Distributional transformations, orthogonal polynomials, and Stein characterizations

LARRY GOLDSTEIN[13] and GESINE REINERT[23]

Abstract

A new class of distributional transformations is introduced, characterized by equations relating function weighted expectations of test functions on a given distribution to expectations of the transformed distribution on the test function's higher order derivatives. The class includes the size and zero bias transformations, and when specializing to weighting by polynomial functions, relates distributional families closed under independent addition, and in particular the infinitely divisible distributions, to the family of transformations induced by their associated orthogonal polynomial systems. For these families, generalizing a well known property of size biasing, sums of independent variables are transformed by replacing summands chosen according to a multivariate distribution on its index set by independent variables whose distributions are transformed by members of that same family. A variety of the transformations associated with the classical orthogonal polynomial systems have as fixed points the original distribution, or a member of the same family with different parameter.

1 Introduction

The zero bias transformation was introduced in [13]. This mapping enjoys properties similar to those of the well known size biased transformation (see e.g. [15]) on non-negative variables, but can be applied to mean zero random variables. One main feature of the zero bias transformation is that its unique fixed point is the mean zero normal distribution, and for this reason it has been applied in Stein's method for the purpose of normal approximation

[1]Department of Mathematics, University of Southern California, Los Angeles CA 90089-2532, USA

[2]Department of Statistics, University of Oxford, Oxford OX1 3TG, UK, supported in part by EPSRC grant no. GR/R52183/01

[3]This work was partially completed while the authors were visiting the Institute for Mathematical Sciences, National University of Singapore, in 2003. The visit was supported by the Institute.

([13], [11], [12], and [14]). The zero bias transformation is also related to the K_i function in the work [16], [5], [22], and [6] and others; for a good overview see [8].

We place the classical size bias transformation and the zero bias transformation in a broader context, showing that both are a particular case of transforming a given distribution X into $X^{(P)}$ through the use of a measurable 'biasing' function P. To be more precise, for the given X and P let $\mathcal{C}^m$ denote the collection of functions whose m^{th} derivative exists and is measurable on $\mathbf{R}$, and suppressing X on the left hand side, set

$$\mathcal{F}^m(P) = \{F \in \mathcal{C}^m : E|P(X)F(X)| < \infty\}.$$

We consider transformations characterized by

$$EP(X)F(X) = \alpha EF^{(m)}(X^{(P)}) \quad \text{for all } F \in \mathcal{F}^m(P), \tag{1}$$

where necessarily $\alpha = (m!)^{-1}EP(X)X^m$ when $X^m \in \mathcal{F}(P)$; we insist $\alpha > 0$. We coin this distribution the $X - P$ biased distribution. For discrete distributions the differential operator is replaced by the difference operator; see Sections 4.3 and 4.4.

Theorem 2.1 provides our most general conditions on the existence of transformations characterized by (1), where the 'biasing' function P is only required to have m sign changes and satisfy certain orthogonality and positivity conditions; we call m the order of the resulting transformation. In the particular case where P is a polynomial, the sign change condition can be expressed in terms of the roots and order of the polynomial P, and the orthogonality properties in terms of moments. For example, for each $m = 0, 1, \ldots$ there exists a distributional transformation which is defined using the Hermite polynomial of order m as the biasing function, and whose domain are those distributions whose first $2m$ moments match those of the mean zero normal; the case $m = 1$ corresponds to the zero bias transformation of [13].

Theorem 2.1 in Section 2 shows distributional transformations exist in great generality. In Section 3 we find that there is considerable additional structure for the families of transformations induced by orthogonal polynomial systems, especially those corresponding to families of distributions which are closed under addition of independent variables. Corresponding to the Normal, Gamma, Poisson, Binomial and Beta-type distributions, in Section 4 we study the family of transformations defined using the Hermite, Laguerre, Charlier, Krawtchouk, and Gegenbauer polynomials, and obtain high order Stein type characterizing equations.

Our work here is in the spirit of [9], where other fundamental connections between Stein equations and orthogonal polynomials were first described. The approach in [9] is iterated in [24], combining it with well-known connections between orthogonal polynomials and birth and death processes, and used as in [17] to describe solutions of Stein equations.

We first review some well known facts regarding the size bias transformation, which is the simplest and best known of all these distributional transformations. For non-negative X with $0 < EX = \mu < \infty$, the X-size biased distribution X^s is defined by the characterizing equation

$$EXF(X) = \mu EF(X^s) \quad \text{for all } F \in \mathcal{F}^0(X). \tag{2}$$

One key feature of the sized bias transformation is the following. If $X_1, \ldots, X_n$ are independent non-negative variables with finite positive expectations $EX_i = \mu_i$ and

$$W = \sum_{i=1}^{n} X_i,$$

then a variable with the W-size biased distribution can be constructed by replacing a variable X_i, chosen with probability proportional to μ_i, by an independent variable X_i^s having the X_i-size biased distribution. In other words, letting

$$P(I = i) = \frac{\mu_i}{\sum_{j=1}^{n} \mu_j}$$

be independent of $X_1, \ldots, X_n$, the variable

$$W^s = W - X_I + X_I^s \tag{3}$$

has the W-size biased distribution. Letting $x^+ = \max(0, x)$, size biasing is the case of (1) with biasing function $P(x) = x^+$. This transformation is of order zero, as there are $m = 0$ sign changes of x^+ on $\mathbf{R}$, and has $\alpha = EX^+$; when $X \geq 0$ we have $X^+ = X$ resulting in the usual characterization (2).

The zero bias transformation [13] was motivated by the similarity between the size bias transformation and the Stein equation [21] for the mean zero normal distribution. In particular, Stein's identity says that $Z \sim \mathcal{N}(0, \lambda)$ if and only if

$$EZF(Z) = \lambda EF'(Z) \quad \text{for all } F \in \mathcal{F}^1(Z). \tag{4}$$

Comparing (4) to (2), for a mean zero, positive variance λ variable X, we say that X^z has the X-zero biased distribution if

$$EXF(X) = \lambda EF'(X^z) \quad \text{for all } F \in \mathcal{F}^1(X). \tag{5}$$

Note that (5) for zero biasing is the same as (2) for size biasing, but with variance replacing mean, and F' replacing F. That the normal distribution with variance λ is the unique fixed point of the zero bias transformation follows immediately from the characterization (4). It was shown in [13] that the zero bias distribution X^z exists for all X that have mean zero and finite positive variance. Its existence follows also from Theorem 2.1, as the special case of (1) for the function $P(x) = x$, having $m = 1$ sign changes on $\mathbf{R}$, and α equal to the variance λ of X.

The zero bias transformation was introduced and used in [13] to obtain bounds of order n^{-1} in normal approximations for smooth test functions under third order moment conditions, in the presence of dependence induced by simple random sampling. In [11] it is used to provide bounds to the normal distribution for hierarchical sequences generated by the iteration of a so called averaging function, in [12] for normal approximation in combinatorial central limit theorems with random permutations having distribution constant over cycle type, and in [14] the extension of the zero bias transformation to higher dimension is considered.

