ACTUARIAL RESEARCH CLEARING HOUSE
1992 VOL. 1

Hierarchical Bayesian Whittaker Graduation

Stuart Klugman, Drake University

1. Introduction

The Whittaker method of graduation has been known and used for a long time and has remained popular due to its possession of a number of ideal properties. They include being non-parametric and having an easy to understand foundation. The latter means that it makes sense and thus the user of the method has a good idea of what it can and cannot do. As well, there is a statistical derivation available that uses Bayesian notions. A problem with the derivation is that it is more intuitive than precise and as such does not provide a useful frame of reference for the graduator. Regardless of the point of view, the graduation cannot be completed until the smoothing parameter is selected and this has always relied on the judgement of the analyst.

In this paper, two tasks will be undertaken. The first is to replace the ad-hoc Bayesian derivation of the method with a formal Bayesian specification. The second is to show that with this specification it is possible to complete the graduation without making an arbitrary selection of the smoothing parameter. The ideas will be illustrated with an example.

2. The Formal Bayesian Derivation

The model to be used here is a special case of the hierarchical normal linear model introduced by Lindley and Smith (1972). An excellent analysis of this model can be found in Berger (1985). This model is also described in great detail in Klugman (1991) which also contains the numerical algorithms used in this paper. A specific application to Whittaker graduation (though it is applied to a completely different problem in that paper) appears in Gersh and Kitigawa (1988).

The model begins with a description of the observed values. Let them be the $n \times 1$ vector $\boldsymbol{x}$ which in most mortality studies contains the crude mortality rates at successive ages. As in the Bayesian graduation method proposed by Kimeldorf and Jones (1967) we assume that $\boldsymbol{x}$ has a multivariate normal distribution:

$$\boldsymbol{x} \sim N(\boldsymbol{\theta}, \sigma^2 V)$$

where $\boldsymbol{\theta}$ is a vector that contains the true values and $\sigma^2 V$ is the covariance matrix that represents the sampling variation in the observed values. In most applications V is a diagonal matrix. We assume that it is known and given the binomial nature of the observations, a reasonable value for the ith diagonal element is $x_i(1-x_i)/e_i$ where e_i is a measure of the exposure associated with x_i. It is possible that the relationship between exposure and sample size is known and therefore the value of σ^2 is also known. This will be known as Case 1 and without loss of generality, σ^2 will be taken to be 1 in this case. It is common in actuarial studies to base exposure on something other than lives (for example, amounts) and in those cases the variance may be at best proportional to V and the constant may not be available from the data. (A discussion of these issues can be found in Klugman (1981).) Case 2 will be used to identify the situation where σ^2 is unknown.

Our prior knowledge about $\boldsymbol{\theta}$ is also expressed as a multivariate normal distribution:

$$\boldsymbol{\theta} \sim N(\boldsymbol{\mu}, \tau^2 Q)$$

where $\boldsymbol{\mu}$ is a prior best guess as to the mortality rates (likely to be a previous study of a similar population) and Q is a known covariance matrix. Here it is almost certain that τ^2 will be unknown (although, if known, the model becomes that used by Kimeldorf and Jones). To correspond to Whittaker graduation, the matrix Q must have a particular form. Begin by considering an arbitrary matrix K with n columns. We have

$$K\boldsymbol{\theta} \sim N(K\boldsymbol{\mu}, \tau^2 KQK').$$

It may be the case that while we have difficulty expressing a prior opinion about $\boldsymbol{\theta}$, we are able to express one about $K\boldsymbol{\theta}$. Suppose K is the matrix that computes zth differences. Our prior opinion may be that the population under study has the same third differences as our reference population but with a modest amount of fluctuation. Also, we may believe that these fluctuations are independent and have a common variance, so that KQK' should be the identity matrix.

For example, if $z = 1$, our prior opinion is that (locally) $\boldsymbol{\theta}$ is equal to $\boldsymbol{\mu}$ plus a constant. This allows for a change in level, while retaining the inherent smoothness from $\boldsymbol{\mu}$. With $z = 2$ the same interpretation would hold except that differences in the two vectors could now be a linear function of age. We will see that the more traditional form of the Whittaker formula obtains when $\boldsymbol{\mu} = \mathbf{0}$. Here the interpretation is that $\boldsymbol{\theta}$ itself should (locally) be a $z-1$ degree polynomial.

