Analysis of Mortality Data using Smoothing Spline Poisson Regression

N. D. Shyamalkumar
Dept. of Stat. & Act. Sci., The University of Iowa

Abstract

We study a smoothing spline Poisson regression model for the analysis of mortality data. Being a non-parametric approach it is intrinsically *robust*, that it is a penalized likelihood estimation method makes available an approximate Bayesian confidence interval and importantly the software **gss**, its implementation on the freely available statistical package **R**, makes it easily accessible to the user. All of this make it an attractive alternative to (usually computationally intensive) fully Bayesian analysis while avoiding the complexity of high-dimensional prior specification.

1. Introduction

Estimation of mortality rates from data is a well studied problem by not only actuaries but also by demographers and statisticians. There is an extensive literature with procedures adopting either a purely mathematical, frequentist or a Bayesian approach. It is now well accepted though that the problem is inherently statistical and hence a *graduation* type mathematical approach which does not recognize this variability is no longer in vogue.

As a vestige of the mathematical methods, the earlier statistical methods employed a two step procedure towards graduation. First, they calculated *raw* mortality rates as a ratio of the number of deaths to the **exposure**. Second, they found a smooth set of **graduated** rates *close* to the *raw* mortality rates; this step is known as graduation. The books by Benjamin and Pollard (1993) and London (1985) are good sources for examples of such early methods.

The first step supposedly made suitable adjustments to account for the actuarial studies not being simple binomial experiments. The latter due to not only migration but also due to insured population dynamics. I say supposedly as contrary to the intent behind some of the exposure calculations the *raw* mortality rates turned out to be statistically **biased** and asymptotically **inconsistent** estimates of the underlying mortality rates, see Elveback

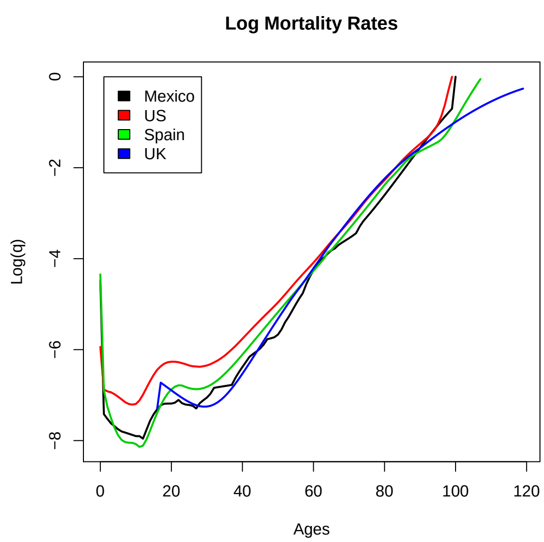
(1958), Breslow and Crowley (1974) and Hoem (1984). More importantly violation of the likelihood principle forced them to make unrealistic assumptions on migration and resort to 'non-standard' reasoning. A more direct approach to calculating *raw* mortality rates is via maximum likelihood (ML) estimation, see Broffitt (1984) for an excellent coverage. By adhering to the likelihood principle, ML estimation avoids the above mentioned problems in **exposure** calculations. Moreover, it indirectly derives the 'correct' **exposure** but by subscribing to a sound statistical reasoning. This is skillfully expressed by Boom (1984) in discussing Broffitt (1984): *This is intuitively superior to the concept of exposure familiar to us in the actuarial estimation/Balducci's assumption combination, since the awkward nonsense of having to expose the already dead still further to the 'risk' of death is avoided.*

Once the first step is framed in a statistical manner it is only natural to combine it with the second step, i.e. simultaneously arriving at smooth 'graduated' rates from the raw data. For example, when using the maximum likelihood estimation method this will result in the unconstrained maximization becoming constrained optimization. Unlike the case of the first step, experience gained in graduation is used in current day models. A well known graduation technique is the use of a mathematical formula, i.e. by using a parametric model. Gompertz (1825) gave the first such model, namely

$$\mu_x = Bc^x \tag{1}$$

which was later generalized by Makeham to

$$\mu_x = A + Bc^x \tag{2}$$



The above two models fail to address infant mortality and the accident hump, both common features of current mortality curves. A relatively more recent parametric model which addresses such features is that of Helligman and Pollard (1980), namely

$$\frac{q_x}{1 - q_x} = A^{(x+B)^C} + D\exp\left[-E\left(\log\left[\frac{x}{F}\right]\right)^2\right] + GH^x \tag{3}$$

An excellent example of maximum likelihood estimation of mortality rates using a parametric model (a generalization of the Gompertz-Makeham formula) can be found in Forfar, McCutcheon and Wilkie (1988).