The zero bias transformation enjoys a property similar to (3) for size biasing. In particular, it was shown in [13] that a sum of independent mean zero variables with finite variances can be zero biased by replacing one variable chosen with probability proportional to its variance by an independent variable from that summands zero biased distribution. Precisely, let $X_1, \ldots, X_n$ be independent mean zero variables with variance $\lambda_i = EX_i^2 > 0$,

$$W = X_1 + \cdots + X_n,$$

and I a random index, independent of $X_1, \ldots, X_n$ with distribution

$$P(I = i) = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}. \tag{6}$$

Then

$$W^z = W - X_I + X_I^z \tag{7}$$

has the W-zero biased distribution, where X_i^z is a variable independent of $X_j, j \neq i$ having the X_i zero biased distribution. This construction is extended to the families of transformations associated with orthogonal polynomial in Theorem 3.1. In particular, in Section 3 we see that for higher order transformations sums of independent variables are transformed by replacing multiple variables chosen according to some distribution (e.g. multinomial, multivariate hypergeometric) with independent variables possessing distributions transformed by the same family.

In Section 2 we give the moment and sign change conditions on P which guarantee the existence of the $X - P$ distribution and provide an explicit construction. In Section 3 we treat the special case where P is a member of a family of orthogonal polynomials. The generalization to higher order of the 'replace one variable' zero and size bias constructions is based on the identity (25) expressing an orthogonal polynomial of a sum as a sum of like polynomials with summands having no larger order, and is given in Section 3. In Sections 4.1, 4.2, 4.3, 4.4 and 4.5 we treat the Hermite, Laguerre, Charlier, Krawtchouk and Gegenbauer polynomials, corresponding to the Normal, Gamma, Poisson, Binomial and Beta-type distributions respectively. Special instances of the Beta-type distributions we consider are the uniform $\mathcal{U}[-1,1]$, the arcsine, and the semi-circle distribution.

2 Transformations in General

We begin our study with the following existence and uniqueness theorem for the types of distributional transformations under consideration. We say the measurable function P on $\mathbf{R}$ is positive on an interval I if $P(x) \geq 0$ for all $x \in I$ with strict inequality for at least one x, and similarly for P negative on I. We say P has exactly $m = 0, 1, \ldots$ sign changes if $\mathbf{R}$ can be partitioned into $m + 1$ disjoint subintervals with non-empty interior such that P alternates sign on successive intervals. Though the choices for the endpoints of such intervals may be somewhat arbitrary when there are intervals where P is zero, we will nevertheless say that a sign change occurs at the interval boundaries; the uniqueness guaranteed by Theorem 2.1 shows that the $X - P$ biased distribution constructed in the proof of Theorem 2.1 is the same for all interval boundary choices, and Example 2.1 gives some additional explanation of this phenomenon in the context of a particular example. We note that for existence in general, regarding boundedness, the orthogonality conditions required by Theorem 2.1 are only relative to P and required only up to a finite order; such conditions may not impose boundedness on any of the power moments of X, as illustrated in Example 2.1.

Theorem 2.1 *Let X be a random variable, $m \in \{0, 1, 2, \ldots\}$ and P a measurable function with exactly m sign changes, positive on its rightmost interval and*

$$\frac{1}{m!} E X^k P(X) = \alpha \delta_{k,m} \quad k = 0, \ldots, m, \tag{8}$$

with $\alpha > 0$. Then there exists a unique distribution for a random variable $X^{(P)}$ such that

$$EP(X)F(X) = \alpha E F^{(m)}(X^{(P)}) \quad \textit{for all } F \in \mathcal{F}^m(P). \tag{9}$$

Theorem 2.1 says that X is in the domain of the distributional transformation of order m defined using the 'biasing' function P having m sign changes when the powers of X smaller than m are orthogonal to $P(X)$ in the $L^2(X)$ sense, that is, when $P(X) \in \{1, X, \ldots, X^{m-1}\}^\perp$, and $EX^m P(X) > 0$. As noted above, the existence of both the size and zero bias transformations are both special cases.

Proof of Theorem 2.1. We give an explicit construction of the variate $X^{(P)}$. By replacing P by P/α, it suffices to prove the theorem for $\alpha = 1$. Label the points where the m sign changes of P occur as $r_1, \ldots, r_m$, and let

$$Q(x) = \prod_{i=1}^{m} (x - r_i), \tag{10}$$

adopting the usual convention that an empty product is 1. By construction Q and P have the same sign, so letting μ_X denote the distribution of X,

$$d\mu_Y(y) = \frac{1}{m!} Q(y) P(y) d\mu_X(y) \tag{11}$$

is therefore a measure, and since (8) with $k = m$ implies that $EQ(X)P(X) = m!$, a probability measure. Now with Y and $\{U_i\}_{i \geq 1}$ mutually independent with Y having distribution μ_Y and U_j having distribution function u^i on $[0, 1]$, with $r_0 = Y$ and $r_{m+1} = 0$, we claim that

$$X^{(P)} = \sum_{k=1}^{m+1} \left(\prod_{i=k}^{m} U_i \right) (r_{k-1} - r_k) \tag{12}$$

satisfies (9), thus proving the existence of the $X-P$ biased distribution.

We begin by noting that for any F for which either side below exists,

$$EF(Y)=\frac{1}{m!}EF(X)Q(X)P(X), \tag{13}$$

and so for $k=0,\ldots,m$, letting

$$R_k(x)=\prod_{i=k+1}^{m}(x-r_i),$$

a polynomial of degree $m-k$, by (13) and (8) we have

$$E(1/\prod_{i=1}^{k}(Y-r_i))=\frac{1}{m!}ER_k(X)P(X)=\delta_{m-k,m}. \tag{14}$$

We show the claim by induction. In particular, for $k\geq 1$ letting

$$V_k=\prod_{i=k}^{m}U_i, \quad W_k=\sum_{j=k}^{m+1}V_j(r_{j-1}-r_j), \tag{15}$$

and taking $X^{(P)}$ as in (12), we show that for all $F\in\mathcal{C}_c^\infty$, the collection of infinitely differentiable functions with compact support, and $k=0,\ldots,m$

$$EF^{(m)}(X^{(P)})=k!\,E\left\{\frac{F^{(m-k)}(V_{k+1}(Y-r_{k+1})+W_{k+2})}{V_{k+1}^k\prod_{i=1}^{k}(Y-r_i)}\right\}. \tag{16}$$

We see the expectation on the right hand exists since F and all its derivatives are bounded, V_{k+1} is independent of Y for all k, $EU_i^{-k}<\infty$ for $i\geq k+1$, and use of (14).

The case $k=0$ is the statement that $X^{(P)}=V_1(Y-r_1)+W_2$, which follows from definitions (12) and (15). Assume (16) holds for some $0\leq k<m$. Using $V_{k+1}=U_{k+1}V_{k+2}$ in (16) and taking expectation over U_{k+1}, with density $(k+1)u_{k+1}^k$, we obtain

$$\begin{aligned} &EF^{(m)}(X^{(P)})\\ &= (k+1)!E\int_0^1\left\{\frac{F^{(m-k)}(u_{k+1}V_{k+2}(Y-r_{k+1})+W_{k+2})}{u_{k+1}^kV_{k+2}^k\prod_{i=1}^{k}(Y-r_i)}\right\}u_{k+1}^k du_{k+1}.\end{aligned}$$

Cancelling u_{k+1}^k and integrating, we obtain

$$(k+1)!E\left\{\frac{F^{(m-(k+1))}(V_{k+2}(Y-r_{k+1})+W_{k+2})-F^{(m-(k+1))}(W_{k+2})}{V_{k+2}^{k+1}\prod_{i=1}^{k+1}(Y-r_i)}\right\}. \tag{17}$$

Using the independence of V_{k+2} and Y for any k, and that W_{k+2} is independent of Y for all $k \geq 0$, the second term in the expectation (17) vanishes by (14), since $k+1 \geq 1$. The induction is completed by noting that definitions (15) give that $V_{k+2}(Y - r_{k+1}) + W_{k+2} = V_{k+2}(Y - r_{k+2}) + W_{k+3}$.

