Rayner, J. C. W.; Thas, O.; De Boeck , B. "A generalised Emerson recurrence relation"
Australian and New Zealand Journal of Statistics Vol. 50, Issue 3, p. 235-240 (2008)

Available from: http://dx.doi.org/10.1111/j.1467-842X.2008.00514.x

A GENERALISED EMERSON RECURRENCE RELATION

J. C. W. RAYNER[1]*, O. THAS[2] AND B. De BOECK[2]
University of Newcastle and Ghent University

Summary

The Emerson (Biometrics, vol. 24, pp. 695-701, 1968) recurrence relation has many important applications in statistics. However the original derivation applied only to discrete distributions. In the following, a simple derivation is given that generalises the Emerson recurrence relation to any distribution for which the necessary expectations exist. A modern application is outlined.

Key words: Hermite polynomials; orthonormal polynomials; spherical Legendre polynomials.

1. Introduction

The Emerson (1968) recurrence relation provides a useful result that enables a set of orthonormal polynomials to be built up from two initial polynomials. However the Emerson derivation was for discrete distributions only. Modern applications require an efficient method of calculation of orthonormal polynomials for arbitrary distributions. Such a method is provided here, and is demonstrated by deriving the Hermite and spherical Legendre polynomials. A modern application, to smooth tests of goodness of fit, is outlined in the final section. In modern data driven application of these tests, test statistics of moderate order (up to order 10, say) are required; these in turn require orthonormal polynomials of similar order. See http://biomath.ugent.be/~othas/recurrence for R (2008) code implementing the results here.

Suppose $\{h_r(x)\}$ is the set of *orthonormal* polynomials with respect to a real random variable X in the sense that $E\{h_r(X)\, h_s(X)\} = 0$ for $r \neq s$ and $E\{h_r^2(X)\} = 1$ with $r, s = 0, 1, 2, \ldots$. In the goodness of fit application discussed in Section 4, this distribution is specified by the null hypothesis. This distribution may depend on nuisance parameters but these are suppressed in the subsequent notation. Here E denotes expectation with regard to this distribution. In the following, we will always assume that all the required expectations exist.

In general we write $E(X) = \mu$ and $\mu_r = E\{(X - \mu)^r\}$ for $r = 2, 3, \ldots$. We shall assume $\{h_r^*(x)\}$ is a set of *orthogonal* polynomials with $h_0^*(x) = 1$ for all x. It will be convenient to write $c_r = E[\{h_r^*(X)\}^2]$ for $r = 0, 1, 2, \ldots$. Thus if $h_r(x) = h_r^*(x)/\sqrt{c_r}$ for all x and $r = 0, 1, 2, \ldots$, then $\{h_r(x)\}$ is the corresponding set of *orthonormal* polynomials. The initial polynomials are $h_0(x) = h_0^*(x) = 1$ for all x, $h_1^*(x) = (x - \mu)$ and $h_1(x) = (x - \mu)/\sqrt{\mu_2}$. Clearly $c_0 = 1$ and $c_1 = \mu_2$.

The Emerson (1968) recurrence relation uses the fact that the $h_r(x)$ are polynomials. The resulting process is more efficient than the more general Gram-Schmidt method in the sense that

*Author to whom correspondence should be addressed.
[1] School of Mathematical and Physical Sciences, University of Newcastle, NSW 2308, Australia.
e-mail: John.Rayner@newcastle.edu.au
[2] Department of Applied Mathematics, Biometrics and Process Control, Ghent University, B-9000 Gent, Belgium.
Acknowledgements. The authors gratefully acknowledge the financial support from the IAP research network P6/03 of the Belgian Government (Belgian Science Policy). We also wish to thank Jennifer Gordon, Ian Hansel and Paul Rippon, all of whom read the manuscript and made helpful suggestions. The insightful comments of the reviewers are also acknowledged.

fewer numerical operations are required to generate the orthonormal polynomials. In expectation form, the Emerson (1968) recurrence relation is

$$h_r^*(x) = [x - \mathrm{E}\{ X h_{r-1}^2(X)\}]\, h_{r-1}(x) - \mathrm{E}\{X h_{r-2}(X) h_{r-1}(X)\}\, h_{r-2}(x). \tag{1}$$

This equation does *not* assume that the mean μ is zero. With, for example, $r = 2$, (1) generates $h_2^*(x) = \{(x-\mu)^2 - (\mu_3/\mu_2)(x-\mu) - \mu_2\}/\sqrt{\mu_2}$. Direct calculation gives $c_2 = (\mu_4 - \mu_3^2/\mu_2 - \mu_2^2)/\mu_2$, but direct calculation becomes increasingly more complex. Instead c_r can be conveniently calculated by recurrence. See Theorem 2 below.

