1-a_1^2}{(1+a_1)^2-4a_1x^2}.$$

Studden (1978) obtained the limiting design for estimating the [nq]th coefficient of the polynomial, $0 < q < 1$, when $n \to \infty$. The density of the limiting design is given by

$$\frac{q}{\pi(q^2+(1-q^2)x^2)\sqrt{1-x^2}} .$$

By comparison, we see

$$a_1 = \frac{q-1}{q+1} \quad \text{and} \quad p_2 = \frac{q}{q+1}.$$

Example 2.8.5: Given the canonical moments $\frac{1}{2}, P_2\ \frac{1}{2}, P_4, \frac{1}{2}, \frac{1}{2} \cdots$ we have $\varphi_4^*(z) = \varphi_5^*(z) = \ldots = (1-a_1z^2)-a_3z^2(z^2-a_1)$. So Corollary 2.8.3 gives

$$P(\theta) = \sigma'(\theta) = \frac{1}{\pi}\frac{(1-a_1^2)(1-a_3^2)}{1+a_1^2(1-a_3)^2+a_3^2-2a_1(1-a_3)\cos 2\theta+2a_1a_3(1-a_3)\cos 2\theta-2a_3\cos 4\theta}.$$

Then density on [-1,1] is given by

$$\frac{1}{16}\frac{1}{\pi}\frac{1}{\sqrt{1-x^2}}\frac{(1-a_1^2)(1-a_3^2)}{-a_3x^4+[a_3-a_1(\frac{1-a_3}{2})^2]x^2+(\frac{1+a_1}{2})^2(\frac{1-a_3}{2})^2} .$$

The orthogonal polynomials are given by

$$T_n(x)-a_1(1-a_3)T_{n-2}(x)-a_3T_{n-4}(x), \quad n \geq 4.$$

Example 2.8.6: Given $p_1, p_2, p_3, \ldots, p_m, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \ldots,$ we know that $\varphi_m^*(e^{i\theta}) = \varphi_{m+1}^*(e^{i\theta}) = \ldots.$ So

$$P(\theta) = \sigma'(\theta) = \frac{1}{\pi}\frac{(1-a_0^2)\ \ldots\ (1-a_{m-1}^2)}{|\varphi_m^*(e^{i\theta})|^2} .$$

The orthogonal polynomials are given by

$$\sum_{i=n-m}^{n} a_i T_i(x), \quad n \geq m$$

where a_i is the corresponding coefficients in $|\varphi_n(e^{i\theta})|^2$.

Recall that in Section 4, for given a continued fraction expansion of $\int_0^1 \frac{d\xi(x)}{z-x}$, we can write down the continued fraction expansion for $\int_0^1 \frac{x(1-x)d\xi(x)}{z-x}$, $\int_0^1 \frac{xd\xi(x)}{z-x}$ and $\int_0^1 \frac{(1-x)d\xi(x)}{z-x}$. By transformation of variable, we obtain similar results for the interval [-1, 1]. More explicitly, given

$$\int_{-1}^{1} \frac{d\xi(x)}{z-x} = \frac{1}{z+1-2\zeta_1} - \frac{4\zeta_1\zeta_2}{z+1-2\zeta_2-2\zeta_3} - \frac{4\zeta_3\zeta_4}{z+1-2\zeta_4-2\zeta_3} - \tag{2.8.4}$$

we have

$$\int_{-1}^{1} \frac{(1-x^2)d\xi(x)}{z-x} = \frac{1-c_2}{z+1-2\gamma_2-2\gamma_3} - \frac{4\gamma_3\gamma_4}{z+1-2\gamma_4-2\gamma_5} - \frac{4\gamma_5\gamma_6}{z+1-2\gamma_6-2\gamma_7} - \ldots$$

$$\int_{-1}^{1} \frac{(1+x)d\xi(x)}{z-x} = \frac{1+c_1}{z+1-2\zeta_1-2\zeta_2} - \frac{4\zeta_2\zeta_3}{z+1-2\zeta_3-2\zeta_4} - \frac{4\zeta_4\zeta_5}{z+1-2\zeta_5-2\zeta_6} - \cdots$$

$$\int_{-1}^{1} \frac{(1-x)d\xi(x)}{z-x} = \frac{1-c_1}{z+1-2\gamma_2} - \frac{4\gamma_2\gamma_3}{z+1-2\gamma_3-2\gamma_4} - \frac{4\gamma_4\gamma_5}{z+1-2\gamma_5-2\gamma_6} - \cdots$$