In order to continue, we must verify that Q can be obtained. For Whittaker graduation K has $n-z$ rows and rank $n-z$. For the normal distribution to apply, Q must be positive definite and must be factorable as LL' with L being lower triangular with positive diagonal elements. Our task reduces to finding L such that $(KL)(KL)' = I$. Regardless of the values of n and z there will be at least one solution for Q. For the time being it is not necessary to specify a particular solution.

To implement the analysis there are three key distributions. The first is identical to the one obtained from the Kimeldorf-Jones model:

$$\boldsymbol{\theta} \mid \boldsymbol{x}, \sigma^2, \tau^2 \sim N(\boldsymbol{\theta}^*, V^*)$$

$$\boldsymbol{\theta}^* = (W + \frac{\sigma^2}{\tau^2}R)^{-1}(W\boldsymbol{x} + \frac{\sigma^2}{\tau^2}R\boldsymbol{\mu}) \quad \text{and} \quad V^* = \sigma^2(W + \frac{\sigma^2}{\tau^2}R)^{-1}$$

$$\text{where} \quad W = V^{-1} \quad \text{and} \quad R = Q^{-1}.$$

With σ^2 and τ^2 known, $\boldsymbol{\theta}^*$ is the posterior mean and is thus the Bayes estimate. If we set $R = KK'$ and $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\theta}^*$ becomes the standard Whittaker solution. However, R as defined here is singlular and so cannot be the inverse of Q. So, strictly speaking, the standard Whittaker solution is not a special case of the model being used here. In the next section we will see how close we can come to obtaining the standard solution.

If either σ^2 or τ^2 are unknown, we need two more distributions. Removing the dependence on $\boldsymbol{\theta}$ produces:

$$\boldsymbol{x} \mid \sigma^2, \tau^2 \sim N(\boldsymbol{\mu}, \sigma^2 V + \tau^2 Q).$$

This leads to the posterior distribution:

$$f(\sigma^2, \tau^2 \mid \boldsymbol{x}) \propto |\sigma^2 V + \tau^2 Q|^{-1/2} exp[-(\boldsymbol{x} - \boldsymbol{\mu})'(\sigma^2 V + \tau^2 Q)^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/2] f(\sigma^2, \tau^2)$$

where $f(\sigma^2, \tau^2)$ is the prior distribution on the parameters σ^2 and τ^2. This distribution must be subjectively supplied by the analyst. Guidelines will be given in a later section. Should σ^2 be known, it would not be included in the prior distribution and its known value would be substituted in the above density.

The posterior distribution of $\boldsymbol{\theta}$ and its moments must then be obtained by integration:

$$f(\boldsymbol{\theta} \mid \boldsymbol{x}) = \int \int f(\boldsymbol{\theta} \mid \boldsymbol{x}, \sigma^2, \tau^2) f(\sigma^2, \tau^2 \mid \boldsymbol{x}) d\sigma^2 d\tau^2$$

$$E(\theta_i \mid \boldsymbol{x}) = \int \int \theta_i^* f(\sigma^2, \tau^2 \mid \boldsymbol{x}) d\sigma^2 d\tau^2$$

$$Var(\theta_i \mid \boldsymbol{x}) = E[Var(\theta_i \mid \boldsymbol{x}, \sigma^2, \tau^2)] + Var[E(\theta_i \mid \boldsymbol{x}, \sigma^2, \tau^2)]$$

$$= \int \int V_{ii}^* f(\sigma^2, \tau^2 \mid \boldsymbol{x}) d\sigma^2 d\tau^2 + \int \int (\theta_i^*)^2 f(\sigma^2, \tau^2 \mid \boldsymbol{x}) d\sigma^2 d\tau^2 - [E(\theta_i \mid \boldsymbol{x})]^2.$$

These integrals must be done by numerical means.

