Orthogonal polynomials in Stein's method

Document status and date:
Published: 01/01/1999

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

Download date: 16. Nov. 2023

Orthogonal Polynomials in Stein's Method

Wim Schoutens[*]

Katholieke Universiteit Leuven

Department of Mathematics

Celestijnenlaan 200 B

B-3001 Leuven

Belgium

Abstract

We discuss Stein's method for Pearson's and Ord's family of distributions. We give a systematic treatment, including the Stein equation, its solution and smoothness conditions. A key role in the analysis is played by the classical orthogonal polynomials.

AMS Subject Classification 60J30, 60H05.

Keywords STEIN'S METHOD, ORTHOGONAL POLYNOMIALS, PEARSON'S CLASS, ORD'S CLASS, APPROXIMATION, DISTRIBUTIONS, MARKOV PROCESSES

Running Head STEIN'S METHOD & ORTHOGONAL POLYNOMIALS

Mailing Address WIM SCHOUTENS, KATHOLIEKE UNIVERSITEIT LEUVEN, DEPARTMENT OF MATH-

EMATICS, CELESTIJNENLAAN 200 B, B-3001 LEUVEN, BELGIUM

E-mail Wim.Schoutens@wis.kuleuven.ac.be

1 Introduction

Stein's Method [28] provides a way of finding approximations to the distribution, ρ say, of a random variable, which at the same time gives estimates of the approximation error involved. The strenghts of the method are that it can be applied in many circumstances in which dependence plays a part. In essence the method is based on a defining equation, or equivalently an operator, of the distribution ρ and a related Stein equation. Up to now is was not clear which equation to take. One could think at a lot of equations. We show how for a broad class of distributions, there is one equation who has a special role. We give a systematic treatment, including the Stein equation, its solution and smoothness conditions.

A key tool in Stein's theory is the generator method developed by Barbour [4]. Barbour suggested employing for the operator of the Stein equation, the generator of a Markov process. In this generator method, one thus looks for a defining equation for ρ, which is related to the generator of a Markov process. For a given distribution there may be various Markov processes who fit in Barbour's method. However, up to now, it was still not clear which Markov process to take to obtain good results. We show how for a broad class of distributions there is a special Markov process, a birth and death process or a diffusion, which takes a leading role in the analysis.

Furthermore, a key role is played by the classical orthogonal polynomials. P. Diaconis and S. Zabell already mentioned this connection [11]. It turns out that the defining operator is based on a hypergeometric difference or differential equation, which lies at the heart of the classical orthogonal polynomials. Furthermore, the spectral representation of the transition probabilities of the Markov process involved will be in terms of orthogonal polynomials closely related to the distribution to be approximated. This systematic treatment together with the introduction of orthogonal polynomials in the analysis seems to be new. Furthermore some earlier uncovered examples like the Beta, the Student's t, and the Hypergeometric distribution are now worked out.

2 Prelimaries

2.1 Birth and Death Processes and Diffusions

A *birth and death process* $\{X_t, t \geq 0\}$ is a Markov process on the state space $S \subset \{0, 1, 2, \ldots\}$ with stationary transition probabilities, i.e. $P_{ij}(t) \equiv \Pr(X_{t+s} = j | X_s = i)$, $i, j \in S$, is not depending on s, and with infinitesimal generator $\mathcal{A}$ given by

$$\mathcal{A}f(i) = \lambda_i f(i+1) - (\lambda_i + \mu_i)f(i) + \mu_i f(i-1), \quad i \in S, \tag{1}$$

for all bounded real-valued functions $f \in \mathcal{B}(S)$ and where we take $\lambda_i, \mu_i > 0$ for i not on the boundary of S. On the boundary of S we must have $\lambda_i, \mu_i \geq 0$. We will always work with $S = \mathcal{N} = \{0, 1, 2, \ldots\}$, in which case we set $\mu_0 = 0$, or take $S = \{0, 1, \ldots, N\}$ with N a positive integer, and in which case $\mu_0 = \lambda_N = 0$. Furthermore we do not allow the existence of an absorbing state, i.e. a boundary state with birth parameter and death parameter equal to zero. The parameters λ_i and μ_i are called, respectively, the *birth and death rates*. It can be shown that the limits $\lim_{t \to \infty} P_{ij}(t) = p_j$, $j \in S$, exist and are independent of the initial state i. It turns out that the p_j are given by $p_j = \pi_j / \sum_{k \in S} \pi_k$, $j \in S$, where $\pi_j = \lambda_0 \lambda_1 \ldots \lambda_{j-1}/(\mu_0 \mu_1 \ldots \mu_j)$, $j \in S \setminus \{0\}$, and $\pi_0 = 1$. In order that the sequence $\{p_j\}$ defines a distribution we must have $\sum_k \pi_k < \infty$ and then clearly $\sum_k p_k = 1$. We say that $\{p_j\}$ is the *limiting stationary distribution*. If $\sum_k \pi_k = \infty$ then all p_j are zero and we do not have a limiting stationary distribution.

In the analysis of birth and death processes, a prominent role is played by a sequence of polynomials $\{Q_n(x), n \in S\}$, called *birth-death polynomials*. They are determined uniquely by the recurrence relation

$$-x Q_n(x) = \mu_n Q_{n-1}(x) - (\lambda_n + \mu_n) Q_n(x) + \lambda_n Q_{n+1}(x), \quad n \in S,$$

together with $Q_{-1}(x) = 0$ and $Q_0(x) = 1$. Karlin and McGregor [15] [23] proved that the transition

function P can be represented as

$$P_{ij}(t) = \pi_j \int_0^\infty e^{-xt} Q_i(x) Q_j(x) d\phi(x), \quad i, j \in S, t \geq 0, \tag{2}$$

where ϕ is a positive Borel measure with total mass 1 and with support on the non-negative real axis; ϕ is called the *spectral measure* of P. Taking $t = 0$ in (2) one easily sees that the polynomials $\{Q_n(x), n \in S\}$ are orthogonal with respect to ϕ. In the examples we will encounter a variety of birth and death processes. Other examples can be found in the litterature, see for example [13], [26] and [30].

Another class of Markov processes will appear also in the analysis: *Diffusions* with state space $S = (a, b)$, $-\infty \leq a < b \leq +\infty$. We refer to [7] for a general introduction. Suppose $\mathcal{A}$ is the generator of the diffusion. In [7] a clear proof is given of the fact that $\mathcal{A}$ is of the form:

$$\mathcal{A}f(x) = \mu(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x),$$

where $\mu(x)$ is called the drift coefficient and $\sigma^2(x) > 0$ the diffusion coefficient. We will highlight the spectral representation for some of diffusion processes in the examples (see also [21]).

2.2 Stein's Method

2.2.1 Normal Approximation and Poisson Approximation

In 1972, Stein [28] published a method to prove Normal approximation. It is based on the fact that a random variable Z has a Standard Normal distribution N(0, 1) if and only if for all differentiable functions f such that $E|f'(X)| < \infty$, where X has a Standard Normal distribution N(0, 1), $E[f'(Z) - Zf(Z)] = 0$.

Hence, it seems reasonable that if $E[f'(W) - Wf(W)]$ is small for a large class of functions f, then the distribution of W is close to the Standard Normal distribution. Suppose we wish to estimate the difference between the expectation of a smooth function h with respect to the random variable W and $E[h(Z)]$, where Z has a Standard Normal distribution. Stein [28] showed that for

any smooth, real-valued bounded function h there is a function $f = f_h$ solving the now called *Stein equation* for the Standard Normal

$$f'(x) - xf(x) = h(x) - E[h(Z)], \tag{3}$$

with Z a Standard Normal random variable. The unique bounded solution of the above equation is given by

$$f_h(x) = \exp(x^2/2) \int_{-\infty}^{x} (h(y) - E[h(Z)]) \exp(-y^2/2) dy.$$

Then we estimate

$$E[f'_h(W) - W f_h(W)] \tag{4}$$

and hence $E[h(W)] - E[h(Z)]$. The next step is to show that the quantity (4) is small. In order to do this we will use the structure of W. For instance, it might be that W is a sum of independent random variables. In addition we will use some smoothness conditions on f_h. Stein showed the following inequalities

$$||f_h|| \leq \sqrt{\frac{\pi}{2}} ||h - E[h(Z)]||, \tag{5}$$

$$||f'_h|| \leq \sup(h) - \inf(h),$$

$$||f''_h|| \leq 2||h'||, \tag{6}$$

where $|| \cdot ||$ denotes the supremum norm.