A statistical paradigm compatible with the likelihood principle is the Bayesian paradigm which has recently gained popularity due to significant increase in accessible computing power. The Bayesian paradigm by explicitly allowing for incorporation of prior beliefs through the prior distribution is particular attractive for mortality rate estimation. Moreover, by delivering the posterior and predictive distributions, a Bayesian analysis empowers a user to answer a myriad of questions. The earlier Bayesian approaches, see Kimeldorf and Jones (1967) and Hickman and Miller (1977), worked with multivariate normal likelihood and prior. As a result of using conjugate priors these model resulted in multivariate normal posteriors. Smoothness constraints were loosely imposed using positive correlation structures. Though computationally easy, prior specification remained an issue with such models. Through reparametrization Broffitt (1984,86 and 88) demonstrated that monotonicity and convexity type constraints could be easily imposed on the estimates and moreover one could continue to use conjugate type priors (keeping computational complexity minimal). Carlin (1992) shows that by the use of monte carlo techniques one could get have the choice of using non-conjugate priors while retaining the ability to impose constraints of the above type. The recent article by Dellaportas, Smith and Stavropoulos (2001) combines the power of monte carlo techniques with the technique of graduation by a mathematical formula. In this way they are able to impose very specific functional form for the estimated mortality curve.

The goal of this work is to propose smoothing spline Poisson regression method for graduation of mortality data. Unlike other similar but parametric methods, this is a nonparametric which brings along a robustness which is appealing. Also, not being a Bayesian approach it avoids the issue of specification of high-dimensional priors. Nevertheless, the equivalent Bayes model provides approximate Bayesian posterior confidence intervals which makes the method very appealing. And finally the availability of the package **gss** on a free statistical environment makes the method very accessible to actuaries. In this article we have a rather simple aggregate mortality example where the power of the method proposed is not brought out. But the real advantage of the method lies when looking at graduation problems involving select and ultimate rates and/or multiple underwriting classes.

The next section briefly discusses the penalized likelihood method. Then in the following section we discuss our model and in the final section we give details of our numerical example. Before we end this section, we refer to Berger (1993) for a lucid discussion and implications of the likelihood principle alluded to above. The problems that arise by abandoning it are well expressed in Pratt (1962); we end this section with an example from it. *An engineer draws a random sample of electron tubes and measures the plate voltages under certain conditions with a very accurate voltmeter, accurate enough so that measurement error is negligible compared with the variability of the tubes. A statistician examines the measurements, which look normally distributed and vary from 75 to 99 volts with a mean of 87 and a standard deviation of 4. He makes the ordinary normal analysis, giving a confidence interval for the true* *mean. Later he visits the engineer's laboratory, and notices that the voltmeter used reads as far as 100, so the population appears to be* **censored**. *this necessitates a new analysis, if the statistician is orthodox. However, the engineer says that he has another meter. equally accurate and reading to 100 volts, which he would have used if any voltage had ever been over 100. this is a relief to the orthodox statistician, because it means that the population was effectively uncensored after all. But the next day the engineer telephones and says, "I just discovered my high-range voltmeter was not working the day I did the experiment you analyzed for me." the statistician ascertains that the engineer would not have held up the experiment until the meter was fixed, and informs him that a new analysis will be required. The engineer is astounded. he says, "But the experiment turned out to be the same as if the high-range voltmeter had been working. I obtained the precise voltages of my sample anyway, so I learned exactly what I would have learned if the high-range meter had been available. next you will be asking about my oscilloscope."*

2. Penalized Likelihood Estimation

Penalized likelihood estimation is a likelihood based estimation method which traces its origin back to Whittaker (1923) (discrete case) and Kimeldorf and Wahba (1970a,70b,71) and Good and Gaskins (1971). The central idea is to estimate a function of interest η on a domain $\mathcal{X}$ by

$$L(\eta|\text{data}) + \frac{\lambda}{2} J(\eta) \tag{4}$$

where $L(\eta|\text{data})$ is the negative log-likelihood and $J(\eta)$ is a quadratic roughness penalty with a low dimensional null-space $\mathcal{N}_J = \{f \in \mathcal{H} : J(f) = 0\}$. The minimizer above can be seen to be the restricted maximum likelihood estimator in a model space $\mathcal{M}_\rho = \{f : J(f) = \rho\}$, for some $\rho > 0$. Moreover, λ above can be seen to be the Lagrange multiplier.