3. Practical considerations

To proceed we need a prior distribution. A general form that provides some computational efficiency is the inverse gamma distribution. That is:

$$f(\sigma^2, \tau^2) \propto (\sigma^2)^{-p} e^{-r/\sigma^2} (\tau^2)^{-q} e^{-s/\tau^2}.$$

The computational efficiency comes when the following change of variable is made. Let $\alpha = \sigma^2$ and $\lambda = \sigma^2/\tau^2$. At this point the two cases must be separated. The priors become:

$$f_1(\lambda) \propto \lambda^{q-2} e^{-s\lambda}$$

$$f_2(\alpha, \lambda) \propto \alpha^{-p-q+1} \lambda^{q-2} e^{-(r+s\lambda)/\alpha}.$$

The required integrals for the two cases are

1: $\int_0^\infty g(\lambda) |V + Q/\lambda|^{-1/2} exp[-(\boldsymbol{x} - \boldsymbol{\mu})'(V + Q/\lambda)^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/2] \lambda^{q-2} e^{-s\lambda} d\lambda$

2: $\int_0^\infty \int_0^\infty \alpha^\delta g(\lambda) |\alpha V + \alpha Q/\lambda|^{-1/2}$

$$\times exp[-(\boldsymbol{x} - \boldsymbol{\mu})'(\alpha V + \alpha Q/\lambda)^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/2] \alpha^{-p-q+1} \lambda^{q-2} e^{-(r+s\lambda)/\alpha} d\alpha d\lambda$$

$$= \int_0^\infty \int_0^\infty g(\lambda) |V + Q/\lambda|^{-1/2}$$

$$\times exp\{-\alpha^{-1}[(\boldsymbol{x} - \boldsymbol{\mu})'(V + Q/\lambda)^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/2 + r + s\lambda]\} \alpha^{-p-q-n/2+\delta+1} \lambda^{q-2} d\alpha d\lambda$$

$$\propto \int_0^\infty g(\lambda) |V + Q/\lambda|^{-1/2} \lambda^{q-2} [(\boldsymbol{x} - \boldsymbol{\mu})'(V + Q/\lambda)^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/2 + r + s\lambda]^{-p-q-n/2+\delta+2} d\lambda.$$

The value of δ is either 0 or 1, the value 1 being used when the integrand is V_{ii}^* or for obtaining the posterior mean of α.

In order to complete the calculations it is necessary that the integrals exist. This will always be the case when $r > 0$ and $s > 0$. However, if $s = 0$ then $q < 1 - k$ is required for the kth moment of λ to exist and if $r = 0$ then $p < 1$ for the moments to exist. The case with $r = s = 0$ is the most useful as this provides an improper prior. While there has been a great deal of discussion concerning the appropriate way to specify an improper prior for it to reflect prior ignorance, there has been little agreement. For variances, the values $p = q = 1$ have been recommended, but these will not produce legitimate posterior densities. For the example, with case 2, I have selected $p = 0$ and $q = -2$. These are the largest integer values that will produce a posterior variance for λ. Also, the marginal prior densities for both α and λ are constant, which is not an unreasonable choice. With the parameters set this way, the case 2 integral becomes:

$$\int_0^\infty g(\lambda)|V + Q/\lambda|^{-1/2}\lambda^{-4}[(\boldsymbol{x} - \boldsymbol{\mu})'(V + Q/\lambda)^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/2]^{-n/2+\delta+2}d\lambda.$$

The calculations in the examples were done using adaptive Gaussian integration as discussed in Klugman (1991). Because the integrals all run from zero to infinity, the integration was done over successive finite segments until the contribution of the most recent segment was negligible.

The remaining practical matter is to specify the matrix Q. As indicated in the previous section, there is more than one version which will satisfy the requirement concerning the taking of differences. For this paper the matrix L is obtained as follows:

1. In column 1 place the constants 1, 1, ...
2. In column 2 place the linear sequence 0, 1, 2, ..., $n - 1$.
3. In column 3 place the quadratic sequence 0, 0, 1, 3, 6, ..., $(n-1)(n-2)/2$.
4. Continue until z columns have been so placed. The general approach is to begin column j with $j - 1$ zeros and then a 1. The rest of the column is filled out so that jth differences are 0.
5. For columns $z + 1$ through n, repeat the previous column but shift the entries down one row.