In this way we can bound the distance of W from the Normal, in terms of a test function h; the immediate bound of the distance is one of the key advantages of Stein's method compared to moment generating functions or characteristic functions.

Chen [8] applied Stein's idea in the context of Poisson approximation. The Stein equation for the Poisson distribution $P(\mu)$ is now a difference equation:

$$\mu f(x) - xf(x-1) = h(x) - E[h(Z)], \tag{7}$$

where h is a bounded real-valued function defined on the set of the non-negative integers and Z has a Poisson distribution $P(\mu)$. The choice of the left hand side of equation (7) is based on the fact that a random variable W on the set of the non-negative integers has a Poisson distribution with parameter μ if and only if for all bounded real-valued functions f on the integers $E[\mu f(W) - W f(W-1)] = 0$. The solution of the Stein equation (7) for the Poisson distribution $P(\mu)$ is given by:

$$f_h(x) = x! \mu^{-x-1} \sum_{k=0}^{x} (h(k) - E[h(Z)]) \frac{\mu^k}{k!}, \quad x \geq 0.$$

This solution is the unique, except at $x < 0$, bounded solution; the value $f_h(x)$ for negative x does not enter into consideration and is conventionally taken to be zero. In [5] one finds the following estimates of the smoothness for f_h by an analytic argument:

$$||f_h|| \leq ||h|| \min(1, \mu^{-1/2}) \text{ and } ||\Delta f_h|| \leq ||h|| \min(1, \mu^{-1}),$$

where $\Delta f(x) = f(x+1) - f(x)$.

2.2.2 General Approach

For an arbitrary distribution ρ, the general procedure is: Find a good characterization of the desired distribution ρ in terms of an equation, that is of the type

$$Z \text{ is a r.v. with distribution } \rho \text{ if and only if } E[Af(Z)] = 0,$$

for all smooth functions f, where A is an operator associated with the distribution ρ. (Thus in the standard normal case $Af(x) = f'(x) - xf(x)$, $x \in \mathbb{R}$.) We will call such an operator a *Stein operator*. Next assume Z to have distribution ρ, and consider the Stein equation

$$h(x) - E[h(Z)] = Af(x). \tag{8}$$

For every smooth h, find a corresponding solution f_h of this equation. For any random variable W, $E[h(W)] - E[h(Z)] = E[Af_h(W)]$. Hence, to estimate the proximity of W and Z, it is sufficient to estimate $E[Af_h(W)]$ for all possible solutions of (8).

7

However, in this procedure it is not completely clear which characterizing equation for the distribution to choose (one could think of a whole set of possible equations). The aim is to be able to solve (8) for a sufficiently large class of functions h, to obtain convergence in a known topology.

2.2.3 Barbour's Generator Method

A key tool in Stein's theory is the generator method developed by Barbour [4]. Replacing f by f' in the Stein equation for the Standard Normal (3) gives $f''(x) - xf'(x) = h(x) - E[h(Z)]$. If we set $\mathcal{A}_1 f(x) = f''(x) - xf'(x)$, this equation can be rewritten as $\mathcal{A}_1 f = h(x) - E[h(Z)]$. The key advantage is that $\mathcal{A}_1$ is also the generator of a Markov process, the Ornstein-Uhlenbeck process, with Standard Normal stationary distribution.

If we replace f by $\Delta f = f(x+1) - f(x)$ in the Stein equation for the Poisson distribution (7), we get $\mu f(x+1) - (\mu+x)f(x) + xf(x-1) = h(x) - E[h(Z)]$. If we set $\mathcal{A}_2 f(x) = \mu f(x+1) - (\mu+x)f(x) + xf(x-1)$, this equation can be rewritten as $\mathcal{A}_2 f(x) = h(x) - E[h(Z)]$. Again we see that $\mathcal{A}_2$ is a generator of a Markov process, an immigration-death process, with stationary distribution the Poisson distribution. Indeed from (1) we see that $\mathcal{A}_2$ is the generator of a birth and death process with constant birth (or immigration) rate $\lambda_i = \mu$ and linear death rate $\mu_i = i$.

This also works for a broad class of other distributions ρ. Barbour suggested employing for an operator the generator of a Markov process. So for a random variable Z with distribution ρ, we are looking for an operator $\mathcal{A}$, such that $E[\mathcal{A}f(Z)] = 0$ and for a Markov process $\{X_t, t \geq 0\}$ with generator $\mathcal{A}$ and with unique stationary distribution ρ. We will call such an operator $\mathcal{A}$ a *Stein-Markov operator* for ρ. The associated equation will be called the *Stein-Markov equation*

$$\mathcal{A}f(x) = h(x) - E[h(Z)]. \tag{9}$$

This method will be in the following called *the generator method*.

However, for a given distribution ρ, there may be various operators $\mathcal{A}$ and Markov processes with ρ as stationary distributions. We will provide a general procedure to obtain for a large class

of distributions one such process.

In this framework, for a bounded function h, the solution to the Stein-Markov equation (9) may be given by $f_h(x) = -\int_0^\infty (T_t h(x) - E[h(Z)])dt$ where Z has distribution ρ, X_t is a Markov process with generator $\mathcal{A}$ and stationary distribution ρ and $T_t h(x) = E[h(X_t)|X_0 = x]$.

2.2.4 Stein Operators and Stein-Markov Operators

We summarize some Stein-Markov operators $\mathcal{A}$ and Stein operators A for some well-known distributions in the next tables, where we set $q = 1 - p$. For more details see [2], [3], [6], [8], [24], [28] and references cited therein.

Table 1: Stein operators

Name	Notation	$\mathcal{A}f(x)$
Normal	N$(0,1)$	$f''(x) - xf'(x)$
Poisson	P(μ)	$\mu f(x+1) - (x+\mu)f(x) + xf(x-1)$
Gamma	G$(a,1)$	$xf''(x) + (a+1-x)f'(x)$
Pascal	Pa(γ, μ)	$\mu(x+\gamma)f(x+1) - (\mu(x+\gamma)+x)f(x) + xf(x-1)$
Binomial	Bin(N,p)	$p(N-x)f(x+1) - (p(N-x)+qx)f(x) + qxf(x-1)$

Table 2: Stein-Markov operators

Name	Notation	$Af(x)$
Normal	N$(0,1)$	$f'(x) - xf(x)$
Poisson	P(μ)	$\mu f(x+1) - xf(x)$
Gamma	G$(a,1)$	$xf'(x) + (a+1-x)f(x)$
Pascal	Pa(γ, μ)	$\mu(x+\gamma)f(x) - xf(x-1)$
Binomial	Bin(N,p)	$p(N-x)f(x) - qxf(x-1)$

Note that the Stein-Markov and the Stein operators are of the form

$$\mathcal{A}f(x) = s(x)f''(x) + \tau(x)f'(x) \text{ and } Af(x) = s(x)f'(x) + \tau(x)f(x)$$

in the continuous case and of the form

$$
\begin{aligned}
\mathcal{A}f(x) &= s(x)\Delta\nabla f(x) + \tau(x)\Delta f(x) \\[2mm]
&= (s(x) + \tau(x))f(x+1) - (2s(x) + \tau(x))f(x) + s(x)f(x-1) \\[2mm]
Af(x) &= s(x)\nabla f(x) + \tau(x)f(x) = (s(x) + \tau(x))f(x) - s(x)f(x-1)
\end{aligned}
$$

in the discrete case, where the $s(x)$ and $\tau(x)$ are polynomials of degree at most two and one respectively and $\nabla f(x) = f(x) - f(x-1)$. Furthermore the above distributions satisfy equations with the same ingredients $s(x)$ and $\tau(x)$. In the continuous case the density (or weight) function $\rho(x)$ of the distribution ρ satisfies the differential equation $(s(x)\rho(x))' = \tau(x)\rho(x)$ and in the discrete case the probabilities $\Pr(Z = x) = p_x$ satisfy the difference equation $\Delta(s(x)p_x) = \tau(x)p_x$.