In the setting here the two initial polynomials are normalised and generate through (1) a polynomial that is not normalised. This is then normalised and fed into (1) to produce the next polynomial and so on. The Emerson (1968) derivation is in terms of an arbitrary discrete distribution, but the method given here can be applied to any distribution provided the necessary expectations exist. Hence the new equation can fairly be called a generalised Emerson recurrence relation.

2. A generalised Emerson recurrence relation

We now derive a recurrence relation for an underpinning distribution for which all expectations involved exist.

We subsequently assume that, for all $r \geq 2$,

$$h_r^*(x) = a_{r,r} x^r + a_{r-1,r} x^{r-1} + a_{r-2,r} x^{r-2} + \ldots + a_{1,r} x + a_{0,r}.$$

The $a_{i,r}$ are defined for $i = 0, 1, \ldots, r$, for $r = 2, 3, \ldots$. It is convenient to set the boundary condition $a_{-1,j} = 0$ for $j = 0, 1, \ldots, r$.

If all orthonormal polynomials up to the $(r-1)$th are known, then Theorem 1 enables calculation of the rth orthogonal polynomial.

Theorem 1. *If $h_0(x) = 1$ for all x and $h_1(x) = (x-\mu)/\sqrt{\mu_2}$, then for $r = 2, 3, \ldots$, $h_r^*(x)$, defined by (1), is orthogonal to $h_0(x), h_1(x), \ldots, h_{r-1}(x)$.*

Proof. Define, for $r = 2, 3, \ldots$

$$h_r^*(x) = x h_{r-1}(x) + b_{r-1,r} h_{r-1}(x) + b_{r-2,r} h_{r-2}(x) + \ldots + b_{0,r} h_0(x), \tag{2}$$

in which $b_{i,j}$ are defined for integers i and j with $i = 0, 1, \ldots, j-1$, for $j = 2, 3, \ldots$. The $x h_{r-1}(x)$ term on the right hand side of (2) ensures $h_r(x)$ is of degree r, while the $b_{i,r}$, $i = 0, 1, \ldots, r-1$ are to be chosen so that $h_r^*(x)$ is orthogonal to $h_0(x), h_1(x), \ldots, h_{r-1}(x)$. Orthogonality gives

$$0 = \mathrm{E}\{h_i(X)\, h_r^*(X)\} = \mathrm{E}\{X\, h_i(X)\, h_{r-1}(X)\} + b_{i,r} \text{ for } i = 0, 1, \ldots, r-1,$$

using the normality of the $h_i(x)$, $i = 0, 1, \ldots, r-1$. For $i = 0, 1, \ldots, r-3$, $x h_i(x)$ is a polynomial of degree at most $r-2$, and hence expressible as a linear combination of $h_0(x), h_1(x), \ldots, h_{r-2}(x)$. This will necessarily be orthogonal to $h_{r-1}(x)$. It follows that, for $r \geq 2$, we have $b_{i,r} = 0$ for $i = 0, 1, \ldots, r-3$, and that $b_{r-2,r}$ and $b_{r-1,r}$ are given by

$$b_{r-2,r} = -\mathrm{E}\{X h_{r-1}(X) h_{r-2}(X)\} \text{ and } b_{r-1,r} = -\mathrm{E}\{X h_{r-1}^2(X)\}.$$

This gives (1). The orthogonality is inherent in the construction. □

Corollary. *For $r = 2, 3, \ldots$ and $i = 0, 1, \ldots, r$,*

$$a_{i,r} = \frac{a_{i-1,r-1}}{\sqrt{c_{r-1}}} + \frac{b_{r-1,r}a_{i,r-1}}{\sqrt{c_{r-1}}} + \frac{b_{r-2,r}a_{i,r-2}}{\sqrt{c_{r-2}}}.$$

with boundary conditions $a_{0,0} = 1$, $c_0 = 1$, $a_{1,1} = 1/\sqrt{\mu_2}$, $a_{0,1} = -\mu/\sqrt{\mu_2}$, $c_1 = 1$.