Now applying (16) for $k = m$ and using $V_{m+1} = 1, W_{m+2} = 0$ and $r_{m+1} = 0$ we obtain

$$EF^{(m)}(X^{(P)}) = m!E\left\{\frac{F(Y)}{Q(Y)}\right\} = EP(X)F(X)$$

by (13). That is, the equality in (9) holds for all $F \in \mathcal{C}_c^\infty$.

For $F \in \mathcal{F}^m(X)$, by replacing F by

$$F(x) - \sum_{j=0}^{m-1} \frac{F^{(j)}(0)}{j!} x^j$$

if necessary, we may assume, in light of (8), that $F^{(j)}(0) = 0$ for $j = 0, \ldots, m-1$, and hence,

$$\text{with} \quad If = \int_0^x f, \quad F(x) = I^m f \quad \text{for some measurable function } f.$$

Since $F = F_1 - F_2$ where $F_1(x) = I^m f^+$ and $F_2(x) = I^m f^-$, it suffices by linearity to consider $f \geq 0$. Letting $0 \leq f_n \uparrow f$ we have $I^m f_n = F_n \uparrow F$, and hence the equality in (9) holds for $F \in \mathcal{F}^m(X)$ using the monotone and dominated convergence theorems on the right and left sides of (9), respectively.

The distribution $X^{(P)}$ is unique since (9) holds for all $F \in \mathcal{C}_c^\infty$, which is separating. ■

The existence of the $X^{(P)}$ distribution also follows from the Riesz representation theorem upon demonstrating the positivity of the linear operator T defined by

$$Tf = EP(X)F(X) \quad \text{with } F(x) = I^m f$$

over $f \in \mathcal{C}_c^0$, the space of continuous functions with compact support. The signed measure $d\mu = P d\mu_X$ has the property $\int x^j d\mu = EX^j P(X) = 0$ for $j = 0, 1, \ldots, m-1$, and now the sign change property of P allows us, when on the finite interval $[a, b]$, to invoke Theorem 5.4 in Chapter XI of [19] (see also Example 1.4 in Chapter XI) to conclude T is positive and hence $Tf = \int_a^b f d\mu^{(m)}$ for some measure $\mu^{(m)}$, which is a probability measure since

$EX^m P(X) = m!$. This argument is similar to the one used in [13] to prove the existence of the zero bias distribution for a mean zero, finite variance X by noting that when $f \geq 0$ the function $F = If$ is non-decreasing, and hence X and $F(X)$ are positively correlated, and so the operator

$$Tf = EXF(X) \geq EXEF(X) = 0$$

is positive.

Example 2.1 *Consider the application of Theorem 2.1 where $P(x)$ has exactly $m = 1$ sign change at $r_1 = 0$. Then for the non constant X to be in the domain of the transformation characterized by*

$$EP(X)F(X) = \alpha EF'(X^{(P)}) \tag{18}$$

we require $EP(X) = 0$ and $\alpha = EXP(X) > 0$. We have $Q(x) = x$ in (10) and, recalling the X variable in the proof was rescaled to have $\alpha = 1$, the Y distribution in (11) is

$$d\mu_Y(y) = xP(x)d\mu_X(y)/\alpha.$$

From (12) with $m = 1$, $r_0 = Y, r_2 = 0$ and U_j with distribution function u^j on $[0,1]$

$$X^{(P)} = \sum_{k=1}^{m+1} \left(\prod_{i=k}^{m} U_i \right) (r_{k-1} - r_k) = U_1(r_0 - r_1) + (r_1 - r_2) = U_1 Y.$$

Hence $X^{(P)}$ is absolutely continuous, and one can directly verify that its density is given by

$$f^{(P)}(x) = \alpha^{-1} E[P(X); X > x]. \tag{19}$$

When $\int_0^x P(u)du$ is finite for all x and $c = \int \exp(-\alpha^{-1} \int_0^x P(u)du)dx < \infty$, the transformation (18) has a fixed point at the distribution with density

$$f(x) = c^{-1} \exp\left(-\frac{1}{\alpha} \int_0^x P(u)du \right);$$

for instance, when $P(x) = x$, f is the mean zero normal density with variance α.

Taking P to be the sign function

$$P(x) = \mathbf{1}(x > 0) - \mathbf{1}(x < 0)$$

provides an example of a transformation given by a discontinuous P, and shows that generally the orthogonality conditions may not reduce to restrictions on the moments of X, in particular, (8) for $k = 0$ requires X to have median 0. If in addition $\alpha = E|X|$ is finite, imposed by (8) for $k = 1$, Theorem 2.1 gives that X is in domain of the transformation characterized by (18). The density of the transformed variables are, by (19),

$$f^{(P)}(x) = \begin{cases} P(X > x)/E|X| & x > 0 \\ P(X < x)/E|X| & x < 0. \end{cases} \tag{20}$$

For this choice of P the Y distribution in (11) becomes

$$d\mu_Y(y) = |y| d\mu_X(y)/E|X|,$$

which is the $|X|$ size biased distribution. Hence, the $X-P$ biased distribution is obtain by multiplying $Y \sim \mu_Y$ by an independent $\mathcal{U}[0,1]$ variable. The transformation has a fixed point at the Laplace distribution with density

$$f(x) = \frac{1}{2\alpha} exp\left(-\frac{1}{\alpha}|x|\right).$$

Taking $P(x) = \mathbf{1}(x > 1) - \mathbf{1}(x < -1)$ gives a transformation having domain those variables X with $\alpha = E(|X|\mathbf{1}(|X| > 1)) < \infty$ and satisfying

$$P(X > 1) = P(X < -1). \tag{21}$$

Since $P(x) = 0$ in the set $[-1,1]$ the sign change can be said to occur at point in $(-1,1)$ and the polynomial Q in the proof of Theorem 2.1 can be taken to be

$$Q(x) = x - r_1 \quad \textit{for any } r_1 \in (-1,1).$$

As assured by uniqueness, the distribution constructed in the proof of Theorem 2.1 does not depend on choice of r_1; in fact, in this case (21) implies that the dmu_Y distribution in (11) is the same for all $r_1 \in (-1,1)$.

3 Transformations using orthogonal polynomials

We consider a system of polynominals orthogonal with respect to a non-trivial family of distributions $Z_\lambda \sim \mathcal{L}_\lambda$ indexed by a real parameter λ.

Condition 3.1 *For some $m \geq 0$, the polynomials $\{P_\lambda^k(x)\}_{0 \leq k \leq m}$ are monic, have degree k, are orthogonal with respect to the distributional family $Z_\lambda \sim \mathcal{L}_\lambda$, and satisfy $E[P_\lambda^k(Z_\lambda)]^2 > 0$.*

Note that since P_λ^k is monic and orthogonal it has k distinct roots and is positive as $x \to \infty$ (e.g. [1]); furthermore, we have

$$EZ_\lambda^k P_\lambda^k(Z_\lambda) = E[P_\lambda^k(Z_\lambda)]^2, \quad k = 0, \ldots, m.$$

When studying transformations using an implicit family of orthogonal polynomials, we index the transformed distribution by say, $X_\lambda^{(k)}$, that is, by the parameter λ and order k of the polynomial.