The following example illustrates the relevant matrices for $n = 6$ and $z = 2$. The standard K matrix for Whittaker graduation is

$$K = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

and the L matrix as described above is

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 1 & 3 & 2 & 1 & 0 & 0 \\ 1 & 4 & 3 & 2 & 1 & 0 \\ 1 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}.$$

Then

$$Q = LL' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 3 & 6 & 9 & 12 & 15 \\ 1 & 4 & 9 & 15 & 21 & 27 \\ 1 & 5 & 12 & 21 & 31 & 41 \\ 1 & 6 & 15 & 27 & 41 & 56 \end{bmatrix}.$$

And

$$R = Q^{-1} = \begin{bmatrix} 3 & -3 & 1 & 0 & 0 & 0 \\ -3 & 6 & -4 & 1 & 0 & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & 0 & 1 & -4 & 5 & -2 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}.$$

In the orginal form of Whittaker graduation, the matrix R is found as $K'K$. The matrix shown above differs only slightly in the first two rows. Gersh and Kitigawa (1988) provide another solution for the matrix Q which differs only slightly from the one presented here.

4. An Example

In order to verify that the computations involved are feasible, a large example was selected. The data are from the 1975-80 Basic Tables (Society of Actuaries, 1985). I have selected the male ultimate values for graduation. Deaths and exposures (in dollars) were available for ages 15 through 100. The table was graduated by the Whittaker method with $\lambda = 18$ and W as the identity matrix and smoothness based on second differences. Because the values at ages 85-100 were deemed

Table 1

Graduation of the 1975-80 Basic Male Ultimate Table

Age	Crude	V	Basic	Bayes	Age	Crude	V	Basic	Bayes
15	0.51	1.1358	0.78	0.32	58	9.66	1.0452	9.74	9.68
16	0.98	1.5771	0.94	0.59	59	10.36	1.1843	10.75	10.65
17	1.31	1.8994	1.09	0.89	60	11.72	1.4275	11.89	11.83
18	1.09	1.4704	1.22	1.13	61	13.31	1.6871	13.16	13.19
19	1.70	2.1556	1.31	1.29	62	14.93	2.0083	14.54	14.65
20	1.18	1.3942	1.37	1.35	63	16.21	2.3249	16.02	16.16
21	1.26	1.2993	1.40	1.39	64	17.92	2.7656	17.62	17.76
22	1.65	1.6663	1.41	1.42	65	19.23	3.4928	19.38	19.49
23	1.32	1.3487	1.40	1.41	66	21.41	4.2860	21.32	21.41
24	1.34	1.3963	1.38	1.38	67	23.51	5.1109	23.49	23.56
25	1.42	1.4910	1.34	1.34	68	26.19	6.0792	25.92	25.96
26	1.44	1.5292	1.29	1.27	69	28.09	7.0014	28.66	28.63
27	0.98	1.0237	1.24	1.19	70	31.51	8.4798	31.77	31.64
28	1.12	1.1196	1.20	1.16	71	36.19	10.508	35.26	34.97
29	1.27	1.1760	1.17	1.17	72	38.05	12.120	39.15	38.63
30	1.27	0.9725	1.14	1.15	73	41.66	14.562	43.51	42.67
31	1.04	0.6963	1.12	1.10	74	45.13	17.334	48.32	47.15
32	1.08	0.6030	1.11	1.08	75	49.87	21.492	53.50	52.05
33	1.07	0.4654	1.12	1.08	76	76.90	52.039	58.76	57.32
34	1.09	0.3850	1.14	1.13	77	60.06	33.205	63.63	62.85
35	1.22	0.3547	1.17	1.21	78	65.84	41.585	68.62	68.73
36	1.34	0.3381	1.22	1.27	79	72.14	51.989	74.07	75.00
37	1.26	0.2889	1.28	1.30	80	77.44	62.853	80.16	81.67
38	1.35	0.2784	1.36	1.35	81	89.10	82.216	86.94	88.71
39	1.42	0.2669	1.45	1.44	82	99.09	105.75	94.34	96.09
40	1.60	0.2728	1.56	1.57	83	100.41	128.46	102.38	103.73
41	1.72	0.2693	1.70	1.71	84	111.14	171.96	111.38	111.61
42	1.84	0.2730	1.87	1.85	85	115.27	226.89	121.52	119.67
43	2.02	0.2876	2.07	2.04	86	135.31	326.09	132.98	127.85
44	2.30	0.3141	2.31	2.29	87	145.23	445.96	145.59	136.10
45	2.59	0.3405	2.58	2.56	88	157.93	625.72	159.30	144.35
46	2.80	0.3498	2.89	2.88	89	166.47	854.42	174.06	152.57
47	3.33	0.3933	3.24	3.26	90	193.18	1245.4	189.73	160.72
48	3.63	0.4065	3.61	3.65	91	196.02	1769.8	205.75	168.77
49	4.13	0.4418	4.02	4.06	92	211.32	2618.4	221.75	176.72
50	4.37	0.4519	4.45	4.46	93	220.65	3648.8	236.82	184.56
51	5.00	0.5018	4.92	4.94	94	239.24	5558.0	249.48	192.29
52	5.42	0.5357	5.44	5.44	95	300.13	9633.7	257.33	199.93
53	5.91	0.5826	6.00	6.01	96	337.22	24935	257.44	207.48
54	6.90	0.6782	6.61	6.66	97	205.80	35660	249.21	214.98
55	7.13	0.7035	7.28	7.30	98	313.63	88534	236.50	222.42
56	8.02	0.8062	8.01	8.03	99	69.12	39278	220.76	229.85
57	9.06	0.9469	8.83	8.83	100	256.25	329585	207.72	237.28