Proof. This follows by equating coefficients in

$$h_r^*(x) = \sum_{i=0}^{r} a_{i,r}x^i = xh_{r-1}(x) + b_{r-1,r}h_{r-1}(x) + b_{r-2,r}h_{r-2}(x).$$ □

Theorem 2. *Assume that the polynomials* $h_0^*(x)$, $h_1^*(x)$, … , $h_{r-1}^*(x)$ *and the constants* c_0, c_1, … , c_{r-1} *are known, and that the necessary moments required in the following exist. Write* $\mu_r' = E(X^r)$ *for* $r = 1, 2, \ldots$ *. For* $r = 2, 3, \ldots$ *, the quantities required in* (1)*, and* c_r*, can be obtained from*

$$E\{Xh_{r-1}^2(X)\} = -b_{r-1,r} = \frac{\sum_{j=0}^{r-1}\sum_{k=0}^{r-1} a_{j,r-1}a_{k,r-1}\mu'_{j+k+1}}{c_{r-1}},$$

$$E\{Xh_{r-1}(X)h_{r-2}(X)\} = -b_{r-2,r} = \frac{\sum_{j=0}^{r-1}\sum_{k=0}^{r-2} a_{j,r-1}a_{k,r-2}\mu'_{j+k+1}}{\sqrt{c_{r-1}c_{r-2}}} \quad \textit{and}$$

$$c_r = \frac{\sum_{j=0}^{r-1}\sum_{k=0}^{r-1} a_{j,r-1}a_{k,r-1}\mu'_{j+k+2}}{c_{r-1}} - b_{r-1,r}^2 - b_{r-2,r}^2 .$$

Proof. Since $h_{r-1}(x) = \sum_{i=0}^{r-1} a_{i,r-1}x^i / \sqrt{c_{r-1}}$, we have

$$h_{r-1}^2(x) = \frac{\sum_{j=0}^{r-1}\sum_{k=0}^{r-1} a_{j,r-1}a_{k,r-1}x^{j+k}}{c_{r-1}}.$$

Using this equation $\mathrm{E}\{Xh_{r-1}^2(X)\}$ may be found, and $\mathrm{E}\{Xh_{r-1}(X)h_{r-2}(X)\}$ may be derived similarly. Thus, knowing the $a_{i,j}$s and c_js up to degree $r - 1$, we can find $\mathrm{E}\{Xh_{r-2}(X)h_{r-1}(X)\}$ and $\mathrm{E}\{Xh_{r-1}^2(X)\}$, and then $b_{r-2,r}$ and $b_{r-1,r}$, and thereby $h_r^*(x) = xh_{r-1}(x) + b_{r-1,r}h_{r-1}(x) + b_{r-2,r}h_{r-2}(x)$. With $h_r^*(x)$ now specified, the normalising constant can be found by squaring and taking expectations in (2) giving, as required,

$$c_r = \mathrm{E}[\{h_r^*(X)\}^2] = \mathrm{E}\{X^2h_{r-1}^2(X)\} - b_{r-1,r}^2 - b_{r-2,r}^2 .$$

In this expression $\mathrm{E}\{X^2h_{r-1}^2(X)\}$ is calculated in the same manner as the other expectations. □

The algorithm that is implemented at the web site given in Section 1 is essentially given by the corollary to Theorem 1 and Theorem 2. Since $h_0(x)$ and $h_1(x)$ are given by the boundary conditions, Theorem 2 gives the $b_{i,j}$s that are required to calculate $h_2^*(x)$, and also c_2, that is required to normalise it. Now $h_1(x)$ and $h_2(x)$ are known, and can be used in the same way to calculate $h_3(x)$, and so on.

In applying the recurrence procedure with expectations approximated numerically, the error first introduced by approximating is propagated and augmented by subsequent approximations. In the examples we have assessed, such as the Hermite polynomials considered below, the first ten polynomials are correct to ten decimal places. However, by the 15th polynomial, accuracy has declined markedly.

3. Examples

We now show that the generalised Emerson equation (1) generates the Hermite and Spherical Legendre polynomials. Of course, both of these are based on continuous, rather than discrete, weight functions.