Applying Theorem 2.1 in this framework, we obtain the following

Corollary 3.1 *Let Condition 3.1 be satisfied with $EZ_\lambda^{2m} < \infty$, and for $0 \leq k \leq m$ set*

$$\alpha_\lambda^{(k)} = \frac{1}{k!} EZ_\lambda^k P_\lambda^k(Z_\lambda). \tag{22}$$

Then for all $X \in \mathcal{M}_\lambda^k$, where

$$\mathcal{M}_\lambda^k = \{X : EX^j = EZ_\lambda^j, \quad 0 \leq j \leq 2k\},$$

there exists a random variable $X_\lambda^{(k)}$ such that for all $F \in \mathcal{F}^k(P_\lambda^k)$

$$EP_\lambda^k(X)F(X) = \alpha_\lambda^{(k)} EF^{(k)}(X_\lambda^{(k)}). \tag{23}$$

Proof: By Condition 3.1 and orthogonality we have for $0 \leq j \leq k \leq m$,

$$\frac{1}{k!} EX^j P_\lambda^k(X) = \frac{1}{k!} EZ_\lambda^j P_\lambda^k(Z_\lambda) = \frac{1}{k!} EP_\lambda^j(Z_\lambda) P_\lambda^k(Z_\lambda) = \alpha_\lambda^{(k)} \delta_{j,k}$$

using $X \in \mathcal{M}_\lambda^k$. Now invoke Theorem 2.1. ∎

We say the family of distributions Z_λ is closed under independent addition if for independent $Z_{\lambda_i} \sim \mathcal{L}_{\lambda_i}, i = 1, 2$ we have $Z_{\lambda_1} + Z_{\lambda_2} \sim \mathcal{L}_{\lambda_1+\lambda_2}$. There is special structure when the transformation function in Theorem 2.1 is a member of an orthogonal polynomial system corresponding to such a family. In particular, the following Theorem 3.1 generalizes (3) and (7) in showing how a sum of independent variables can be P_λ^m transformed by replacing a randomly chosen collection in the sum by variables with distributions transformed using the same orthogonal polynomial system.

For $n = 1, 2, \ldots$, consider a multi-index $\mathbf{m} = (m_1, \ldots, m_n)$, and with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ and $\mathbf{x} = (x_1, \ldots, x_n)$ let

$$m = |\mathbf{m}| = \sum_{i=1}^n m_i, \quad \lambda = \sum_{i=1}^n \lambda_i$$

and set

$$\alpha_{\boldsymbol{\lambda}}^{(\mathbf{m})} = \prod_{i=1}^n \alpha_{\lambda_i}^{(m_i)} \quad \text{and} \quad P_{\boldsymbol{\lambda}}^{\mathbf{m}}(\mathbf{x}) = \prod_{i=1}^n P_{\lambda_i}^{m_i}(x_i). \tag{24}$$

Theorem 3.1 *Let $Z_\lambda, \lambda > 0$ be a family of random variables closed under independent addition with $EZ_\lambda^{2m} < \infty$, and suppose the associated orthogonal polynomials $\{P_\lambda^k(x)\}_{0 \le k \le m}$ satisfies Condition 3.1 and, for some weights $c_{\mathbf{m}}$, the identity*

$$P_\lambda^m(w) \;\; = \sum_{\mathbf{m}:|\mathbf{m}|=m} c_{\mathbf{m}} P_{\boldsymbol{\lambda}}^{\mathbf{m}}(\mathbf{x}), \tag{25}$$

where $P_{\boldsymbol{\lambda}}^{\mathbf{m}}(\mathbf{x})$ is given in (24) and $w = x_1 + \cdots + x_n$. Then $\alpha_\lambda^{(m)}$ and $\alpha_{\boldsymbol{\lambda}}^{(\mathbf{m})}$ defined in (22) and (24) respectively, satisfy

$$\alpha_\lambda^{(m)} = \sum_{\mathbf{m}:|\mathbf{m}|=m} c_{\mathbf{m}} \alpha_{\boldsymbol{\lambda}}^{(\mathbf{m})}, \tag{26}$$

and we may consider the variable $\mathbf{I}$, independent of all other variables, with distribution

$$P(\mathbf{I} = \mathbf{m}) = c_{\mathbf{m}} \frac{\alpha_{\boldsymbol{\lambda}}^{(\mathbf{m})}}{\alpha_\lambda^{(m)}}, \quad |\mathbf{m}| = m. \tag{27}$$

Furthermore, for any positive $\lambda_1, \ldots, \lambda_n$ *and independent variables* $X_1, \cdots, X_n$ *with*

$$X_i \in \mathcal{M}^m_{\lambda_i} \quad \textit{and} \quad W = \sum_{i=1}^n X_i,$$

the variable

$$W^{(m)}_\lambda = \sum_{\mathbf{m}:|\mathbf{m}|=m} (X_i)^{(I_i)}_{\lambda_i}$$

has the $W - P^m_\lambda$ *distribution.*

Proof. Since $X_i \in \mathcal{M}^m_{\lambda_i}$, we have for $0 \le k \le 2m$, and independent $Z_{\lambda_i} \sim \mathcal{L}_{\lambda_i}$ and $Z_\lambda \sim \mathcal{L}_\lambda$, that

$$EW^k = E(\sum_{i=1}^n X_i)^k = E(\sum_{i=1}^n Z_{\lambda_i})^k = EZ^k_\lambda.$$

Hence $W \in \mathcal{M}^m_\lambda$, and the $W^{(m)}$ distribution exists by Corollary 3.1. Equality (26) follows by multiplying (25) by $W^m = (\sum_i X_i)^m$ taking expectation, and using independence and orthogonality.

By (26), for any $F \in \mathcal{C}^\infty_c$,

$$\alpha^{(m)}_\lambda EF^{(m)}(W^{(m)}_\lambda) \quad = \quad E\sum_{\mathbf{m}} c_{\mathbf{m}} \alpha^{(\mathbf{m})}_{\boldsymbol{\lambda}} F^{(m)}(W^{(m)}_\lambda). \tag{28}$$

Using independence and successively applying the identity

$$\alpha^{(m_i)}_{\lambda_i} EF^{(q)}((X_i)^{(m_i)}_{\lambda_i} + y) = EP^{m_i}_{\lambda_i}(X_i)F^{(q-m_i)}(X_i + y), \tag{29}$$

we see that the right hand side of (28) is equal to

$$E\sum_{\mathbf{m}} c_{\mathbf{m}} P^{\mathbf{m}}_{\boldsymbol{\lambda}}(\mathbf{X})F(W) = EP^m_\lambda(W)F(W), \tag{30}$$

by (25). Comparing (28) to (30) we have

$$\alpha^{(m)}_\lambda EF^{(m)}(W^{(m)}_\lambda) = EP^m_\lambda(W)F(W),$$

for all $F \in \mathcal{C}^\infty_c$, and hence $W^{(m)}_\lambda$ has the W-P^m_λ biased distribution. ∎

For the possibly infinite system of monic polynomials $\{P_\lambda^m(x)\}$ orthogonal with respect to $\mathcal{L}_\lambda$, define the generating function

$$\phi_t(x,\lambda) = \sum_{m\geq 0} P_\lambda^m(x)\frac{t^m}{m!}. \tag{31}$$

Though the constants $\alpha_\lambda^{(m)}$ can be found using $F(x) = x^m$ in (23), squaring (31) and taking expectation using orthogonality gives the alternative method

$$E[\phi_t(Z_\lambda,\lambda)]^2 = \sum_{m\geq 0} \alpha_\lambda^{(m)}\frac{t^{2m}}{m!}. \tag{32}$$

Theorem 3.2 applies in the special cases considered in Sections 4.1 through 4.4.