An example of the above formulation is the cubic smoothing spline, see Klugman, Panjer and Willmott (2004). It arises from the regression problem

$$Y_i = \eta(x_i) + \epsilon_i, \quad i = 1, \ldots, n \tag{5}$$

where ϵ_i are independent random variables with $\epsilon_i \sim N(0, \sigma_i^2)$. Let us assume that $x_i \in [0, 1]$, $i = 1, \ldots, n$. Then with the quadratic roughness penalty taken to be

$$\int_0^1 \ddot{\eta}(x)^2 \mathrm{d}x, \quad \text{where } \ddot{\eta} \text{ is the second derivative of } \eta. \tag{6}$$

Now (**4**) in this problem is equivalent to

$$\sum_{i=1}^{n} \left(\frac{Y_i - \eta(x_i)}{\sigma_i} \right)^2 + \frac{\lambda}{2} \int_0^1 \ddot{\eta}(x)^2 \mathrm{d}x \tag{7}$$

and from elementary properties of spline we see that the minimizer has to be a cubic spline which is called the cubic smoothing spline. Observe that the null space $\mathcal{N}_J$ mentioned above in this case is the two dimensional space of all linear functions on $[0,1]$.

The above setup naturally leads to functional spaces which are Reproducing Kernel Hilbert Spaces (RKHS) as the optimization problem then becomes tractable. This is so as the log likelihood, being a smooth function of the evaluation functionals, to be a continuous functional requires that so are the evaluations functionals and this is same as requiring the function space (a Hilbert space) to be a RKHS. Moreover, it is then natural to take the smoothness penalty J to be one induced by a semi-inner product on the RKHS. This general setup is useful in extending the method to multivariate function estimation and allows one to consider interesting classes of smooth functions, see Gu (2002). This is particularly interesting for graduating a select and ultimate table or/and while considering at various underwriting.

Of particular interest is the equivalence of the penalized likelihood method to a Bayes model with a Gaussian prior on $\mathcal{H} \ominus \mathcal{N}_J$ (with the covariance related to the reproducing kernel) and a diffuse prior on $\mathcal{N}_J$. This Bayes model makes available Bayesian posterior confidence intervals. And these become approximate Bayesian confidence intervals in the case of non-Gaussian response as it involves quadratic approximation of the likelihood (Laplace's method).

Smoothing parameter selection is an important practical issue in penalized likelihood estimation. There are different score based methods, the scores being asymptotically close to

$$\frac{1}{n}\sum_{i=1}^{n}(\eta_\lambda(x_i) - \eta(x_i))^2. \tag{8}$$

The idea being that the smoothing parameter is chosen to minimize the above. For details see Gu (2002). Interestingly, using the generalized cross validation based score for choosing the smoothing parameter results in the pointwise Bayesian confidence intervals having good across-the-function coverage, see Gu (2002) and references therein for more details.

3. The Model

Let D_x, for $x = 0, 1, \ldots$, be the number of deaths observed at age (according to any of the common used definition) x. Also let E_x, for $x = 0, 1, \ldots$, be the total amount of time the persons were under observation at age x. We assume that D_x is a Poisson distributed random variable with parameter $E_x\lambda(x)$, i.e.

$$\Pr(D_x = k) := \frac{1}{k!}(\lambda(x)E_x)^k \exp\{-E_x\lambda(x)\}. \tag{9}$$

Such a Poisson distribution has been assumed by many, for e.g. see Forfar, McCutcheon and Wilkie (1988). Assuming the constant force of mortality within integral ages results in the same likelihood as the above Poisson likelihood, see Hoem (1984). This fact has been used in the Bayesian analysis of Broffitt (1988). Also for an argument that the Poisson distribution assumption results from using the above constant force assumption, see Scott (1982, 1984). In this connection also interesting are the articles Sverdup (1965) and Borgan (1984). In the following we will assume the constant force within integral ages.