The Hermite Polynomials

For the standard normal distribution, the polynomials with leading coefficient unity are denoted by $He_r(x)$ in Abramowitz & Stegun (1970, Chapter 22). By Abramowitz & Stegun (1970, 22.2.15), $\mathrm{E}\{He_r^2(X)\} = r!$, and from Abramowitz & Stegun (1970, 22.7.14),

$$He_r(x) = x\, He_{r-1}(x) - (r-1)\, He_{r-2}(x).$$

Multiplying throughout by $He_r(x)$ and taking expectations gives $\mathrm{E}\{X\, He_r(X)\, He_{r-1}(X)\} = r!$ for all r. The first four orthogonal polynomials generated by this system are

$$He_0(x) = 1 \text{ for all } x,\ He_1(x) = x,\ He_2(x) = (x^2 - 1) \text{ and } He_3(x) = (x^3 - 3x).$$

We now show these $He_r(x)$ satisfy (1). Put $h_r(x) = He_r(x)/\sqrt{r!}$ and note that $He_{r-1}^2(x)$ is an even function, so $\mathrm{E}\{XHe_{r-1}^2(X)\}$ is the expectation of an odd function using an even probability density function, and so is zero. Since from above $\mathrm{E}\{X\, He_{r-1}(X)\, He_{r-2}(X)\} = (r-1)!$, it follows that

$$\mathrm{E}\{X\, h_{r-1}(X)\, h_{r-2}(X)\} = \sqrt{(r-1)}.$$

Now the right hand side of (1) reduces to

$$(x-0)\,\frac{He_{r-1}(x)}{\sqrt{(r-1)!}} - \sqrt{r-1}\,\frac{He_{r-2}(x)}{\sqrt{(r-2)!}} = \frac{\{x\,He_{r-1}(x) - (r-1)\,He_{r-2}(x)\}}{\sqrt{(r-1)!}} = \frac{He_r(x)}{\sqrt{(r-1)!}},$$

using $He_r(x) = x\, He_{r-1}(x) - (r-1)\, He_{r-2}(x)$. Thus the Hermite polynomials are the solution of (1) corresponding to the standard normal distribution.

The Spherical Legendre Polynomials

For the uniform distribution over (–1, 1), Abramowitz & Stegun (1970, Chapter 22) give detail for the spherical Legendre polynomials $\{P_n(z)\}$; particularly note Abramowitz & Stegun (1970, 22.2.10, 22.3.8 and 22.7.10). These polynomials are orthogonal but not normalised. A recurrence relation that generates them is

$$P_r(z)\, r = z\, P_{r-1}(z)\,(2r-1) - P_{r-2}(z)\,(r-1).$$

The first four spherical Legendre polynomials are given by

$$P_0(z) = 1 \text{ for all } z,\ P_1(z) = z,\ 2\,P_2(z) = (3z^2 - 1) \text{ and } 2\,P_3(z) = (5z^3 - 3z).$$

We now show that the $P_r(z)$ are the solution to (1). If the defining recurrence relation above is multiplied throughout by $P_{r-2}(z)$ and expectations taken, we find

$$\mathrm{E}\{Z\, P_{r-1}(Z)\, P_{r-2}(Z)\} = \frac{r-1}{(2r-1)(2r-3)}.$$

Also, by the same argument as in the Hermite example, E{ $ZP_{r-1}^2(Z)$} = 0. Moreover adjusting Abramowitz & Stegun (1970, 22.2.10) to account for their different weight function gives E{ $P_r^2(Z)$} = 1/(2r + 1). Substituting shows that $P_r(z)$ is a solution of (1) if and only if

$$r\, a_{r,\, r} = (2r - 1)\, a_{r-1,\, r-1}.$$

For the spherical Legendre polynomials this follows from Abramowitz & Stegun (1970, 22.3.8), which gives

$$a_{r,\, r} = \binom{2r}{r} 2^{-r}.$$

A quick way to generate the normalised Legendre polynomials {$\pi_r(x)$} with weight function the uniform (0, 1) distribution is to generate the spherical Legendre polynomials, normalise by using E{ $P_r^2(Z)$} = 1/(2r + 1) and then replace z by 2x – 1. The first four normalised Legendre polynomials are:

$$\pi_0(x) = 1,\ \pi_1(x) = (2x - 1)\sqrt{3},\ \pi_2(x) = (6x^2 - 6x + 1)\sqrt{5} \text{ and } \pi_3(x) = (20x^3 - 30x^2 + 12x - 1)\sqrt{7}.$$