Theorem 3.2 *If the polynomial generating function $\phi_t(x,\lambda)$ in (31) satisfies*

$$\phi_t(w,\lambda) = \prod_{i=1}^n \phi_t(x_i,\lambda_i) \tag{33}$$

for $w = x_1 + \cdots + x_n$ and $\lambda = \lambda_1 + \cdots + \lambda_n$, then (25), and hence (27), in Theorem 3.1 are satisfied respectively by

$$c_{\mathbf{m}} = \binom{m}{\mathbf{m}} \quad \textit{and} \quad P(\mathbf{I} = \mathbf{m}) = \binom{m}{\mathbf{m}}\frac{\alpha_{\boldsymbol{\lambda}}^{(\mathbf{m})}}{\alpha_\lambda^{(m)}}, \quad |\mathbf{m}| = m.$$

Proof: Rewriting (33) ,

$$\begin{aligned}
\sum_{m\geq 0}\frac{t^m}{m!}P_\lambda^m(w) &= \prod_{i=1}^n \sum_{m_i\geq 0} P_{\lambda_i}^{m_i}(x_i)\frac{t^{m_i}}{m_i!} \\
&= \sum_{m_1,\cdots,m_n} P_{\boldsymbol{\lambda}}^{\mathbf{m}}(\mathbf{x})\frac{t^{m_1+\cdots+m_n}}{m_1!\cdots m_n!} \\
&= \sum_{m\geq 0}^{\infty}\frac{t^m}{m!}\sum_{\mathbf{m}=m}\binom{m}{\mathbf{m}}P_{\boldsymbol{\lambda}}^{\mathbf{m}}(\mathbf{x}),
\end{aligned}$$

giving (25) with the values claimed. ■

We also note that squaring (25) and taking expectation, using independence and orthogonality, results in

$$\alpha_\lambda^{(m)} = \sum_{|\mathbf{m}|=m}\binom{m}{\mathbf{m}}^{-1} c_{\mathbf{m}}^2\alpha_{\boldsymbol{\lambda}}^{\mathbf{m}}, \tag{34}$$

so that the conclusion of Theorem 3.2 can also be seen to hold by equating coefficients of (26) and (34) when $\alpha_{\boldsymbol{\lambda}}^{\mathbf{m}}$ takes on sufficiently many values.

We end this section with a result about the potential for iterated biasing.

Theorem 3.3 *Let Condition 3.1 be satisfied, and suppose that the the distributional family at Z_λ is closed under transformation with respect to $P_\lambda^k(x)$, that is, there exists $\mu(\lambda,k)$ such that*

$$(Z_\lambda)_\lambda^{(k)} = Z_{\mu(\lambda,k)}.$$

Then if $X \in \mathcal{M}_\lambda^m$ we have $X_\lambda^{(k)} \in \mathcal{M}_{\mu(k,\lambda)}^{m-k}$ for $k \le m$. In particular for non-negative j with $0 \le k+j \le m$, the distribution $(X_\lambda^{(k)})_{\mu(\lambda,k)}^{(j)}$ exists.

Proof. Let $0 \le j \le 2(m-k)$ and $F(x) = x^{k+j}/(k+j)_k$, where $(x)_k = x(x-1)\cdots(x-k+1)$. Then

$$\begin{aligned} &\alpha_\lambda^{(k)} E(X_\lambda^{(k)})^j = \alpha_\lambda^{(k)} EF^{(k)}(X_\lambda^{(k)}) = EP_\lambda^k(X)F(X) \\ = \;& EP_\lambda^k(Z_\lambda)F(Z_\lambda) = \alpha_\lambda^{(k)} EF^{(k)}((Z_\lambda)_\lambda^{(k)}) = \alpha_\lambda^{(k)} E(Z_{\mu(\lambda,k)})^j. \end{aligned}$$

Thus the first $2(m-k)$ moments of $X_\lambda^{(k)}$ match those of $Z_{\mu(\lambda,k)}$, and the existence of the distribution $(X_\lambda^{(k)})_{\mu(\lambda,k)}^{(j)}$ follows from Corollary 3.1. ∎

4 Special Orthogonal Polynomial Systems

In Sections 4.1 - 4.5 we specialize to the classic Hermite, Laguerre, Charlier, Krawtchouk and Gegenbauer orthogonal polynomial systems, corresponding to the Normal, Gamma, Poisson, Binomial and a Beta like family, respectively. All these families correspond to a collection of orthogonal polynomials satisfying Condition 3.1, and except for the last case, have a generating function which satisfies (33). The Normal and Poisson distributions are fixed points of their associated transformations. In the Gamma, Binomial and Beta-type cases the transformations map to the same family, but with a shifted parameter. For further connections between probability distributions and such polynomial system generating functions, see [2] and [3].

4.1 Hermite Polynomials

For $\sigma^2 = \lambda > 0$, define the collection of Hermite polynomials $\{H_\lambda^m(x)\}_{m\geq 0}$ through the generating function

$$e^{xt-\frac{1}{2}\lambda t^2} = \sum_{m=0}^{\infty} H_\lambda^m(x)\frac{t^m}{m!}, \tag{35}$$

or equivalently, the Rodriguez formula

$$H_\lambda^m(x) = (-\lambda)^m e^{\frac{x^2}{2\lambda}} \frac{d^m}{dx^m} e^{-\frac{x^2}{2\lambda}}. \tag{36}$$

These polynomials are orthogonal with respect to the normal distribution $\mathcal{N}(0,\lambda)$ with density $(2\pi\lambda)^{-1/2}\exp(-x^2/(2\lambda))$.

For $F \in \mathcal{C}_c^\infty$ and $Z_\lambda \sim \mathcal{N}(0,\lambda)$, applying the Rodriguez formula (36) we have

$$\begin{aligned} EH_\lambda^m(Z_\lambda)F(Z_\lambda) &= \int_{-\infty}^{\infty} (-\lambda)^m e^{\frac{x^2}{2\lambda}} \left(\frac{d^m}{dx^m} e^{-\frac{x^2}{2\lambda}}\right) F(x) \frac{e^{-\frac{x^2}{2\lambda}}}{\sqrt{\lambda 2\pi}} dx \\ &= \int_{-\infty}^{\infty} (-\lambda)^m \left(\frac{d^m}{dx^m} e^{-\frac{x^2}{2\lambda}}\right) F(x) \frac{1}{\sqrt{\lambda 2\pi}} dx \\ &= \lambda^m \int_{-\infty}^{\infty} F^{(m)}(x) \frac{e^{-\frac{x^2}{2\lambda}}}{\sqrt{\lambda 2\pi}} dx \\ &= \lambda^m EF^{(m)}(Z_\lambda). \end{aligned} \tag{37}$$

Hence,

$$(Z_\lambda)_\lambda^{(m)} = Z_\lambda,$$

that is, for each $m = 0, 1, \ldots,$ the normal $Z_\lambda \sim \mathcal{N}(0,\lambda)$ is a fixed point of the m^{th} order transformation induced by $H_\lambda^m(x)$.

From (37) we see that $\alpha_\lambda^{(m)} = \lambda^m$, which we could find alternatively using (32) and

$$E[e^{Z_\lambda t - \frac{1}{2}\lambda t^2}]^2 = e^{\lambda t^2} = \sum_{m\geq 0} \lambda^m \frac{t^{2m}}{m!}.$$

Now since the generating function (35) satisfies the conditions of Theorem 3.2, the distribution of the random index I in Theorem 3.1 is multinomial

Mult$(m, \boldsymbol{\lambda})$. For zero biasing and $m = 1$, this multinomial distribution reduces to the 'pick an index proportional to variance' as specified in (6).