The crude, basic, and Bayes rates have been multiplied by 1000. The entries in the V column are proportional to the variances.

insufficiently smooth, they were empirically adjusted to force second differences to be constant. The data, along with the original (prior to the adjustment) and the Bayesian graduation appear in Table 1.

Also included are the diagonal elements of V. These were obtained as indicated earlier but to simplify the presentation as well as make the matrix calculations stay within reason, they were multiplied by 830,076,080,000 to make the average of their reciprocals equal to 1. This also puts them on the same scale as used in the official graduation.

The first step was do to four integrals in order to obtain some posterior moments. Using $g(\lambda) = 1$ and $\delta = 0$ provides the proportionality constant needed to produce a density. It turned out to be $2.4083(10)^{133}$. Keeping $\delta = 0$ and using $g(\lambda) = \lambda$ and then $g(\lambda) = \lambda^2$ produces the numerators for the first two moments of λ. Dividing by the proportionality constant produced a posterior mean of 2.1872 and second moment of 5.5294 for a posterior standard deviation of .8634. The final integral used $\delta = 1$ and $g(\lambda) = 1$. Dividing the result by the proportionality constant produced the posterior mean of α, 0.0000137347.

If doing 86 integrals simultaneously is a problem, the Bayesian solution could be approximated by using the traditional Whittaker formula with $\lambda = 2.1872$. To get the posterior mean, the function $g(\lambda)$ now must produce the 86 element vector of graduated values. This was done to produce the values in Table 1.

References

Berger, J. (1985). *Bayesian Inference in Statistical Analysis, 2nd ed.*, New York: Springer-Verlag.

Gersh, W. and Kitagawa, G. (1988). "Smoothness Priors in time Series," in *Bayesian Analysis of Time Series and Dynamic Models*, J. Small, ed., New York: Marcel Dekker.

Kimeldorf, G. and Jones, D. (1967), "Bayesian Graduation," *Transactions of the Society of Actuaries*, **19**, 66-.

Klugman, S. (1981). "On the Variance and Mean Squared Error of Decrement Estimators," *Transactions of the Society of Actuaries*, **33**, 301-.

Klugman, S. (1991). *Bayesian Statistics in Actuarial Science with Emphasis on Credibility*, Boston: Kluwer Academic Publishers, to appear.

Lindley, D. and Simth, A. (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society, Series B*, **34**, 1-41.