4. Smooth tests of goodness of fit

Suppose we wish to test if a random sample $X_1, \dots, X_n$ comes from a distribution with probability (density) function $f(x; \beta)$, where β is a $q \times 1$ vector of nuisance parameters, such as the rate when testing for an exponential distribution. First define a *smooth alternative of order k* by

$$g_k(x;\, \theta, \beta) = C(\theta, \beta) \exp\left\{ \sum_{i=q+1}^{q+k} \theta_i h_i(x; \beta) \right\} f(x; \beta).$$

Here {$h_i(x; \beta)$} is a set of orthonormal polynomials on $f(x; \beta)$, $\theta_1, \dots, \theta_k$ are real-valued parameters, $\theta = (\theta_1, \dots, \theta_k)^T$, and $C(\theta; \beta)$ is a normalising constant (assumed to exist) that ensures that $g_k(x; \beta)$ sums (integrates) to one. One approach is to use the score test of H: $\theta = 0$ against K: $\theta \neq 0$. If the densities $f(x; \beta)$ as β ranges over a well-defined parameter space form an exponential family, using Rayner & Best (1989, Theorem 6.1.1) the test statistic is

$$V_{q+1}^2 + \dots + V_{q+k}^2 \text{ in which } V_r = \frac{\sum_{j=1}^{n} h_r\left(X_j; \hat{\beta}\right)}{\sqrt{n}}.$$

Here $\hat{\beta}$ is the maximum likelihood estimator of β, with estimating equations $V_1 = V_2 = \dots = V_q = 0$.

For densities not in an exponential family the score test statistic is more complicated, but nevertheless involves the orthonormal polynomials {$h_i(x; \beta)$}. For discrete distributions such as the zero inflated Poisson (see Thas & Rayner, 2005), the orthonormal polynomials can be generated using the Emerson (1968) recurrence relation. For standard distributions, they may be found in sources such as Abramowitz & Stegun (1970). However for distributions such as the extreme-value and the Laplace, these avenues are not available. See, for example, Best, Rayner & Thas (2007).

To be explicit, the usual extreme-value distribution has two parameters: a location parameter α and a dispersion parameter β. If we let X have the extreme-value distribution, then the standardized extreme-value variable $T = (X - \alpha)/\beta$ has probability density function

$$f_T(t) = \exp\{-t - \exp(-t)\},\ -\infty < t < \infty.$$

For this standardised distribution, the orthonormal polynomials to order four are

$$h_0(t) = 1,\ h_1(t) = (\sqrt{6})(t - \gamma)/\pi,\ h_2(t)\sqrt{8.392} = (6/\pi^2)\{(t - \gamma)^2 - 1.462(t - \gamma) - \pi^2/6\},$$
$$h_3(t)\sqrt{20} = (6\sqrt{6}/\pi^3)\{(t - \gamma)^3 - 4.662(t - \gamma)^2 - 2.069(t - \gamma) + 5.265\}, \text{ and}$$
$$h_4(t)\sqrt{219.72} = (36/\pi^4)\{(t - \gamma)^4 - 9.693(t - \gamma)^3 + 10.792(t - \gamma)^2 + 31.160(t - \gamma) - 9.060\},$$

where γ is Euler's constant, approximately 0.57722.

The data driven smooth tests of goodness of fit advocated by, for example, Ledwina (1994), Kallenberg & Ledwina (1997) and Claeskens & Hjort (2004), all require orthonormal polynomials up to order at least 10. The results of this paper make their methodology far more practical than previously.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1970). *Handbook of Statistical Functions*. New York: Dover.

BEST, D.J., RAYNER, J.C.W. & THAS, O. (2007). Comparison of five tests of fit for the extreme value distribution. *J. Stat. Theory Pract.* **1**, 89–99.

CLAESKENS, G. & HJORT, N. (2004). Goodness of fit via non-parametric likelihood ratios. *Scand. J. Statist.* **31**, 487–513.

EMERSON, P.L. (1968). Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics* **24**, 695–701.

KALLENBERG, W. & LEDWINA, T. (1997). Data-driven smooth tests when the hypothesis is composite. *J. Amer. Statist. Assoc.* **92**, 1094–1104.

LEDWINA, T. (1994). Data-driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* **89**, 1000–1005.

R DEVELOPMENT CORE TEAM. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

RAYNER, J.C.W. & BEST, D.J. (1989). *Smooth Tests of Goodness of Fit*. New York: Oxford University Press.

THAS, O. and RAYNER, J.C.W. (2005). Smooth tests for the zero inflated Poisson distribution. *Biometrics* **61**, 808-815.