This brings us to the Pearson class of continuous distributions and Ord's family of discrete distributions

3 Stein's Method for Pearson's and Ord's Family

In 1895 K. Pearson introduced his famous family of frequency curves. The elements of this family arise by considering the possible solutions to the differential equation

$$\rho'(x) = \frac{(x + a_0)\rho(x)}{b_0 + b_1 x + b_2 x^2} = \frac{q(x)\rho(x)}{p(x)}. \tag{10}$$

There are in essence five basic solutions, depending on whether the polynomial $p(x)$ in the denominator is constant, linear or quadratic and, in the latter case, on whether the discriminant, $D = b_1^2 - 4b_0 b_2$, of $p(x)$ is positive, negative or zero. It is easy to show that the Pearson family is closed under translation and scale change. Thus the study of the family can be reduced to differential equations that result after an affine transformation of the independent variable.

1. If $\deg(p(x)) = 0$, then $\rho(x)$ can be reduced after change of variable to a Standard Normal density.

2. If $\deg(p(x)) = 1$, then the resulting solution may be seen to be the family of Gamma distributions.

3. If $\deg(p(x)) = 2$ and $D = 0$, then the density is of the form $\rho(x) = Cx^{-\alpha}\exp(-\beta/x)$, where C is the appropriate normalizing constant.

4. If $\deg(p(x)) = 2$ and $D < 0$, then the density $\rho(x)$ can be brought into the form $\rho(x) = C(1 + x^2)^{-\alpha}\exp(\beta\arctan(x))$, where again C is the appropriate normalizing constant; in particular, the t-distributions are a rescaled subfamily of this class.

5. If $\deg(p(x)) = 2$ and $D > 0$, the density $\rho(x)$ can be brought into the form $\rho(x) = Cx^{\alpha-1}(1 - x)^{\beta-1}$, where C is the appropriate normalizing constant; the Beta densities clearly belong to this class.

In what follows we will suppose that in the continuous case we have a distribution ρ on an interval (a, b), with a and b possible infinite, with a second moment, a distribution function $F(x)$ and a density function $\rho(x)$, but we will find it more convenient to work with an equivalent form of the differential equation (10). We assume that our density function $\rho(x)$ satisfies:

$$(s(x)\rho(x))' = \tau(x)\rho(x), \tag{11}$$

for some polynomials $s(x)$ of degree at most two and $\tau(x)$ of exact degree one. The equivalence between (10) and (11) can easily be seen by setting $p(x) = s(x)$ and $q(x) = \tau(x) - s'(x)$.

Furthermore we will make the following assumptions on $s(x)$:

$$s(x) > 0, \quad a < x < b \text{ and } s(a), s(b) = 0 \text{ if } a, b \text{ is finite.} \tag{12}$$

Note that because $\rho(x) \geq 0$ and $\int_a^b \rho(y)dy = 1$, we have that $\tau(x)$ is not a constant and is a decreasing linear function. Indeed, suppose it was non-constant and increasing and denote the only

11

zero of $\tau(x)$ by l, then we would have for $x < l$,

$$\int_a^x \rho(y)dy \leq \int_a^x \frac{\tau(y)}{\tau(x)}\rho(y)dy = \frac{\int_a^x (s(y)\rho(y))'dy}{\tau(x)} = \frac{s(x)\rho(x)}{\tau(x)} < 0,$$

which is impossible. For a similar reason, $\tau(x)$ can not be constant.

The only zero of $\tau(x)$, l say, is just $E[Z]$, where Z has distribution ρ. This can be seen by calculating,

$$E[\tau(Z)] = \int_a^b \tau(y)\rho(y)dy = \int_a^b (s(y)\rho(y))'dy = s(y)\rho(y)|_a^b = 0.$$

Ord's family comprises all the discrete distributions that satisfy

$$\frac{\Delta p_x}{p_x} = \frac{p_{x+1} - p_x}{p_x} = \frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2} = \frac{q(x)}{p(x)}, \tag{13}$$

where $p_x = \mathrm{Pr}(Z = x)$ and x takes values in $S = \{a, a+1, \ldots, b-1, b\}$, with a, b possible infinite and where we set for convenience $p_x = 0$ for $x \notin S$.

So we suppose that we have a discrete distribution ρ on S with a finite second moment, but also here we prefer to work with an equivalent form of the difference equation (13). We assume that our probabilities p_x satisfy:

$$\Delta(s(x)p_x) = \tau(x)p_x, \tag{14}$$

for some polynomials $s(x)$ of degree at most two and $\tau(x)$ of exact degree one. The equivalence between (14) and (13) can easily be seen by using $\Delta(s(x)p_x) = s(x+1)\Delta p_x + p_x \Delta s(x)$, and setting $p(x) = s(x+1)$ and $q(x) = \tau(x) - \Delta s(x)$. In this way we can also rewrite the difference equation (14) as

$$\frac{p_{x+1}}{p_x} = \frac{s(x) + \tau(x)}{s(x+1)}. \tag{15}$$

Furthermore we will make the following assumptions on $s(x)$:

$$s(a) = 0 \text{ if } a \text{ is finite}, \quad s(x) > 0, \quad a < x \leq b. \tag{16}$$

Note again, that because $p_x \geq 0$ and $\sum_{i=a}^b p_i = 1$, that $\tau(x)$ is not a constant and is a decreasing linear function. and that the only zero of $\tau(x)$, l say, is just the mean of the distribution ρ. For a complete description of Ord's family, we refer to [14].

12

We start with a characterization of a distribution ρ with density $\rho(x)$ satisfying (11). We set $\mathcal{C}_1$ equal to the set of all real bounded piecewise continuous functions on the interval (a, b) and set $\mathcal{C}_2$ equal to the set of all real continuous and piecewise continuously differentiable functions f on the interval (a, b), for which the function $g(z) \equiv |s(z)f'(z)| + |\tau(z)f(z)|$ is bounded. We have the following theorem.

Theorem 1 *Suppose we have a random variable X on (a, b) with density function $\tilde{\rho}(x)$ and finite second moment and that $\rho(x)$ satisfies (11). Then $\tilde{\rho}(x) = \rho(x)$ if and only if for all functions $f \in \mathcal{C}_2$, $E[s(X)f'(X) + \tau(X)f(X)] = 0$.*

Proof: First assume X has density function $\rho(x)$. Then

$$
\begin{aligned}
E[s(X)f'(X) + \tau(X)f(X)] &= \int_a^b f'(x)(s(x)\rho(x))dx + \int_a^b f(x)\tau(x)\rho(x)dx \\
&= \left. f(x)s(x)\rho(x) \right|_a^b - \int_a^b f(x)(s(x)\rho(x))'dx + \int_a^b f(x)\tau(x)\rho(x)dx \\
&= -\int_a^b f(x)\tau(x)\rho(x)dx + \int_a^b f(x)\tau(x)\rho(x)dx = 0
\end{aligned}
$$

Conversely, suppose we have a random variable X on (a, b) with density function $\tilde{\rho}(x)$ and finite second moment such that for all functions $f \in \mathcal{C}_2$, $E[s(X)f'(X) + \tau(X)f(X)] = 0$. Then

$$
\begin{aligned}
0 &= E[s(X)f'(X) + \tau(X)f(X)] = \int_a^b (s(x)f'(x) + \tau(x)f(x))\tilde{\rho}(x)dx \\
&= \left. f(x)\tilde{\rho}(x)s(x) \right|_a^b - \int_a^b (s(x)\tilde{\rho}(x))'f(x)dx + \int_a^b \tau(x)f(x)\tilde{\rho}(x)dx \\
&= -\int_a^b (s(x)\tilde{\rho}(x))'f(x)dx + \int_a^b \tau(x)f(x)\tilde{\rho}(x)dx.
\end{aligned}
$$

But this means that for all functions $f \in \mathcal{C}_2$, $\int_a^b (s(x)\tilde{\rho}(x))'f(x)dx = \int_a^b (\tau(x)\tilde{\rho}(x))f(x)dx$. So $\tilde{\rho}(x)$ satisfies the differential equation $(s(x)\tilde{\rho}(x))' = \tau(x)\rho(x)$, which uniquely defines the density $\rho(x)$. In conclusion we have $\tilde{\rho}(x) = \rho(x)$. $\bullet$

In the discrete case, we have a similar characterization of a distribution ρ with probabilities p_x satisfying (14) and (16). We set $\mathcal{C}_3$ equal to the set of all real-valued functions f on the integers such that f is zero outside S and the function $g(x) \equiv |s(x)\nabla f(x)| + |\tau(x)f(x)|$ is bounded and

13

where $\nabla f(x) = f(x) - f(x-1)$. We have the following theorem, we omit the prove of it because it is completly along the same lines as the proof of the continuous version.