Example 2.8.7: Apply the preceding reasoning to Example 2.8.2, we have

$$\frac{1}{\pi}\int_{-1}^{1} \frac{\sqrt{1-x^2}dx}{z-x} = \frac{\frac{1}{2}}{z} - \frac{\frac{1}{4}}{z} - \frac{\frac{1}{4}}{z} - \cdots$$

$$\frac{1}{\pi}\int_{-1}^{1} \frac{\sqrt{\frac{1+x}{1-x}}\,dx}{z-x} = \frac{1}{z-\frac{1}{2}} - \frac{\frac{1}{4}}{z} - \frac{\frac{1}{4}}{z} - \cdots$$

$$\frac{1}{\pi}\int_{-1}^{1} \frac{\sqrt{\frac{1-x}{1+x}}}{z-x} = \frac{1}{z+\frac{1}{2}} - \frac{\frac{1}{4}}{z} - \frac{\frac{1}{4}}{z} - \cdots$$

Denote their orthogonal polynomials by $U_n(x)$, $W_n(x)$ and $V_n(x)$ respectively. It is obviously from the above that they satisfy the same recursion formula but with different initial conditions i.e. $U_1(x) = x$, $W_1(x) = x - \frac{1}{2}$ and $V_1(x) = x + \frac{1}{2}$.

The following theorem gives the weighted Tchebycheff polynomials for some weight functions.

Theorem 2.8.3: Given a positive polynomial P(t) of degree q on [-1, 1], then

(i) $$\min_{\{b_k\}} \max_{-1\le x\le 1} \frac{|t^m+b_1t^{m-1} +\cdots+ b_m|}{\sqrt{P(t)}} \qquad m \ge [\tfrac{q}{2}]$$

is attained if $t^m + b_1t^{m-1} +\cdots+ b_m$ is the m-th orthogonal polynomial with respect to $\frac{1}{\pi P(t)\sqrt{1-t^2}}$

(ii) $$\min_{\{b_k\}} \max_{-1\le x\le 1} \frac{|t^m+b_1t^{m-1}+\ldots+b_m|}{\sqrt{P(t)}}\sqrt{1-t^2}, \quad m \ge [\tfrac{q}{2}],$$

is attained if $t^m + b_1t^{m-1} +\ldots+ b_m$ is the m-th orthogonal polynomial with respect to

$$\frac{1}{\pi}\frac{1}{P(t)}\sqrt{1-t^2}.$$

(iii) $$\min_{\{b_k\}} \max_{-1\le x\le 1} \frac{|t^m+b_1t^{m-1}+\ldots+b_m|}{\sqrt{P(t)}}\sqrt{1-t}, \quad m \ge [\tfrac{q}{2}],$$

is attained if $t^m + b_1t^{m-1} +\ldots+ b_m$ is the m-th orthogonal polynomial with respect to

$$\frac{1}{\pi}\frac{1}{P(t)}\sqrt{\frac{1-t}{1+t}},$$

(iv) $$\min_{\{b_k\}} \max_{-1\le x\le 1} \frac{|t^m+h_1t^{m-1}+\ldots+b_m|}{\sqrt{P(t)}}\sqrt{1+t}, \quad m \ge [\tfrac{q}{2}],$$

is attained if $t^m + b_1t^{m-1} +\ldots+ b_m$ is the m-th orthogonal polynomial with respect to

$$\frac{1}{\pi}\frac{1}{P(t)}\sqrt{\frac{1+t}{1-t}}.$$

Proof: See Krein and Nudelman (1977).

The following will show how the limiting properties of the sequence of canonical moments determine the properties of the measure.

Theorem 2.8.4: If $\sum_{k=0}^{\infty} |a_k| < \infty$, then $\sigma(\theta)$ is absolutely continuous and $P(\theta) = \sigma'(\theta)$ is continuous and positive.

Proof: See Geronimus (1961a).

Observing that $a_k = 2p_{k+1} - 1$ and the relationship between the measure ξ on the interval $[-1, 1]$ and the measure σ on the circle, we have

Corollary 2.8.4: If $\sum_{k=1}^{\infty} |p_k - \frac{1}{2}| < \infty$, then ξ is absolutely continuous and $w(x) = \xi'(x)$ is continuous and positive on $(-1, 1)$.