Lastly, we indicate two ways in which the classical Stein equation can be generalized to the Hermite case. With $\mathcal{N}h = Eh(Z)$, the standard normal expectation of h, both the equations

$$f'(x)H_1^{m-1}(x) - H_1^m(x)f(x) = h(x) - \mathcal{N}h \tag{38}$$

and

$$f^{(m)}(x) - H_1^m(x)f(x) = h(x) - \mathcal{N}h \tag{39}$$

reduce to the usual Stein equation when $m = 1$ (see [21], [22])

$$f'(x) - xf(x) = h(x) - \mathcal{N}h, \tag{40}$$

and in particular the expectations on the left hand sides of each evaluated at a random variable W are zero for all $f \in C_c^\infty$ if and only if W is standard normal.

4.2 Laguerre Polynomials

For $\lambda > 0$, let $\{L_\lambda^m(x)\}_{m \geq 0}$ be the collection of Laguerre polynomials defined by the generating function

$$(1+t)^{-\lambda} \exp\left\{\frac{xt}{1+t}\right\} = \sum_{m=0}^{\infty} L_\lambda^m(x) \frac{t^m}{m!}, \tag{41}$$

or equivalently, the Rodriguez formula

$$L_\lambda^m(x) = (-1)^m x^{-\lambda+1} e^x \frac{d^m}{dx^m} x^{\lambda+m-1} e^{-x}, \tag{42}$$

which are orthogonal with respect to the Gamma distribution with parameter λ, having density $x^{\lambda-1}e^{-x}/\Gamma(\lambda), x > 0$.

For $F \in \mathcal{C}_c^\infty$ and Z_λ with this density, applying the Rodriguez formula (42) yields

$$EL_\lambda^m(Z_\lambda)F(Z_\lambda) = \int_0^\infty (-1)^m x^{-\lambda+1} e^x \left(\frac{d^m}{dx^m} x^{\lambda+m-1} e^{-x}\right) F(x) \frac{x^{\lambda-1}e^{-x}}{\Gamma(\lambda)} dx$$

$$
\begin{aligned}
&= \frac{(-1)^m}{\Gamma(\lambda)} \int_0^\infty \left(\frac{d^m}{dx^m} x^{\lambda+m-1} e^{-x} \right) F(x) dx \\
&= \frac{\Gamma(\lambda+m)}{\Gamma(\lambda)} \int_0^\infty F^{(m)}(x) \frac{x^{\lambda+m-1} e^{-x}}{\Gamma(\lambda+m)} dx \\
&= (\lambda)^m E F^{(m)}(Z_{\lambda+m}), \qquad (43)
\end{aligned}
$$

where $(\lambda)^m$ is the rising factorial,

$$
(\lambda)^m = \lambda(\lambda+1)\cdots(\lambda+m-1) = \frac{\Gamma(\lambda+m)}{\Gamma(\lambda)}.
$$

Hence

$$
(Z_\lambda)_\lambda^{(m)} = Z_{\lambda+m}.
$$

From (43) we see that $\alpha_\lambda^{(m)} = (\lambda)^m$, which we could find alternatively using (32) and

$$
E\left[(1+t)^{-\lambda} \exp\left(\frac{Z_\lambda t}{1+t}\right)\right]^2 = (1-t^2)^{-\lambda} = \sum_{m\geq 0} (\lambda)^m \frac{t^{2m}}{m!}.
$$

Since the generating function (41) satisfies the conditions of Theorem 3.2, the random index I in Theorem 3.1 has distribution

$$
P(I = \mathbf{m}) = \binom{m}{\mathbf{m}} \frac{\prod_{i=1}^n (\lambda_i)^{m_i}}{(\lambda)^m} = \frac{\prod_{i=1}^n \binom{\lambda_i+m_1-1}{m_i}}{\binom{\lambda+m-1}{m}},
$$

which we recognize as the multivariate hypergeometric distribution with parameters m and $\lambda_1+m_1-1, \ldots, \lambda_n+m_n-1$, see [18], p.301.

Though the Gamma is not a fixed point of the Laguerre transformations as the normal is for the Hermites, nevertheless there exist Stein equations for the Gamma paralleling (40) for the normal which can be used for studying distributional approximations for the Gamma family; for details, see [20]. In particular, we have the Stein characterization that $X \sim \Gamma(\lambda, 1)$ if and only if

$$
E(X-\lambda)f(X) = EXf'(X)
$$

for all smoooth functions f. Using that $L_\lambda^1(x) = x - \lambda$, the $X^{(1)}$ order one Laguerre transformation is characterized by

$$
E(X-\lambda)f(X) = \lambda E f'(X^{(1)})
$$

for all smooth functions f. Comparing these two equations we see that $X \sim \Gamma(\lambda, 1)$ if and only if for all smooth functions f,

$$EXf'(X) = \lambda E f'(X^{(1)});$$

in other words, $X \sim \Gamma(\lambda, 1)$ if and only if $X^{(1)}$, the first order Laguerre transformation of X, equals its size bias transformation X^s.

4.3 Charlier Polynomials

For $\lambda > 0$, let $\{C_\lambda^m(x)\}_{m \geq 0}$ be the collection of Charlier polynomials defined by the generating function

$$e^{-\lambda t}(1+t)^x = \sum_{m=0}^{\infty} C_\lambda^m(x) \frac{t^m}{m!}, \tag{44}$$

or, equivalently, with $(x)_k = x(x-1)\cdots(x-k+1)$, the falling factorial,

$$C_\lambda^m(x) = \sum_{k=0}^{m} \binom{m}{k} (x)_k (-\lambda)^{m-k}, \tag{45}$$

giving a family orthogonal with respect to the Poisson distribution $\mathcal{P}(\lambda)$ with mass function $e^{-\lambda}\lambda^k/k!, k = 0, 1, \ldots$. From (45) one can derive the Rodriguez formula

$$C_\lambda^m(x) = (-1)^m \Gamma(x+1) \lambda^{m-x} \nabla^m \left(\frac{\lambda^x}{\Gamma(x+1)} \right), \tag{46}$$

where $\nabla f(x) = f(x) - f(x-1)$, the backward difference.