Theorem 2 *Suppose we have a discrete random variable X on the set S with probabilities $\Pr(X = x) = \tilde{p}_x$ and finite second moment and that p_x satisfies (14). Then $\tilde{p}_x = p_x$ if and only if for all functions $f \in \mathcal{C}_3, E[s(X)\nabla f(X) + \tau(X)f(X)] = 0$.*

In Stein's method we wish to estimate the difference between the expectation of a function $h \in \mathcal{C}_1$ with respect to a continuous random variable W and $E[h(Z)]$, where Z has distribution ρ. To do this, we solve first the so-called Stein equation for the distribution ρ,

$$s(x)f'(x) + \tau(x)f(x) = h(x) - E[h(Z)]. \tag{17}$$

The solution of this Stein equation is given in the next proposition.

Proposition 1 *The Stein equation (17) for the distribution ρ and a function $h \in \mathcal{C}_1$ has as solution*

$$f_h(x) \;=\; \frac{1}{s(x)\rho(x)} \int_a^x (h(y) - E[h(Z)])\rho(y)dy \tag{18}$$

$$\;=\; \frac{-1}{s(x)\rho(x)} \int_x^b (h(y) - E[h(Z)])\rho(y)dy, \tag{19}$$

when $a < x < b$ and $f_h = 0$ elsewhere. This f_h belongs to $\mathcal{C}_2$.

Proof: First note that

$$\begin{aligned}
f_h'(x) \;&=\; \frac{-(s(x)\rho(x))'}{(s(x)\rho(x))^2} \int_a^x (h(y) - E[h(Z)])\rho(y)dy + \frac{h(x) - E[h(Z)]}{s(x)} \\
&=\; \frac{-\tau(x)\rho(x)}{(s(x)\rho(x))^2} \int_a^x (h(y) - E[h(Z)])\rho(y)dy + \frac{h(x) - E[h(Z)]}{s(x)} \\
&=\; \frac{-\tau(x)}{(s(x))^2\rho(x)} \int_a^x (h(y) - E[h(Z)])\rho(y)dy + \frac{h(x) - E[h(Z)]}{s(x)}.
\end{aligned}$$

Next we just substitute the proposed solution (18) into the left hand side of the Stein equation. This gives

$$\begin{aligned}
s(x)f_h'(x) + \tau(x)f_h(x) \;&=\; \frac{-\tau(x)}{s(x)\rho(x)} \int_a^x (h(y) - E[h(Z)])\rho(y)dy + h(x) - E[h(Z)] \\
&\quad + \frac{\tau(x)}{s(x)\rho(x)} \int_a^x (h(y) - E[h(Z)])\rho(y)dy \\
&=\; h(x) - E[h(Z)].
\end{aligned}$$

14

The second expression for f_h follows from the fact

$$\int_a^x (h(y) - E[h(Z)])\rho(y)dy + \int_x^b (h(y) - E[h(Z)])\rho(y)dy = 0.$$

To prove that for $h \in \mathcal{C}_1$, we have $f_h \in \mathcal{C}_2$, we need only show that $g(x) \equiv |s(x)f_h'(x)| + |\tau(x)f_h(x)|$, $a < x < b$, is bounded. We have for $x < l$,

$$\begin{aligned}
g(x) &= |s(x)f_h'(x)| + |\tau(x)f_h(x)| \\
&\leq \left| \frac{\tau(x)}{s(x)\rho(x)} \int_a^x (h(y) - E[h(Z)])\rho(y)dy + h(x) - E[h(Z)] \right| \\
&\quad + \left| \frac{\tau(x)}{s(x)\rho(x)} \int_a^x (h(y) - E[h(Z)])\rho(y)dy \right| \\
&\leq \left| \frac{||h(x) - E[h(Z)]||}{s(x)\rho(x)} \int_a^x \tau(y)\rho(y)dy \right| + ||h(x) - E[h(Z)]|| \\
&\quad + \left| \frac{||h(y) - E[h(Z)]||}{s(x)\rho(x)} \int_a^x \tau(y)\rho(y)dy \right| \\
&\leq 3||h(x) - E[h(Z)]||,
\end{aligned}$$

where $||f(x)|| = \sup_{a < x < b} |f(x)|$. A similar result for $x \geq l$ follows from (19). This proves our proposition. $\bullet$

In the discrete case, we wish to estimate the difference between the expectation of a bounded function h with respect to a random variable W and $E[h(Z)]$, where Z has distribution ρ. To do this, we solve first the so-called Stein equation for the discrete distribution ρ,

$$s(x)\nabla f(x) + \tau(x)f(x) = h(x) - E[h(Z)]. \tag{20}$$

This Stein equation is solved in the next proposition. The proof is completely of the same structure as in the continuous case.

Proposition 2 *The Stein equation (20) for the distribution ρ and a bounded function h has as solution*

$$f_h(x) = \frac{1}{s(x+1)p_{x+1}} \sum_{i=a}^x (h(i) - E[h(Z)])p_i = \frac{-1}{s(x+1)p_{x+1}} \sum_{i=x+1}^b (h(i) - E[h(Z)])p_i, \tag{21}$$

when $a \leq x < b$ and $f_h = 0$ elsewhere. Furthermore, this f_h belongs to $\mathcal{C}_3$.

15

Again suppose first we are in the continuous setting. The next step is to estimate

$$E[s(W)f_h'(W) + \tau(W)f_h(W)] \tag{22}$$

and hence $E[h(W)] - E[h(Z)]$. To show the quantity in (22) is small, it is necessary to use the structure of W. In addition we might require certain smoothness conditions on f_h which would translate into smoothness conditions on h by the following lemma.

Set as before $l = E[Z]$ and remember that l is the only zero of τ. We will need the following positive constant

$$M = \frac{1}{\rho(l)s(l)} \max(F(l), 1 - F(l)).$$

Lemma 1 *Suppose $h \in \mathcal{C}_1$ and let f_h be the solution (18) of the Stein equation given by (17). Then*

$$||f_h(x)|| \leq M||h(x) - E[h(Z)]||.$$

Proof: Note that $\tau(x)$ is positive and decreasing in (a, l) and negative and decreasing in (l, b). So we have for $x < l$,

$$F(x) = \int_a^x \rho(y)dy \leq \int_a^x \frac{\tau(y)}{\tau(x)}\rho(y)dy = \frac{1}{\tau(x)}\int_a^x (s(y)\rho(y))'dy = \frac{s(x)\rho(x)}{\tau(x)}.$$

Similarly, for $x > l$, we have

$$1 - F(x) = \int_x^b \rho(y)dy \leq \frac{-s(x)\rho(x)}{\tau(x)}. \tag{23}$$

Now, for $x \leq l$,

$$|f_h(x)| = \left|\frac{1}{s(x)\rho(x)}\int_a^x (h(y) - E[h(Z)])\rho(y)dy\right| \leq ||h(x) - E[h(Z)]||\frac{\int_a^x \rho(y)dy}{s(x)\rho(x)}. \tag{24}$$

Similarly for $x \geq l$,

$$|f_h(x)| = \left|\frac{-1}{s(x)\rho(x)}\int_x^b (h(y) - E[h(Z)])\rho(y)dy\right| \leq ||h(x) - E[h(Z)]||\frac{\int_x^b \rho(y)dy}{s(x)\rho(x)}. \tag{25}$$

Next we prove that the expressions $\int_a^x \rho(y)dy/(s(x)\rho(x))$ and $\int_x^b \rho(y)dy/(s(x)\rho(x))$, of (24) and (25) attain there maximum at $x = l$. To show this we calculate

$$\left(\frac{\int_a^x \rho(y)dy}{s(x)\rho(x)}\right)' = \frac{-\tau(x)}{\rho(x)(s(x))^2}\int_a^x \rho(y)dy + \frac{1}{s(x)}$$

16

and

$$\left(\frac{\int_x^b \rho(y)dy}{s(x)\rho(x)}\right)' = \frac{-\tau(x)}{\rho(x)(s(x))^2}\int_x^b \rho(y)dy - \frac{1}{s(x)}.$$

Next we use (3) and (23) to obtain,

$$\left(\frac{\int_a^x \rho(y)dy}{s(x)\rho(x)}\right)' \geq 0 \text{ for } x \leq l \text{ and } \left(\frac{\int_x^b \rho(y)dy}{s(x)\rho(x)}\right)' \leq 0 \text{ for } x \geq l.$$

In conclusion we have $||f_h(x)|| \leq M||h - E[h(Z)]||$. •

In the discrete setting, we wish to estimate

$$E[s(W)\nabla f_h(W) + \tau(W)f_h(W)] \tag{26}$$

and hence $E[h(W)] - E[h(Z)]$. We again might require certain smoothness conditions on f_h which would translate into smoothness conditions on h by the following lemma. Let $\tau(x) = cx + d$, $c < 0$ and $l = -d/c$ be the only zero of τ. Now we need the following positive constant

$$M = \frac{1}{p_{\lfloor l \rfloor + 1}s(\lfloor l \rfloor + 1)}\max\left(\sum_{i=a}^{\lfloor l \rfloor} p_i, \sum_{i=\lfloor l \rfloor + 1}^{b} p_i\right).$$

Lemma 2 *Suppose h is a bounded function and let f_h be the solution of the Stein equation given by (20). Then $||f_h(x)|| \leq M||h(x) - E[h(Z)]||$, where $||f(x)|| = \sup_{a \leq x \leq b}|f(x)|$.*

The proof is complete analoguous to the continuous case.