Nevai (1979) defined the quantity

$$d_k = |\alpha_k| + |\frac{\gamma_{k-1}}{\gamma_k} - \frac{1}{2}| + |\frac{\gamma_k}{\gamma_{k+1}} - \frac{1}{2}|.$$

In our notation,

$$d_k = |1-2\zeta_{2k}-2\zeta_{2k+1}| + 2|\sqrt{\zeta_{2k-1}\zeta_{2k}} - \frac{1}{4}| + 2|\sqrt{\zeta_{2k-3}\zeta_{2k-2}} - \frac{1}{4}|.$$

Theorem 2.8.5: If $\sum_{k=1}^{\infty} d_k < \infty$, $d\xi(x) = w(x)dx$ is continuous and positive on $(-1, 1)$.

Proof: See Nevai (1979).

Theorem 2.8.6: $\sum_{k=1}^{\infty} |P_k - \frac{1}{2}| < \infty \Rightarrow \sum_{k=1}^{\infty} d_k < \infty$.

Proof: We want to show $\sum_{k=1}^{\infty} |P_k - \frac{1}{2}| < \infty$ implies $\sum_{n=1}^{\infty} |q_{n-1}p_n - \frac{1}{4}| < \infty$, where $q_0 = 1$. Notice that $q_{n-1}P_n - \frac{1}{4} = P_n - \frac{1}{2} - (P_{n-1}P_n - \frac{1}{4})$ and

$$|P_{n-1}P_n - \frac{1}{4}| \leq |P_{n-1}P_n - \frac{1}{4} - \frac{1}{2}(P_{n-1} - \frac{1}{2}) - \frac{1}{2}(P_n - \frac{1}{2})| + \frac{1}{2}|P_{n-1} - \frac{1}{2}| + \frac{1}{2}|P_{n-1} - \frac{1}{2}|.$$

Now the convergence of $\Sigma|P_{n-1} - \frac{1}{2}||P_n - \frac{1}{2}|$ would imply the convergence of the first term on the right hand side of the last expression. So $\sum|q_{n-1}P_n - \frac{1}{4}| < \infty$. Using the same method, we see $\sum|q_{n-1}P_nq_nP_{n+1} - \frac{1}{16}| < \infty$. This would imply $\Sigma|\sqrt{q_{n-1}P_nq_nP_{n+1}} - \frac{1}{4}| < \infty$ since

$$\frac{q_{n-1}P_nq_nP_{n+1} - \frac{1}{16}}{\sqrt{q_{n-1}P_nq_nP_{n+1}} - \frac{1}{4}} \to \frac{1}{2}.$$

Also notice that

$$1 - 2\zeta_{2n} - 2\zeta_{2n+1} = 2(\frac{1}{4} - \zeta_{2n} + \frac{1}{4} - \zeta_{2n+1}).$$

So the convergence of $\Sigma|1-2\zeta_{2n} - 2\zeta_{2n+1}|$ follows immediately.

Theorem 2.8.7: The followings are equivalent:

(i) $\Sigma|a_k|^2 < \infty$

(ii) $\int_0^{2\pi} \log \sigma'(\theta)d\theta > -\infty$

(iii) $\lim_{n\to\infty} \frac{\Delta_n}{\Delta_{n-1}} = \lim_{n\to\infty} c_0 \prod_{n=0}^{n-1} (1-|a_0|^2) = \frac{1}{\pi} \exp\{\frac{1}{2\pi} \int_0^{2\pi} \log \sigma'(\theta)d\theta\}.$

Proof: See Geronimus (1948).

On the interval [-1, 1], we have

Corollary 2.8.7a: The followings are equivalent.

(i) $\sum_{k=1}^{\infty} |P_k - \frac{1}{2}|^2 < \infty$

(ii) $\int_{-1}^{1} \frac{\log w(x)}{\sqrt{1-x^2}} dx > -\infty$

(iii) $$\lim_{n\to\infty} 2^{2n+1} \frac{|M_n(\xi)|}{|M_{n-1}(\xi)|} = 2\pi \exp\{\frac{1}{\pi} \int_{-1}^{1} \frac{\log w(x)}{\sqrt{1-x^2}} dx\}.$$

Here $w(x)$ denotes the derivative of the measure ξ.

Proof: See Geronimus (1948).