Since the transformations in Theorem 2.1 defined using derivatives of test functions yield absolutely continuous distributions when $m \geq 1$, no discrete distribution will be a fixed point. However, parallel to (9), for an integer valued random variable X we can define the discrete $X - P$ biased distribution via

$$EP(X)F(X) = \alpha E \Delta^m F(X^{(m)}) \quad \text{for all } F \in \mathcal{F}_\Delta(P), \tag{47}$$

where $\Delta f(x) = f(x+1) - f(x)$, and again suppressing dependence on X,

$$\mathcal{F}_\Delta(P) = \{F : \mathbf{R} \to \mathbf{R} : E|P(X)F(X)| < \infty\}.$$

That for all $m = 0, 1 \ldots$ the Poisson $\mathcal{P}(\lambda)$-distribution is a fixed point

$$(Z_\lambda)_\lambda^{(m)} = Z_\lambda$$

of the discrete transformation (47) with P replaced by C_λ^m can be seen as follows. For $Z_\lambda \sim \mathcal{P}(\lambda)$, by the Rodriguez formula (46) and

$$\sum_{k=0}^{\infty} \nabla^m b_k \cdot a_k = (-1)^m \sum_{k=0}^{\infty} b_k \Delta^m a_k, \tag{48}$$

we have

$$\begin{aligned} EC_\lambda^m(Z_\lambda)F(Z_\lambda) &= \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} C_\lambda^m(k)F(k) \\ &= \lambda^m(-1)^m \sum_{k=0}^{\infty} e^{-\lambda}\nabla^m \left(\frac{\lambda^k}{k!}\right) \cdot F(k) \\ &= \lambda^m \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!}\Delta^m F(k) \\ &= \lambda^m E\Delta^m F(Z_\lambda). \end{aligned} \tag{49}$$

From (49) we see that $\alpha_\lambda^{(m)} = \lambda^m$, which we could find alternatively using (32) and

$$E[e^{-\lambda t}(1+t)^{Z_\lambda}]^2 = e^{\lambda t^2} = \sum_{m \geq 0} \lambda^m \frac{t^{2m}}{m!}.$$

Using the existence of the Charlier biased distributions and that (29) holds with derivative replaced by difference, it is easy to see that the argument and hence conclusion of Theorem 3.1 holds in this discrete case. Now since the generating function (44) satisfies the conditions of Theorem 3.2, the distribution of the random index I in Theorem 3.1 is multinomial $\mathrm{Mult}(m, \boldsymbol{\lambda})$, as in the normal case.

As the order one Charlier polynomial is $C_\lambda^1(x) = x - \lambda$, Stein characterizations of the form (38) or (39), with Hermite replaced by Charlier and derivatives replaced by differences, generalize the Stein equation for the Poisson distribution with parameter λ given in [7], and extensively studied for example in [4].

4.4 Krawtchouk Polynomials

With $\lambda = 1, 2, \ldots$ and $p \in (0,1)$ fixed, let $\{K_\lambda^m(x)\}_{0 \le m \le \lambda}$ be the collection of Krawtchouk polynomials defined by the generating function

$$(1+qt)^x (1-pt)^{\lambda - x} = \sum_{m=0}^{\lambda} \frac{t^m}{m!} K_\lambda^m(x) \tag{50}$$

where $p + q = 1$, giving the family of polynomials orthogonal with respect to the Binomial $\mathcal{B}(\lambda, p)$ distribution. In contrast to the previous examples, the Binomial is not infinitely divisible and has support on a bounded set.

Following the approach set out in [3], the polynomials can also be given by the Rodriguez formula

$$K_\lambda^m(x) = \frac{(-1)^m m! \binom{\lambda}{m} p^{m-x} q^x}{\binom{\lambda}{x}} \nabla^m \left\{ \binom{\lambda - m}{x} \left(\frac{p}{q}\right)^x \right\},$$

and so for $Z_\lambda \sim \mathcal{B}(\lambda, p), 0 \le m \le \lambda$ and bounded F,

$$\begin{aligned}
&EK_\lambda^m(Z_\lambda)F(Z_\lambda) \\
&= EF(Z_\lambda) \frac{(-1)^m m! \binom{\lambda}{m} p^{m-Z_\lambda} q^{Z_\lambda}}{\binom{\lambda}{Z_\lambda}} \nabla^m \left\{ \binom{\lambda - m}{Z_\lambda} \left(\frac{p}{q}\right)^{Z_\lambda} \right\} \\
&= \sum_{k=0}^{\lambda} \binom{\lambda}{k} p^k q^{\lambda - k} F(k) \frac{(-1)^m m! \binom{\lambda}{m} p^{m-k} q^k}{\binom{\lambda}{k}} \nabla^m \left\{ \binom{\lambda - m}{k} \left(\frac{p}{q}\right)^k \right\} \\
&= m! \binom{\lambda}{m} p^m q^\lambda (-1)^m \sum_{k=0}^{\lambda} F(k) \nabla^m \left\{ \binom{\lambda - m}{k} \left(\frac{p}{q}\right)^k \right\}.
\end{aligned}$$

Using (48) and letting $(\lambda)_m$ again be the falling factorial, we write the last expression as

$$\begin{aligned}
&(\lambda)_m p^m q^\lambda \sum_{k=0}^{\lambda} \binom{\lambda - m}{k} \left(\frac{p}{q}\right)^k \Delta^m F(k) \\
&= (\lambda)_m (pq)^m \sum_{k=0}^{\lambda} \binom{\lambda - m}{k} p^k q^{\lambda - m - k} \Delta^m F(k) \\
&= \alpha_\lambda^{(m)} E \Delta^m F(Z_\lambda^{(m)}),
\end{aligned}$$

yielding

$$\alpha_\lambda^{(m)} = (\lambda)_m (pq)^m \quad \text{and} \quad (Z_\lambda)_\lambda^{(m)} = Z_{\lambda-m}.$$

Hence, similar to the Gamma family, the Binomial distribution is not a fixed point of its own transformational family, but the transformed distribution is a member of the same family. One can calculate $\alpha_\lambda^{(m)}$ alternatively using (32), (50), and series expansion of

$$E(1+qt)^{2Z_\lambda}(1-pt)^{2\lambda-2Z_\lambda} = (1+pqt^2)^\lambda.$$

As for Example 4.3, the conclusion of Theorem 3.1 holds, and since the generating function (50) satisfies the conditions of Theorem 3.2 the distribution of the random index I in Theorem 3.1 is given by

$$P(\mathbf{I} = \mathbf{m}) \quad = \quad \frac{\binom{m}{\mathbf{m}}(pq)^{\sum m_i}\prod_{i=1}^n (\lambda_i)_{m_i}}{(\lambda)_m (pq)^m} = \binom{m}{\mathbf{m}} \frac{\prod_{i=1}^n (\lambda_i)_{m_i}}{(\lambda)_m} = \frac{\prod_{i=1}^n \binom{\lambda_i}{m_i}}{\binom{\lambda}{m}},$$

which we recognize as the multivariate hypergeometric distribution with parameters m and $\lambda_1, \ldots, \lambda_n$, see [18], p.301.

From [10] we have the Stein characterization that $X \sim \mathcal{B}(\lambda, p)$ if and only if

$$pE(\lambda - X)f(X+1) = qEXf(X),$$

for all functions f for which these expectations exist. Using the first Krawtchouk polynomial is $K_\lambda^1(x) = qx - p(\lambda - x)$, we obtain that the first order Krawtchouk transformation is characterized by

$$qEXf(X) - pE(\lambda - X)f(X) = \lambda pqE\Delta f(X^{(1)}).$$

Combining these equations yields that $X \sim \mathcal{B}(\lambda, p)$ if and only if

$$\begin{aligned} pE(\lambda - X)\Delta f(X) &= pE(\lambda - X)(f(X+1) - f(X)) \\ &= qEXf(X) - pE(\lambda - X)f(X) \\ &= \lambda pqE\Delta f(X^{(1)}). \end{aligned}$$

Putting $g(x) = \Delta f(\lambda - x)$ we see that $X \sim \mathcal{B}(\lambda, p)$ if and only if $\lambda - X^{(1)}$ has the $(\lambda - X)$-size biased distribution, that is, if and only if

$$\lambda - X^{(1)} \sim \mathcal{B}(\lambda - 1, q) + 1, \quad \text{which is equivalent to} \quad X^{(1)} \sim \mathcal{B}(\lambda - 1, p).$$