Another inequality for distributions in Pearson's class is given in the next lemma.

Lemma 3 *Suppose $h \in C_1$ and let f_h be the solution (18) of the Stein equation given by (17). Then $||f_h'(x)|| \leq 2||1/s(x)|| \times ||h(x) - E[h(Z)]||$.*

Proof: Because

$$\begin{aligned}
f_h'(x) &= \frac{-\tau(x)}{(s(x))^2\rho(x)}\int_a^x (h(y) - E[h(Z)])\rho(y)dy + \frac{h(x) - E[h(Z)]}{s(x)} \\
&= \frac{\tau(x)}{(s(x))^2\rho(x)}\int_x^b (h(y) - E[h(Z)])\rho(y)dy + \frac{h(x) - E[h(Z)]}{s(x)},
\end{aligned}$$

we have for $x \leq l$,

$$|f_h'(x)| \leq ||h(x) - E[h(Z)]||\left(\frac{\tau(x)F(x)}{(s(x))^2\rho(x)} + \frac{1}{s(x)}\right) \leq ||h(x) - E[h(Z)]||\left(\frac{2}{s(x)}\right),$$

17

where we used for the last inequality (3).

Similarly, for $x \geq l$ we have

$$|f_h'(x)| \leq ||h(x) - E[h(Z)]|| \left(\frac{-\tau(x)(1 - F(x))}{(s(x))^2 \rho(x)} + \frac{1}{s(x)} \right) \leq ||h(x) - E[h(Z)]|| \left(\frac{2}{s(x)} \right).$$

This ends the proof. ●

A slightly different version for Ord's class is given in the next lemma.

Lemma 4 *Suppose h is a bounded function and let f_h be the solution of the Stein equation given by (20). Then*

$$||\nabla f_h(x)|| \leq \max \left(\frac{1}{|\tau(a)|}, \sup_{a < x \leq b} (2/s(x)) \right) ||h(x) - E[h(Z)]||.$$

Proof: For $x = a$ and $x = b$, we have

$$\begin{aligned}
|\nabla f_h(a)| &= \left| \frac{(h(a) - E[h(Z)])p_a}{s(a+1)p_{a+1}} \right| = \left| \frac{h(a) - E[h(Z)]}{\tau(a)} \right| \leq \frac{||h(x) - E[h(Z)]||}{|\tau(a)|} \\
|\nabla f_h(b)| &= |f_h(b-1)| = \left| \frac{h(b) - E[h(Z)]}{s(b)} \right| \leq \sup_{a < x \leq b} (2/s(x)) ||h(x) - E[h(Z)]||.
\end{aligned}$$

For $a < x < b$, the proof is completely similar as in the continuous case. ●

4   Orthogonal Polynomials and Barbour's Markov Process

After having considered the close relation between the defining difference and differential equations of the distributions involved and their Stein(-Markov) operators, we bring into the analysis some related orthogonal polynomials [9] [25]. The key link in the continuous case will be the differential equation of hypergeometric type which is satisfied by the classical orthogonal polynomials of a continuous variable: $s(x)y'' + \tau(x)y' + \lambda y = 0$, where $s(x)$ and $\tau(x)$ are polynomials of at most second and first degree, respectively, and λ is a constant.

In the discrete case, the link will be the difference equation of hypergeometric type which is satisfied by the classical orthogonal polynomials of a discrete variable: $s(x)\Delta\nabla y(x) + \tau(x)\Delta y(x) +$

$\lambda y(x) = 0$, where $s(x)$ and $\tau(x)$ are again polynomials of at most second and first degree, respectively, and λ is a constant.

Let $Q_n(x)$ be the orthogonal polynomials of degree n with respect to the distribution ρ, then the $Q_n(x)$ satisfy such an equation of hypergeometric type for some specific constants $\lambda_n \neq 0$. But this means that we have

$$\mathcal{A} Q_n(x) = -\lambda_n Q_n(x). \tag{27}$$

In this way we can formally solve the Stein-Markov equation

$$\mathcal{A} f = h(x) - E[h(Z)]. \tag{28}$$

with the aid of orthogonal polynomials. Let $F(x) = \Pr(Z \leq x)$ the involved distribution function and $d_n = \int_S Q_n(x)^2 d\rho(x) \neq 0$, with S the support of ρ. Suppose $h(x) - E[h(Z)] = \sum_{n=0}^{\infty} a_n Q_n(x)$, where we can determine the a_n by

$$a_n = \int_S Q_n(x)(h(x) - E[h(Z)])dF(x)/d_n^2, \quad n \geq 0.$$

Note that $a_0 = Q_0(x) \int_S (h(x) - E[h(Z)])dF(x) = 0$. But then for a given h the solution of (28) is given by

$$f_h(x) = \sum_{n=1}^{\infty} \frac{-a_n}{\lambda_n} Q_n(x).$$

Indeed, we have

$$\mathcal{A} f_h(x) = \mathcal{A} \sum_{n=1}^{\infty} \frac{-a_n}{\lambda_n} Q_n(x) = \sum_{n=1}^{\infty} \frac{-a_n}{\lambda_n} \mathcal{A} Q_n(x) = \sum_{n=1}^{\infty} a_n Q_n(x) = h(x) - E[h(Z)].$$

Another place where the orthogonal polynomials appear is in Barbour's operator method. Recall that we are considering some distribution ρ, continuous or discrete, together with a Stein-Markov operator $\mathcal{A}$ of a Markov process, X_t say.

In the discrete case the operator $\mathcal{A}$ has the form

$$
\begin{aligned}
\mathcal{A} f(x) &= s(x)\Delta \nabla f(x) + \tau(x)\Delta f(x) \\
&= (s(x) + \tau(x))f(x+1) - (2s(x) + \tau(x))f(x) + s(x)f(x-1),
\end{aligned}
$$

19

which is the operator of a birth and death process with birth and death rates $\kappa_n = s(n) + \tau(n)$ and $\mu_n = s(n)$, respectively, if $\kappa_n, \mu_n \geq 0$.

The orthogonal polynomials $Q_n(x)$ of ρ satisfy

$$
\begin{aligned}
\mathcal{A}Q_n(x) &= (s(x) + \tau(x))Q_n(x+1) - (2s(x) + \tau(x))Q_n(x) + s(x)Q_n(x-1) \\
&= -\lambda_n Q_n(x). \tag{29}
\end{aligned}
$$

Suppose we have a duality relation of the form $Q_n(x) = \hat{Q}_x(\lambda_n)$ and that $\hat{Q}_x$ is a polynomial of degree x. Then (29) can be written as

$$
-\lambda_n \hat{Q}_x(\lambda_n) = (s(x) + \tau(x))\hat{Q}_{x+1}(\lambda_n) - (2s(x) + \tau(x))\hat{Q}_x(\lambda_n) + s(x)\hat{Q}_{x-1}(\lambda_n).
$$

Interchanging the role of x and n we clearly see that this results in a three term recurrence equation

$$
-\lambda_x \hat{Q}_n(\lambda_x) = (s(n) + \tau(n))\hat{Q}_{n+1}(\lambda_x) - (2s(n) + \tau(n))\hat{Q}_n(\lambda_x) + s(n)\hat{Q}_{n-1}(\lambda_x).
$$

By Favard's Theorem the $\hat{Q}_n$ must be orthogonal polynomials with respect to some distribution $\tilde{\rho}$ say. Furthermore, note that these polynomials are the birth-death polynomials of the birth and death process X_t. According to the Karlin and McGregor spectral representation (2) we have

$$
P_{ij}(t) \equiv \Pr(X_t = j | X_0 = i) = \pi_j \int_0^\infty e^{-\lambda_y t} \hat{Q}_i(\lambda_y)\hat{Q}_j(\lambda_y)d\tilde{F}(y),
$$

where $\pi_0 = 1$ and $\pi_j = (\kappa_0 \kappa_1 \ldots \kappa_{j-1})/(\mu_1 \mu_2 \ldots \mu_j)$, $j \geq 1$, and $\tilde{F}(x)$ is the distribution function of $\tilde{\rho}$.