Corollary 2.8.7b: (i) $$\sum_{k=0}^{\infty} |a_k|^2 < \infty \Rightarrow \lim_{n\to\infty} (\Delta_n)^{\frac{1}{n+1}}$$

$$= \exp\{\frac{1}{2\pi} \int_0^{2\pi} \log \sigma'(\theta) d\theta\}$$

(ii) $$\sum_{k=1}^{\infty} |P_k - \frac{1}{2}|^2 < \infty \Rightarrow \lim_{n\to\infty} 2^{n+1} |M_n(\xi)|^{\frac{1}{n+1}} = 2\pi \exp\{\frac{1}{\pi} \int_{-1}^{1} \frac{\log w(x)}{\sqrt{1-x^2}} dx\}.$$

Proof: (i) follows from Theorem 2.8.7 (iii) and (ii) follows from Corollary 2.8.7a (iii).

Theorem 2.8.8: Let $\sum_{k=1}^{\infty} |P_k - \frac{1}{2}|^2 < \infty$ and $w(x)$ denote the derivative of ξ. If $\frac{1}{w(x)} \in L_1[-1, 1]$, then

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} P_k^2(x) = \frac{1}{\pi w(x) \sqrt{1-x^2}}$$

for almost every $x \in \tau \subset (-1, 1)$.

Proof: See Nevai (1979).

Theorem 2.8.9: $\Sigma\, d_k < \infty \Rightarrow \Sigma |P_k - \frac{1}{2}|^2 < \infty$.

Proof: See Nevai (1979).

Theorem 2.8.10: The followings are equivalent:

(i) $\lim_{n\to\infty} a_n = 0$

(ii) $\lim_{n\to\infty} \frac{\varphi_{n+1}(z)}{\varphi_n(z)} = z$ uniformly for $z > 1$.

Proof: See Geronimus (1961b).

Corollary 2.8.10: The following are equivalent:

(i) $\lim_{n\to\infty} P_n = \frac{1}{2}$

(ii) $\lim_{n\to\infty} \frac{P_{n+1}(x)}{P_n(x)} = x + \sqrt{x^2-1}$ for $|x + \sqrt{x^2-1}| > 1$.

Proof: See Gerominus (1961b).

Theorem 2.8.11: The following are equivalent:

(i) $\lim_{n\to\infty} P_n = \frac{1}{2}$

(ii) $\lim_{n\to\infty} \zeta_n = \frac{1}{4}$

(iii) $\lim_{n\to\infty} \zeta_{2n} + \zeta_{2n+1} = \frac{1}{2}$, $\lim_{n\to\infty} \zeta_{2n-1}\zeta_{2n} = \frac{1}{16}$

(iv) $\lim_{n\to\infty} d_k = 0$.

Proof: It is obvious that (iii) $\Leftrightarrow$ (iv) and (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii). We want to prove (iii) $\Rightarrow$ (i). Let ξ be a measure with the properties that $\lim_{n\to\infty}(\zeta_{2n}+\zeta_{2n+1}) = \frac{1}{2}$ and $\lim_{n\to\infty} \zeta_{2n-1}\zeta_{2n} = \frac{1}{16}$. Then by Theorem 2.4.11, the limit of the ratio of its orthogonal polynomials is

$$\lim_{n\to\infty} \frac{P_{n+1}(x)}{P_n(x)} = x - \frac{\frac{1}{4}}{x - \frac{\frac{1}{4}}{x - \dots}} = x + \sqrt{x^2-1}$$

By Corollary 2.8.10, this implies $P_n \to \frac{1}{2}$.

We will discuss the asymptotic distribution of the zeros of the orthogonal polynomials on a closed bounded interval.

Let $P_n(x) = (x-x_{1n})(x-x_{2n})\dots(x-x_{nn})$, where $x_{1n} < x_{2n} < \dots < x_{nn}$, be orthogonal with respect to $d\xi$ on $[a,b]$. Let us denote by $N_n(d_\alpha, t)$ the number of integers k for which

$$x_{nn} - x_{kn} \geq t[x_{nn}-x_{1n}]$$

holds. If $\lim_{n\to\infty} n^{-1} N_n(d\xi,t) = \frac{1}{2} - \frac{1}{\pi}\sin^{-1}(2t-1)$, $t \in [0,1]$, we say ξ is regular.