Given the above model we suppose that $\eta(\cdot) = \log \lambda(\cdot)$ lies in the space of all twice continuously differentiable functions and by using the roughness penalty

$$\int \ddot{\eta}(x)^2 \mathrm{d}x, \quad \text{where } \ddot{\eta} \text{ is the second derivative of } \eta. \tag{10}$$

we force the estimate to be a cubic spline. Estimates for q_x are derived using the formula

$$\mathrm{q}_x = 1 - \exp[-\lambda(x)], \tag{11}$$

which results from the constant force assumption.

4. Numerical Example

One of the important reasons that smoothing spline Poisson regression is easy to use is the availability of the package **gss** on the **R** environment. R is an extensible, well documented language and environment with a core group of developers spread out over many countries. And it is freely available at *www.r-project.org*.

As an example, to demonstrate the simplicity of using the package we estimate the mortality rates from the Female, English and Welsh Mortality data (1988-1992). This is the data set used by Dellaportas, Smith and Stavropoulos (2001). The time axis was re-scaled using a square root transformation to make the ratio D_x/E_x more evenly scattered. A few other reasonable time transformation resulted in similar estimates - hence the answer seemed to be robust to the particular choice. The R-code is given below.

```
t<-sqrt((0:74));
pois.fit <- gssanova((d/e)~t,family="poisson",weights=e);
est <- predict(pois.fit,data.frame(t=t),se=TRUE);
plot((0:74),log(d/e),type="l",xlab="Age", ylab="Log Mortality");
lines((0:74),(est$fit),col=2);
lines((0:74),(est$fit+1.96*est$se),col=3);
lines((0:74),(est$fit-1.96*est$se),col=4);
```