The stationary distribution is given by $r_i = \pi_i / \sum_{k=0}^\infty \pi_k$. Note that this distribution is completely defined by the fraction of successive probabilities

$$
\frac{r_{i+1}}{r_i} = \frac{\kappa_i}{\mu_{i+1}} = \frac{\sigma(i) + \tau(i)}{\sigma(i+1)}.
$$

Comparing this with (15) we see that the stationary distribution is indeed our starting distribution ρ. In the examples, we will work out this procedure for some well-known discrete distributions.

In the continuous case the operator $\mathcal{A}$ has the form $\mathcal{A}f(x) = s(x)f''(x) + \tau(x)f'(x)$ which is the operator of a diffusion with drift coefficient $\mu(x) = \tau(x)$ and diffusion coefficient $\sigma^2(x) = 2s(x) > 0$. Recall that the polynomials, $y_n(x)$, which are orthogonal with respect to ρ satisfy $\mathcal{A}y_n(x) = -\lambda_n y_n(x)$, $n \geq 0$. As in the discrete case the orthogonal polynomials involved are eigenfunctions and appear in the spectral representation as shown in the examples section.

5   Examples

1. The Standard Normal distribution $N(0,1)$ and the Ornstein-Uhlenbeck process: The Normal distribution $N(m, \sigma^2)$ with mean $m \in \mathbb{R}$ and variance $\sigma^2 > 0$, has a density function $\rho(x; m, \sigma^2) = \exp(-(x-m)^2/(2\sigma^2))/\sqrt{2\pi\sigma^2}$, $x \in \mathbb{R}$. Clearly we have $\rho'(x; m, \sigma^2)/\rho(x; m, \sigma^2) = (m-x)/x = q(x)/p(x)$, and thus $s(x) = p(x) = \sigma^2$ and $\tau(x) = q(x) + s'(x) = m - x$. So the Stein equation for the $N(m, \sigma^2)$ distribution is given by $\sigma^2 f'(x) + (m-x)f(x) = h(x) - E[h(Z)]$. The Stein operator is given by $Af(x) = \sigma^2 f'(x) + (m-x)f(x)$ and $M = \sqrt{\pi/(2\sigma^2)}$. So for the Standard Normal distribution $N(0,1)$, Lemma 1 recovers the first bound in (5). This case was the starting point of Stein's theory [28]. Suppose we have $\mu(x) = \tau(x) = -x$ and $\sigma^2(x) = 2s(x) = 2$, then we have $\mathcal{A}H_n(x/\sqrt{2}) = -nH_n(x/\sqrt{2})$, where the operator $\mathcal{A}$ is given by $\mathcal{A}f = f''(x) - xf'(x)$. This is the generator of the *the Ornstein Uhlenbeck Process*. The spectral representation for the transition density is given by [21]

$$p(t; x, y) = \frac{e^{-y^2/2}}{\sqrt{2\pi}} \sum_{n=0}^{\infty} e^{-nt} H_n(x/\sqrt{2}) H_n(y/\sqrt{2}) \frac{1}{2^n n!}.$$

where $H_n(x)$ is the *Hermite polynomial* of degree n [22]. The Hermite polynomials $H_n(x/\sqrt{2})$ are orthogonal with respect to the Standard Normal distribution we started with.

2. The Gamma distribution $G(r, \lambda^{-1})$ and the Laguerre diffusion: The Gamma distribution $G(r, \lambda^{-1})$, with $r, \lambda > 0$, has a density function $\rho(x; r, \lambda) = \lambda^r e^{-\lambda x} x^{r-1}/\Gamma(r)$, $x > 0$. Clearly we have $\rho'(x; r, \lambda)/\rho(x; r, \lambda) = (r - 1 - \lambda x)/x = q(x)/p(x)$, and thus $s(x) = p(x) = x$ and $\tau(x) = q(x) + s'(x) = r - \lambda x$. So the Stein equation for the $G(r, \lambda^{-1})$ distribution is given by

$xf'(x) + (r - \lambda x)f(x) = h(x) - E[h(Z)]$. and $\mathcal{A}f = xf''(x) + (r - \lambda x)f'(x)$, which is the generator

of the so-called Laguerre diffusion and has a spectrale representation given by

$$p(t; x, y) = \frac{\lambda^r y^{r-1} e^{-\lambda y}}{\Gamma(r)} \sum_{n=0}^{\infty} e^{-n\lambda t} L_n^{(r-1)}(\lambda x) L_n^{(r-1)}(\lambda y) \frac{\Gamma(n+1)}{\Gamma(n+r)},$$

with L_n the Laguerre polynomial of degree n [22]. It is no co-incidence that also here, the orthogonal

polynomials involved are orthogonal with respect to the Gamma distribution we started with.

3. The Beta distribution B(α, β), $\alpha, \beta > 0$, and the Jacobi diffusion: The Beta distribution B(α, β)

on $(0, 1)$, with parameters $\alpha, \beta > 0$, has a density function $\rho(x; \alpha, \beta) = x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha, \beta)$,

$0 < x < 1$. Clearly we have

$$\frac{\rho'(x; \alpha, \beta)}{\rho(x; \alpha, \beta)} = \frac{(-\alpha - \beta + 2)x + \alpha - 1}{(1-x)x} = \frac{q(x)}{p(x)},$$

and thus $s(x) = p(x) = (1-x)x$ and $\tau(x) = q(x) + s'(x) = -(\alpha + \beta)x + \alpha$. So the Stein equation

for the B(α, β) distribution is given by

$$x(1-x)f'(x) + (\alpha - (\alpha + \beta)x)f(x) = h(x) - E[h(Z)].$$

This Stein equation seems to be new. Suppose we have $\mu(x) = \tau(x) = \frac{1}{2}(\alpha - (\alpha + \beta)x)$ and

$\sigma^2(x) = (1-x)x$, where $0 < x < 1$ and $\alpha, \beta > 0$. Then we encounter the generator of the Jacobi

diffusion: $\mathcal{A}f = \frac{1}{2}(1-x)xf''(x) + \frac{1}{2}(\alpha - (\alpha + \beta)x)f'(x)$. In this case the spectral expansion is in

terms of the *Jacobi polynomials* $P_n(x) \equiv P_n^{(\beta-1, \alpha-1)}(x)$ [22]:

$$p(t; x, y) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)} \sum_{n=0}^{\infty} e^{-n(n+\alpha+\beta-1)t/2} P_n(2x - 1)P_n(2y - 1)\pi_n \tag{30}$$

where

$$\pi_n = \frac{B(\alpha, \beta)(2n + \alpha + \beta - 1)n!\Gamma(n + \alpha + \beta - 1)}{\Gamma(n + \alpha)\Gamma(n + \beta)}.$$

and $B(\alpha, \beta)$ is the *Beta function*.