Theorem 2.8.12: ξ is regular iff

$$\lim_{n\to\infty} \sqrt[n]{P_n(x)} = \frac{1}{2}(x + \sqrt{x^2-1}).$$

Proof: Since $\lim_{n\to\infty} \frac{1}{n} \ln P_n(x) = \int_{-1}^{1} \ln(x-t)d\psi(t)$, where $\psi(t)$ is the limiting distribution of the roots of the orthogonal polynomials. Notice that

$$\frac{d}{dx}\int_{-1}^{1} \ln(x-t)d\psi(t) = \int_{-1}^{1} \frac{d\psi(t)}{x-t}$$

which is the Stieltjes transform of the measure $d\psi(t)$. So

$\lim_{n\to\infty} \sqrt[n]{P_n(x)} = \frac{1}{2}(x + \sqrt{x^2-1})$ iff

$$\int_{-1}^{1} \frac{d\psi(t)}{x-t} = \frac{d}{dx} \ln \frac{1}{2}(x + \sqrt{x^2-1})$$

$$= \frac{1}{\sqrt{x^2-1}}$$

From Example 2.8.2, this is the Stieltjes transform of the arc-sine distribution. Hence the theorem is proved.

In below we will give a restricted form of Walsh's Theorem since we do not want to bring in so many definitions. It should be noted that the original form is more general.

Theorem 2.8.13 (Walsh): Let [a,b] be a closed bounded interval on the real line. If all the zeros of a given system of polynomials $\{P_n(z)\}$ have no limit points outside [a,b] and if $\max_{x\in[a,b]} |P_n(x)| = M_n$, then $\lim_{n\to\infty} \sqrt[n]{(P_n(z))} = \frac{b-a}{4} |\varphi(z)|$ where $\varphi(z) = \frac{2}{b-a} z - \frac{b+a}{b-a} + \sqrt{(\frac{2}{b-a} z - \frac{b+a}{b-a})^2 - 1}$ for all z in the complex plane outside of [a,b], iff $\lim_{n\to\infty} \sqrt[n]{M_n} = \frac{b-a}{4}$.

Remark 2.8.1: If $b = -a = 1$ then

$$\varphi(z) = z + \sqrt{z^2-1}.$$

Let ξ be regular with canonical moments $P_1, P_2, \ldots$ and is defined on [0,1]. Define $\tilde{\xi}$ to be the measure with canonical moments $\frac{1}{2}, P_1, \frac{1}{2}, P_2, \frac{1}{2}, P_3, \ldots$ and is defined on [-1, 1]. It is known that the orthogonal polynomials $\tilde{P}_n(x)$ of $\tilde{\xi}$ on [-1, 1] can be expressed as

$$\tilde{P}_{2n}(x) = P_n(x^2)$$
$$\tilde{P}_{2n+1}(x) = xR_n(x^2)$$

where $P_n(x)$ and $R_n(x)$ is the orthogonal polynomial with respect to $d\xi$ and $xd\xi$ respectively. It is well known that the zeros of $P_n(x)$ and $R_n(x)$ are interlaced, so $xd\xi$ is also regular and hence $\tilde{\xi}$ is also regular. The converse is true also. We are now ready to prove

Theorem 2.8.14: ξ is regular iff $\lim_{n\to\infty} \sqrt[n]{\zeta_1 \cdots \zeta_n} = \frac{1}{4}$.

Proof: Geronimus (1961b) has proved that

$\lim_{n\to\infty} \sqrt[2n]{(4\zeta_1\zeta_2)\cdots(4\zeta_{2n-1}\zeta_{2n})} = \frac{1}{2}$ is sufficient for $\lim_{n\to\infty} \sqrt[n]{M_n} = d$.
Hence $\lim_{n\to\infty} \sqrt[n]{\zeta_1 \cdots \zeta_n} = \frac{1}{4}$ implies that ξ is regular. On the other hand if ξ is regular, then so is $\tilde{\xi}$ according to the above discussion. Now
$\lim_{n\to\infty} \sqrt[2n]{|\tilde{P}_{2n}(z)|} = \frac{1}{2}(z + \sqrt{z^2-1})$ for all $z \notin [-1, 1]$.

But $\tilde{P}_n(0) = \zeta_1 \cdots \zeta_n \neq 0$, so $\sqrt[2n]{|\tilde{P}_{2n}(z)|}$ is analytic at the point 0. Hence

$$\lim_{n\to\infty} \sqrt[2n]{|\tilde{P}_{2n}(0)|} = \frac{1}{2}$$

i.e.

$$\lim_{n\to\infty} \sqrt[n]{\zeta_1 \cdots \zeta_n} = \frac{1}{4}.$$