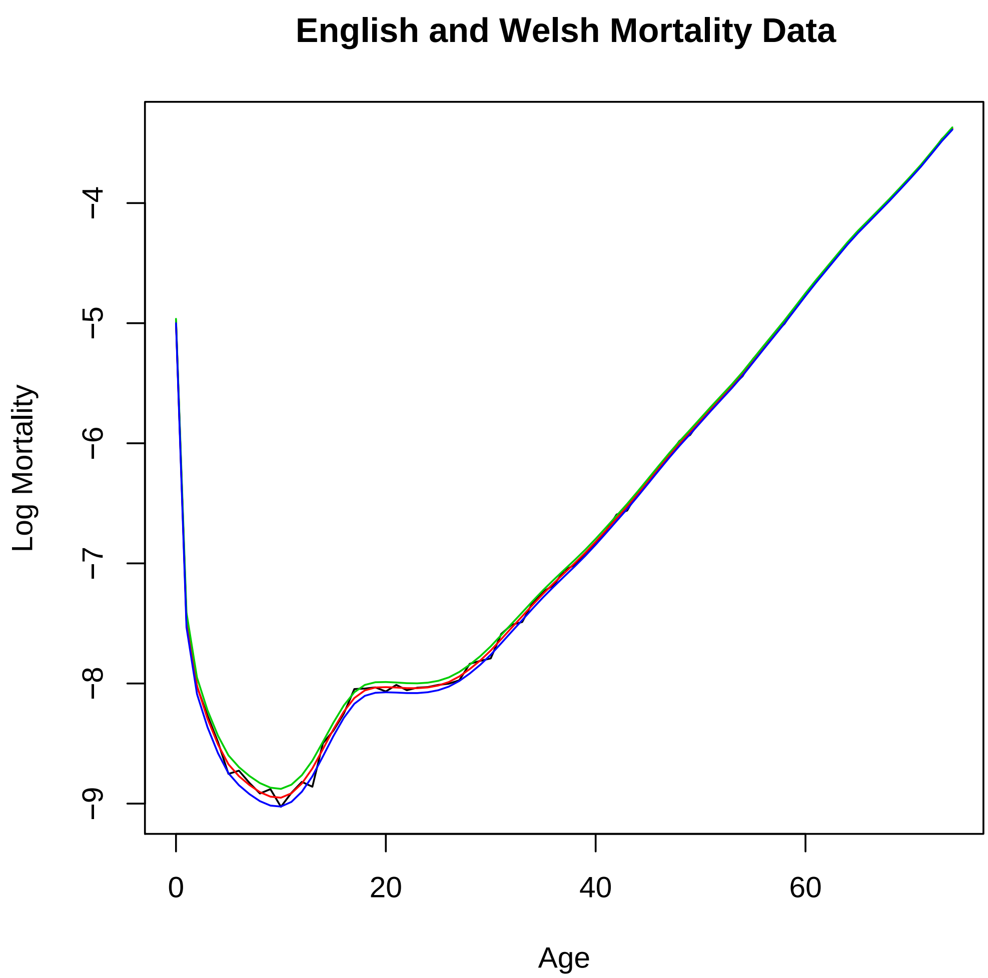


Figure 1 Raw Data (Black), Upper 95%, Lower 95% and Bayes Estimate.

ACKNOWLEDGEMENT

This work was funded by a CKER grant of the Society of Actuaries. I thank Manuel Mendoza and Eduardo Guttierez for valuable discussions. All opinions and errors are mine alone.

REFERENCES

BENJAMIN, B. and POLLARD, J. H. 1993. *The Analysis of Mortality and other Actuarial Statistics*. Third Edition. The Institute of Actuaries and The Faculty of Actuaries, U.K.

BERGER, J. O. 1993. *Statistical Decision Theory and Bayesian Analysis*. Springer, USA.

BORGAN, O. (1984). 'Maximum Likelihood Estimation in a Parametric Counting Process Model, with Applications to Censored Failure Time Data and multiplicative Models'. *Scandinavian Actuarial Journal*, Vol. 11, 1-16.

x	n_x	d_x	x	n_x	d_x	x	n_x	d_x	x	n_x	d_x
0	1682000	11543	19	1822800	592	38	1684600	1516	57	1306200	8068
1	1666400	940	20	1883200	591	39	1707600	1693	58	1314600	8809
2	1644700	538	21	1930400	640	40	1755900	1905	59	1325400	10148
3	1634400	420	22	1964400	623	41	1844500	2207	60	1330600	11390
4	1610000	332	23	2015600	653	42	1837500	2517	61	1332100	12789
5	1581800	250	24	2051700	668	43	1812200	2565	62	1328200	13999
6	1564500	254	25	2078400	689	44	1777100	2918	63	1322300	15528
7	1554700	228	26	2084300	698	45	1699800	3077	64	1323000	17368
8	1549800	208	27	2067000	712	46	1563200	3119	65	1329000	19277
9	1544600	215	28	2021400	799	47	1499200	3369	66	1344200	20991
10	1514300	182	29	1963000	795	48	1453200	3677	67	1370100	23665
11	1482500	200	30	1903800	787	49	1408400	3740	68	1408200	26365
12	1453900	215	31	1844600	935	50	1375500	4130	69	1337400	27664
13	1436700	204	32	1788000	978	51	1365400	4564	70	1249500	28397
14	1443000	294	33	1745500	977	52	1373500	5017	71	1174200	29178
15	1496400	339	34	1714800	1131	53	1361300	5417	72	1098800	30437
16	1576800	412	35	1690300	1219	54	1335900	5786	73	1029500	32146
17	1670500	535	36	1671400	1270	55	1313900	6567	74	1052400	35728
18	1744500	561	37	1668000	1435	56	1306200	7173			

Table 1 English and Welsh Mortality data, females, 1988-1992

Breslow, N. and Crowley, J. 1974. 'A Large Sample Study of the Life table and Product Limit Estimates under Random Censorship'. *Annals of Statistics*, Vol. 2, 437-453.

Broffitt, J. D. (1984) 'Maximum Likelihood Alternatives to Actuarial Estimators of Mortality Rates (with discussion)'. *Transactions of the Society of Actuaries*, Vol. 36, 77-122.

Broffitt, J. D. (1984). 'A Bayes Estimator for Ordered Parameters and Isotonic Bayesian Graduation'. *Scandinavian Actuarial Journal*, 231-47.