4. The Student's t-distribution t_n with $n \in \{1, 2, \ldots\}$ degrees of freedom has a density function

$$\rho(x; n) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R}.$$

22

Clearly we have

$$\frac{\rho'(x;n)}{\rho(x;n)} = \frac{-(n+1)x/n}{1+x^2/n} = \frac{q(x)}{p(x)},$$

and thus clearly $s(x) = p(x) = 1 + (x^2/n)$ and $\tau(x) = q(x) + s'(x) = -((n-1)/n)x$. This means that the Stein equation for the t_n distribution is given by

$$\left(1 + \frac{x^2}{n}\right) f'(x) - \frac{n-1}{n} x f(x) = h(x) - E[h(Z)].$$

This Stein equation seems to be new. Note that Lemma 3 gives us a useful bound on f_h', namely

$$||f_h'|| \le 2||h - E[h(Z)]||.$$

5. The Poisson distribution $P(\mu)$, $\mu > 0$, is given by the probabilities $p_x = e^{-\mu}\mu^x/x!$, $x \in \{0, 1, 2, \ldots\}$. An easy calculation gives $s(x) = x$ and $\tau(x) = \mu - x$. So the Stein operator for the Poisson distribution $P(\mu)$ is given by $Af(x) = x\nabla f(x) + (\mu - x)f(x) = \mu f(x) - xf(x-1)$, which is the same as in (7). This case was studied by [8] and many others [3], [4], [5], [6]. The Poisson distribution $P(\mu)$, $\mu > 0$, has a Stein-Markov operator, $\mathcal{A}$, given by $\mathcal{A}f(x) = \mu f(x+1) - (x+\mu)f(x) + xf(x-1)$. This is the operator of a birth and death process on $\{0, 1, 2 \ldots\}$ with birth and death rates $\kappa_n = \mu$ and $\mu_n = n$, $n \ge 0$ respectively. This birth and death process is the immigration-death process with a constant immigration rate μ and unit per capita death rate [1]. The birth-death polynomials, $Q_n(x)$, for this process are recursively defined by the relations

$$-xQ_n(x) = \mu Q_{n+1}(x) - (\mu + n)Q_n(x) + nQ_{n-1}(x), \tag{31}$$

together with $Q_0(x) = 1$ and $Q_{-1}(x) = 0$. The polynomials which are orthogonal with respect to the Poisson distribution $P(\mu)$ are the Charlier polynomials $C_n(x; \mu)$, which satisfy the following equation of hypergeometric type

$$-nC_n(x; \mu) = \mu C_n(x+1; \mu) - (\mu + n)C_n(x; \mu) + nC_n(x-1; \mu)$$

and are self-dual, i.e. $C_n(x; \mu) = C_x(n; \mu)$. Using this duality relation we obtain the three term recurrence relation of the Charlier polynomials

$$-nC_x(n; \mu) = \mu C_{x+1}(n; \mu) - (\mu + n)C_x(n; \mu) + nC_{x-1}(n; \mu).$$

23

But this is after interchanging the role of x and n exactly of the same form as (31), so we conclude that $Q_n(x) = C_n(x; \mu)$. In this way, using Karlin and McGregor's spectral representation (2), we can express the transition probabilities of our process X_t as

$$P_{ij}(t) = \Pr(X_t = j | X_0 = i) = \frac{\mu^j}{j!} \sum_{x=0}^{\infty} e^{-xt} C_i(x; \mu) C_j(x; \mu) e^{-\mu} \frac{\mu^x}{x!}.$$

6. The Binomial distribution $\mathrm{Bin}(N, p)$ on $\{0, 1, 2, \ldots, N\}$ with parameter $0 < p < 1$ is given by the probabilities $p_x = \binom{N}{x} p^x q^{N-x}$, $x \in \{0, 1, 2, \ldots, N\}$, where $q = 1 - p$. Here $s(x) = qx$ and $\tau(x) = pN - x$. So the Stein operator for the $\mathrm{Bin}(N, p)$ distribution is given by $Af(x) = qx \nabla f(x) + (pN - x)f(x) = p(N - x)f(x) - qxf(x - 1)$. The Binomial distribution $\mathrm{Bin}(N, p)$, $0 < p < 1$, has a Stein-Markov operator, $\mathcal{A}$, given by

$$\mathcal{A}f(x) = p(N - x)f(x + 1) - (p(N - x) + qx)f(x) + qxf(x - 1),$$

where $q = 1 - p$. This is the operator of a birth and death process on $\{0, 1, 2 \ldots, N\}$ with birth and death rates $\kappa_n = p(N - n)$ and $\mu_n = qn$, $0 \le n \le N$ respectively, also called the Ehrenfest Model [19]. In this case the birth-death polynomials, $Q_n(x)$, are recursively defined by

$$-xQ_n(x) = p(N - n)Q_{n+1}(x) - (p(N - n) + qn)Q_n(x) + qnQ_{n-1}(x), \tag{32}$$

together with $Q_0(x) = 1$ and $Q_{-1}(x) = 0$.

The polynomials which are orthogonal with respect to the Binomial distribution $\mathrm{Bin}(N, p)$ are the Krawtchouk polynomials $K_n(x; N, p)$, $n = 0, 1, \ldots, N$. They satisfy the following equation of hypergeometric type

$$-nK_n(x; N, p) = p(N - n)K_n(x + 1; N, p) - (p(N - n) + qn)K_n(x; N, p) + qnK_n(x - 1; N, p)$$

and are self-dual, i.e. $K_n(x; N, p) = K_x(n; N, p)$. This duality relation leads to the three term recurrence relation of the Krawtchouk polynomials

$$-nK_x(n; \gamma, \mu) = p(N - n)K_{x+1}(n; N, p) - (p(N - n) + qn)K_x(n; N, p) + qnK_{x-1}(n; N, p).$$

After interchanging the role of x and n, we conclude that $Q_n(x) = K_n(x; N, p)$. In this way, using Karlin and McGregor's spectral representation (2), we can express the transition probabilities of our process X_t as

$$P_{ij}(t) = \binom{N}{j} p^j q^{-j} \sum_{x=0}^{N} e^{-xt} K_i(x; N, p) K_j(x; N, p) \binom{N}{x} p^x q^{N-x}.$$

7. The Pascal distribution $\text{Pa}(\gamma, \mu)$ with parameters $\gamma > 0$ and $0 < \mu < 1$ is given by $p_x = \binom{x+\gamma-1}{x} \mu^\gamma (1-\mu)^x$, $x \in \{0, 1, 2, \ldots\}$.

So $s(x) = x$ and $\tau(x) = (1-\mu)\gamma - \mu x$. The Stein operator for the $\text{Pa}(\gamma, \mu)$ distribution is thus given by $Af(x) = x \nabla f(x) + ((1-\mu)\gamma - \mu x)f(x) = (1-\mu)(\gamma + x)f(x) - xf(x-1)$. The Pascal distribution $\text{Pa}(\gamma, \mu)$ has a Stein-Markov operator, $\mathcal{A}$, given by

$$\mathcal{A}f(x) = \mu(x+\gamma)f(x+1) - (\mu(x+\gamma) + x)f(x) + xf(x-1),$$

which is the operator of the above described linear birth and death process on $\{0, 1, 2 \ldots\}$ with birth and death rates $\kappa_n = \mu(n+\gamma)$ and $\mu_n = n$, $n \geq 0$ respectively [17]. The birth-death polynomials involved are defined by

$$-xQ_n(x) = \mu(n+\gamma)Q_{n+1}(x) - (\mu(n+\gamma) + n)Q_n(x) + nQ_{n-1}(x), \tag{33}$$

together with $Q_0(x) = 1$ and $Q_{-1}(x) = 0$.

The Meixner polynomials $M_n(x; \gamma, \mu)$, $n = 0, 1, \ldots$, are orthogonal with respect to the Pascal distribution $\text{Pa}(\gamma, \mu)$; they satisfy the following equation of hypergeometric type

$$-nM_n(x; \gamma, \mu) = \mu(n+\gamma)M_n(x+1; \gamma, \mu) - (\mu(n+\gamma) + n)M_n(x; \gamma, \mu) + nM_n(x-1; \gamma, \mu)$$

and are self-dual, i.e. $M_n(x; \gamma, \mu) = M_x(n; \gamma, \mu)$. Using this duality relation we obtain the three term recurrence relation of the Meixner polynomials

$$-nM_x(n; \gamma, \mu) = \mu(n+\gamma)M_{x+1}(n; \gamma, \mu) - (\mu(n+\gamma) + n)M_x(n; \gamma, \mu) + nM_{x-1}(n; \gamma, \mu).$$

Interchanging the role of x and n, we conclude that $Q_n(x) = M_n(x; \gamma, \mu)$. With the spectral representation (2), we express the transition probabilities of X_t as

$$P_{ij}(t) = \frac{\mu^j(\gamma)_j}{j!} \sum_{x=0}^{\infty} e^{-xt} M_i(x; \gamma, \mu) M_j(x; \gamma, \mu)(1 - \mu)^{\gamma} \mu^x(\gamma)_x / x!.$$