4.5 Gegenbauer Polynomials

In this last section, we consider a polynomial system orthogonal with respect to a continuous distribution with compact support. For $\lambda > -\frac{1}{2}$, let (see [3])

$$\alpha_\lambda^{(m)} = \frac{\Gamma(\lambda)\Gamma(2\lambda+m)\Gamma(\lambda+1)}{2^{2m}\Gamma(\lambda+m+1)\Gamma(\lambda+m)\Gamma(2\lambda)},$$

and the collection of Gegenbauer polynomials $G_\lambda^m(x)$ be defined via the Rodriguez formula

$$G_\lambda^m(x) = (-1)^m \alpha_\lambda^{(m)} \frac{\Gamma(\lambda+\frac{1}{2})\Gamma(\lambda+m+1)}{\Gamma(\lambda+1)\Gamma(\lambda+m+\frac{1}{2})}(1-x^2)^{\frac{1}{2}-\lambda}\frac{d^m}{dx^m}\{(1-x^2)^{\lambda+m-\frac{1}{2}}\}.$$

Then $G_\lambda^m(x)$ are monic, have degree m, and satisfy the orthogonality relation

$$\frac{1}{m!}EG_\lambda^k(Z_\lambda)G_\lambda^m(Z_\lambda) = \alpha_\lambda^{(m)}\delta_{k,m} \quad k=0,\ldots,m,$$

where $Z_\lambda \sim g_\lambda$ with

$$g_\lambda(x) = \frac{1}{\sqrt{\pi}}\frac{\Gamma(\lambda+1)}{\Gamma(\lambda+\frac{1}{2})}(1-x^2)^{\lambda-\frac{1}{2}}, \quad |x| \le 1.$$

In particular Corollary 3.1 obtains, proving the existence of the family of Gegenbauer transformations. This family of distribution is a special case of the centered Pearson Type I-distributions, sometimes also called Beta Type I-distributions, see [23], p.150. We note that for $\lambda = 1/2$ we obtain the uniform distribution $\mathcal{U}[-1,1]$, for $\lambda = 0$ the arcsine law, and for $\lambda = 1$ the semi-circle law [25], [26].

Considering the action of the G_λ^m transformation on $Z_\lambda \sim g_\lambda$, for $F \in \mathcal{C}_c^\infty$ we have

$$\begin{aligned}
&EG_\lambda^m(Z_\lambda)F(Z_\lambda) \\
&= \frac{1}{\sqrt{\pi}}\frac{\Gamma(\lambda+1)}{\Gamma(\lambda+\frac{1}{2})}(-1)^m\alpha_\lambda^{(m)}\frac{\Gamma(\lambda+\frac{1}{2})\Gamma(\lambda+m+1)}{\Gamma(\lambda+1)\Gamma(\lambda+m+\frac{1}{2})}\int_{-1}^{1}F(x)\frac{d^m}{dx^m}\{(1-x^2)^{\lambda+m-\frac{1}{2}}\} \\
&= \frac{1}{\sqrt{\pi}}\alpha_\lambda^{(m)}\frac{\Gamma(\lambda+m+1)}{\Gamma(\lambda+m+\frac{1}{2})}\int_{-1}^{1}F^{(m)}(x)(1-x^2)^{\lambda+m-\frac{1}{2}} \\
&= \alpha_\lambda^{(m)}EF^{(m)}(Z_{\lambda+m}),
\end{aligned}$$

yielding

$$(Z_\lambda)_\lambda^{(m)} = Z_{\lambda+m}.$$

Thus, for $\lambda = 0$ we obtain that the first order Gegenbauer transformation of the arcsine distribution is the semi-circle law.

Lastly we note that since the above Beta-type distributions are not invariant under addition, Theorem 3.1 and its construction do not apply. However, as $G_\lambda^1(x) = x$, we recognize the first order Gegenbauer transformation as the zero-bias transformation, so that for sums of independent random variables the construction given in (7) applies.

Acknowledgement. The authors would like to thank the organizers of the program 'Stein's method and applications: A program in honor of Charles Stein' and the Institute of Mathematical Sciences in Singapore for their most generous hospitality, and the excellent meeting where this work was completed. Also we would like to thank an anonymous referee for very helpful comments.

References

[1] Abramowitz and A.A. Stegun, eds. (1972) Handbook of Mathematical Functions. M Dover, New York

[2] Asai, N., Kubo, I., and Kuo, H. (2003). Multiplicative renormalization and generating functions I. *Taiwanese Journal of Mathematics* **7**, 89–101.

[3] Asai, N., Kubo, I., and Kuo, H. (2004). Multiplicative renormalization and generating functions II. To appear in *Taiwanese Journal of Mathematics.*

[4] Barbour, A.D., Holst, L., and Janson, S. (1992). *Poisson Approximation.* Oxford University Press.

[5] Bolthausen, E. (1984). An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **66**, 379–386.

[6] Cacoullos, T., Papathanasiou, V., and Utev, S. (1994). Variational inequalities with examples and an application to the central limit theorem. *Ann. Probab.* **22**, 1607–1618.

[7] Chen, L.H.Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.* **3**, 534–545.

[8] Chen, L.H.Y. and Shao, Q-M. (2003). Stein's method and normal approximation. Tutorial notes for the workshop on Stein's method and Applications. `http://www.ims.nus.edu.sg/Programs/stein/files/tut_shao`

[9] Diaconis, P., and Zabell, S. (1991). Closed form summation for classical distributions: variations on a theme of de Moivre. *Statistical Science* **6**, 284-302.

[10] Ehm, W. (1991). Binomial approximation to the Poisson binomial distribution. *Statistics and Probability Letters* **11**, 7-16.

[11] Goldstein, L. (2003). Normal approximation for hierarchical sequences. *Annals of Applied Probability*, In Press.

[12] Goldstein, L. (2004). Berry Esseen bounds for combinatorial central limit theorems and pattern occurrences, using zero and size biasing. *Preprint.*

[13] Goldstein, L., and Reinert, G. (1997) Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.* **7**, 935-952.

[14] Goldstein, L., and Reinert, G. (2004) Zero biasing in one and higher dimensions, and applications. *Preprint.*

[15] Goldstein, L., and Rinott, Y. (1996). On multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Prob.* **33**, 1–17.

[16] Ho, S.-T., and Chen, L.H.Y. (1978). An L_p bound for the remainder in a combinatorial central limit theorem. *Ann. Probab.* **6**, 231–249.

[17] Holmes. S. (1998). Stein's method for birth and death processes. *Preprint.*

[18] Johnson, N.L., and Kotz, S. (1969). *Discrete Distributions*. Wiley, New York etc.

[19] Karlin, S. and Studden, W.J. (1966). *Tchebycheff Systems, with Applications in Analysis and Statistics*. Interscience, New York.

[20] Luk, H.M. (1994). Stein's method for the gamma distribution and related statistical applications. Ph.D. thesis. University of Southern California, Los Angeles, USA.

[21] Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2**, 583–602. Univ. California Press, Berkeley.

[22] Stein, C. (1986). *Approximate Computation of Expectations*. IMS, Hayward, California.

[23] Kendall, M.G. amd Stuart, S. (1969). *The Advanced Theory of Statistics. Volume 1: Distribution Theory*. Griffin, London.

[24] Schoutens, W. (2001). Orthogonal polynomials in Stein's method. *J. Math. Anal. Appl.* **253**, 515-531.

[25] Wigner, E. P. (1955). Characteristic vectors of bordered matrices with infinite dimension. *Ann. Math.* **62** 548-564.

[26] Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. Math.* **67** 325-328.