Broffitt, J. D. (1986) 'Isotonic Bayesian Graduation with an Additive Prior'. *Advances in Statistical Sciences.* Vol 6, Actuarial Science. Edited by I. B. Macneill and G. J. Umphrey, 19-40. Boston:D. Reidel Publishing Co. 1986.

Broffitt, J. D. (1988). 'Increasing and Increasing Convex Bayesian Graduation (with discussion)'. *Transactions of the Society of Actuaries*, Vol 40, 115-48.

BOOM, H. J. (1984). 'Maximum Likelihood Alternatives to Actuarial Estimators of Mortality Rates (with discussion)'. *Transactions of the Society of Actuaries*, Vol. 36, 77-122.

CARLIN, B. (1992). 'A Simple Monte Carlo approach to Bayesian Graduation'. *Transactions of the Society of Actuaries*, 44, 55-76.

DELLAPORTAS, P., SMITH, A. F. M. and STAVROPOULOS, P. (2000). 'Bayesian Analysis of Mortality Data'. *Journal of the Royal Statistical Society Series A*, Vol. 164, 275-91.

ELVEBACK, L. (1958). 'Estimation of Survivorship in Chronic Disease: the "Actuarial" Method'. *Journal of the American Statistical Association*, Vol. 53, 420-440.

FORFAR, D. O., MCCUTCHEON, J. J. and WILKIE, A. D. (1988). 'On Graduation by Mathematical Formula'. *Journal of the Institute of Actuaries*, 115, 1-135.

GOMPERTZ, B. (1825). 'On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining teh Value of Life Contingencies.' *Phil. Trans. R. Soc.*, Vol 115, 513-585.

GOOD, I. J. and GASKINS, R. A. (1971). 'Nonparametric roughness penalties for probability densities'. *Biometrika*, Vol. 58, 255-277.

GU, C. (2002). *Smoothing Spline ANOVA Models*. Springer, New York.

HELLIGMAN, L. and POLLARD (1980). 'The Age Pattern of Mortality'. *Journal of the Institute of Actuaries*, Vol. 107, 49-80.

HICKMAN, J. C. and MILLER, R. B. (1977). 'Notes on Bayesian Graduation'. *Transactions of the Society of Actuaries*, 29, 1-21.

HOEM, J. M. 1984. 'A Flaw in the Actuarial Exposed-to-Risk Theory'. *Scandinavian Actuarial Journal*, 187-194.

KIMELDORF, G. S. and JONES, D. A. (1967). 'Bayesian Graduation'. *Transactions of the Society of Actuaries*, 19, 66-112.

KIMELDORF, G. S. and Wahba, G. (1970a). 'A Correspondence between Bayesian Estimation of Stochastic Processes and Smoothing by Splines'. *Annals of Mathematical Statistics*, Vol. 41, 495-502.

KIMELDORF, G. S. and Wahba, G. (1970b). 'Spline Functions and Stochastic Processes'. *Sankhya Ser. A.*, Vol. 32, 173-180.

KIMELDORF, G. S. and Wahba, G. (1971). 'Some results on Tchebycheffian Spline Functions'. *J. Math. Anal. Applic.*, Vol. 32, 82-85.

KLUGMAN, S. A., PANJER, H. H. and WILMOTT, G. E. (2004). *Loss Models: From Data to Decisions*. Second Edition, Wiley, New York.

LONDON, D. 1985. *Graduation: The Revision of Estimates*. ACTEX, USA.

PRATT, 1962. 'On the Foundations of Statistical Inference: Discussion'. *Journal of the American Statistical Association*, Vol. 57, 307-326.

SCOTT, W. F. (1982). 'Some Applications of the Poisson Distribution in Mortality Studies', *Transactions of the Faculty of Actuaries*, Vol. 38, 255-263.

SCOTT, W. F. (1984). 'Corrigendum to Some Applications of the Poisson Distribution in Mortality Studies', *Transactions of the Faculty of Actuaries*, Vol. 40, 419-20.

SVERDUP, E. (1965). 'Estimates and Test Procedures in Connection with Stochastic Models for Deaths Recoveries and Transfers between different States'. of Health, *Skandinavisk Aktuar.*, Vol. 48, 184-211.

WHITTAKER, E. T. (1923). On a New Method of Graduation. *Proc. Edinburgh Math. Soc.*, Vol. 41, 63-75.