8. The Hypergeometric distribution $\text{Hyp}(\alpha, \beta, N)$, with parameters $\alpha \geq N$, $\beta > N$ and N a non-negative integer, is given by $p_x = \binom{\alpha}{x}\binom{\beta}{N-x}/\binom{\alpha+\beta}{N}$, $x \in \{0, 1, 2, \ldots, N\}$. An easy calculation gives $s(x) = x(\beta - N + x)$ and $\tau(x) = \alpha N - (\alpha + \beta)x$. So the Stein operator for the $\text{Hyp}(\alpha, \beta, N)$ distribution is given by

$$
\begin{aligned}
Af(x) &= x(\beta - N + x)\nabla f(x) + (\alpha N - (\alpha + \beta)x)f(x) \\
&= (N - x)(\alpha - x)f(x) - x(\beta - N + x)f(x - 1).
\end{aligned}
$$

This Stein equation seems to be new. The Hypergeometric distribution $\text{Hyp}(\alpha, \beta, N)$, has a Stein-Markov operator, $\mathcal{A}$, given by

$$\mathcal{A}f(x) = (\alpha - x)f(x + 1) - ((N - x)(\alpha - x) + x(\beta - N + x))f(x) + x(\beta - N + x)f(x - 1).$$

This is the operator of a birth and death process on $\{0, 1, 2 \ldots, N\}$ with quadratic birth and death rates

$$\kappa_n = (N - n)(\alpha - n) \text{ and } \mu_n = n(\beta - N + n), \quad 0 \leq n \leq N$$

respectively, whih was studied in [30]. The birth-death polynomials involved are recursively defined by the relations

$$
\begin{aligned}
-xQ_n(x) &= (N - n)(\alpha - n)Q_{n+1}(x) \\
&\quad - ((N - n)(\alpha - n) + n(\beta - N + n))Q_n(x) + n(\beta - N + n)Q_{n-1}(x), \tag{34}
\end{aligned}
$$

together with $Q_0(x) = 1$ and $Q_{-1}(x) = 0$. The *Hahn polynomials* $Q_n(x; -\alpha - 1, -\beta - 1, N)$ [22] are orthogonal with respect to Hypergeometric distribution and satisfy the following equation of

hypergeometric type

$$n(n - \alpha - \beta - 1)Q_n(x; -\alpha - 1, -\beta - 1, N) =$$

$$(N - x)(\alpha - x)Q_n(x + 1; -\alpha - 1, \beta - 1, N)$$

$$-((N - x)(\alpha - x) + x(x + \beta - N))Q_n(x; -\alpha - 1, -\beta - 1, N)$$

$$+x(x + \beta - N)Q_n(x - 1; -\alpha - 1, -\beta - 1, N).$$

Furthermore, we have the duality relation

$$Q_n(x; -\alpha - 1, -\beta - 1, N) = R_x(\lambda_n; -\alpha - 1, -\beta - 1, N),$$

where the R_x are the Dual Hahn polynomials [22] and $\lambda_n = n(-\alpha - \beta - 1)$. In what follows we will

often write for notational convenience $R_x(\lambda_n)$ instead of $R_x(\lambda_n; -\alpha - 1, -\beta - 1, N)$.

Using this duality relation we obtain the three term recurrence relation of the Dual Hahn

polynomials:

$$\lambda_n R_x(\lambda_n) = (N - x)(\alpha - x)R_{x+1}(\lambda_n)$$

$$-((N - x)(\alpha - x) + x(x + \beta - N))R_x(\lambda_n) + x(x + \beta - N)R_{x-1}(\lambda_n).$$

But this is after interchanging the role of x and n of the same form as (34), so we conclude that

$Q_n(x) = R_n(-x; -\alpha - 1, -\beta - 1, N)$. Finally, using Karlin and McGregor's spectral representation

(2), we can express the transition probabilities of our process X_t as

$$P_{ij}(t) = \frac{\binom{\alpha}{j}\binom{\beta}{N-j}}{\binom{\beta}{N}} \sum_{x=0}^{N} e^{\lambda_x t} R_i(\lambda_x) R_j(\lambda_x) \frac{(N!)(-N)_x(-\alpha)_x(2x - \alpha - \beta - 1)\binom{N-\alpha-1}{N}}{(-1)^x(x!)(-\beta)_x(x - \alpha - \beta - 1)_{N+1}}.$$

Acknowledgements

References

[1] W.J. Anderson, "Continuous-Time Markov Chains - An Applications-Oriented Approach", Springer-Verlag, New-York, 1991.

[2] A.D. Barbour, Equilibrium distributions for Markov population processes. *Adv. Appl. Prob.* **12** (1980), 591–614.

[3] A.D. Barbour, Stein's method and Poisson process convergence. In *A celebration of Applied Probability (Editor: J. Gani), J. Appl. Prob.* **25A** (1988), 175–184.

[4] A.D. Barbour, Stein's method for diffusion approximations. *Probab. Theory Related Fields* **84** (1990), 297–322.

[5] A.D. Barbour and G.K. Eagleson, Poisson approximation for some statistics based on exchangeable trials. *Adv. Appl. Prob.* **15** (1983), 585–600.

[6] A.D. Barbour, L. Holst and S. Janson, "Poisson Approximation", Clarendon, Oxford, 1992.

[7] R.N. Bhattacharya and E.C. Waymire, "Stochastic Processes with Applications", John Wiley & Sons, New York, 1990.

[8] L.H.Y. Chen, Poisson approximation for dependent trials. *Ann. Probab.* **22** (1975), 1607–1618.

[9] T.S. Chihara, "An Introduction to Orthogonal Polynomials", Gordon and Breach, New York-London-Paris, 1978.

[10] H. Dette, On a generalization of the Ehrenfest urn model. *J. Appl. Prob.* **31** (1994), 930–939.

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

[12] G.K. Eagleson, A duality relation for discrete orthogonal systems. *Studia Scientiarum Mathematicarum Hungarica* **3** (1968), 127–136.

[13] M.E.H. Ismail, J. Letessier and G. Valent, Linear birth and death models and associated Laguerre and Meixner polynomials. *Journal of Approximation Theory* **55** (1988), 337–348.

[14] N.L. Johnson, S. Kotz, and A.W. Kemp, "Univariate Discrete Distributions", 2nd edition. Wiley, New York, 1992.

[15] S. Karlin and J.L. McGregor, The differential equations of birth-and-death processes, and the Stieltjes moment problem. *Trans. Amer. Math. Soc.* **85** (1957), 489–546.

[16] S. Karlin and J.L. McGregor, The classification of birth and death processes. *Trans. Amer. Math. Soc.* **86** (1957), 366–400.

[17] S. Karlin and J.L. McGregor, Linear growth, birth and death processes. *J. Math. Mech.* **7** (1958), 643–662.

[18] S. Karlin and J.L. McGregor, Random walks, *Illinois J. Math.* **8** (1959), 87–118.

[19] S. Karlin and J.L. McGregor, Ehrenfest Urn Models. *J. Appl. Prob.* **19** (1965), 477-487.

[20] S. Karlin and H.M. Taylor, "A First Course in Stochastic Processes", 2nd edition. Academic Press, New York, 1975.

[21] S. Karlin, and H.M. Taylor, "A Second Course in Stochastic Processes", Academic Press, New York, 1981.

[22] R. Koekoek and R.F. Swarttouw, "The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue", Report 98-17, Delft University of Technology, 1998.

[23] W. Ledermann, and G.E.H. Reuter, Spectral theory for the differential equations of simple birth and death processes. *Philos. Trans. Roy. Soc. London, Ser. A* **246** (1954), 321–369.

[24] H.M. Luk, "Stein's Method for the Gamma Distribution and Related Statistical Applications", Ph.D.Thesis. University of Southern California, Los Angeles, USA, 1994.

[25] A.F. Nikiforov, S.K. Suslov, and V.B. Uvarov, "Classical Orthogonal Polynomials of a Discrete Variable", Springer-Verlag, Berlin-Heidelberg, 1991.

[26] P.R. Parthasarathy, R.B. Lenin, W. Schoutens and W. Van Assche, A birth and death process related to the Rogers-Ramanujan Continued Fraction. *J. Math. Anal. Appl.* **224** (1998), 297–315.

[27] W. Schoutens, "Stochastic Processes in the Askey Scheme", Doctoral Thesis. K.U.Leuven, Leuven, 1999.

[28] C. Stein, 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** (1972), 583–602. Univ. California Press, Berkeley.

[29] C. Stein, "Approximate Computation of Expectations", IMS Lecture Notes - Monograph Series, Volume 7. IMS, Hayward, California, 1986.

[30] W. Van Assche, P.R. Parthasarathy and R.B. Lenin, Spectral representation of certain finite birth and death processes. *The Mathematical Scientist* **24** (2) (1999), *to appear.*