SOCIETY OF ACTUARIES

Article from:

North American Actuarial Journal

January 2008 – Vol.12 No.1

Note that (D.5) remains a valid formula for determining the volatility parameter, no matter how short the time interval $[a, b]$ is. Also, one can derive (D.5) without the multiplication rules; see Section 10.13, ''A Layman's Guide to Brownian Motion (Wiener Process),'' in Panjer (1998).

Finally, we wish to point out the following problem in the ''Sample Questions and Solutions'' of Exam MFE/3F:

$$\lim_{n\to\infty} \sum_{j=1}^{n} [X(jT/n) - X((j-1)T/n)]^2 = \sigma^2 T. \tag{D.6}$$

A derivation is as follows. From

$$X(t + h) - X(t) = \mu h + \sigma[W(t + h) - W(t)],$$

we have

$$[X(t + h) - X(t)]^2 = \mu^2 h^2 + 2\mu h\sigma[W(t + h) - W(t)] + \sigma^2[W(t + h) - W(t)]^2.$$

With $h = T/n$,

$$\sum_{j=1}^{n} [X(jT/n) - X((j-1)T/n)]^2$$

$$= \mu^2 T^2/n + 2\mu(T/n)\sigma[W(T) - W(0)] + \sigma^2 \sum_{j=1}^{n} [W(jT/n) - W((j-1)T/n)]^2. \tag{D.7}$$

As $n \to \infty$, the first two terms on the right-hand side of (D.7) become 0, and the sum on the right-hand side of (D.7) becomes T in accordance with formula (20.6) of McDonald (2006, p. 653).

REFERENCES

MCDONALD, ROBERT L. 2006. *Derivatives Markets*. 2nd ed. Boston: Addison-Wesley.

PANJER, HARRY H., ED. 1998. *Financial Economics: With Applications to Investments, Insurance and Pensions*. Schaumburg, IL: Actuarial Foundation.

''Markov Aging Process and Phase-Type Law of Mortality,'' X. Sheldon Lin and Xiaoming Liu, October 2007

JOHNNY S. H. LI* AND ANDREW C. Y. NG†

In recent years there has been increasing interest in stochastic modeling of human mortality. This interesting paper by Dr. Lin and Ms. Liu contributed a new analytical mortality law, which has explicitly taken account of the underlying unobservable physiological processes. In this discussion we further explore the computations involved in estimating the model parameters.

In equation (4.2) in Lin and Liu's paper, we see that the survival function under the proposed phase-type law of mortality involves the matrix exponential of $x\boldsymbol{\Lambda}$. The exponential of an $n \times n$ square matrix $\boldsymbol{A}$ is defined by the usual series expansion

$$e^{\boldsymbol{A}} = \boldsymbol{I} + \boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \frac{1}{3!}\boldsymbol{A}^3 + \cdots + \frac{1}{r!}\boldsymbol{A}^r + \cdots.$$

* Johnny S. H. Li, PhD, ASA, is Assistant Professor in the Department of Statistics and Actuarial Science, University of Waterloo, Ontario, Canada, shli@uwaterloo.ca.

† Andrew C. Y. Ng, PhD, ASA, is Assistant Professor in the Department of Finance, the Chinese University of Hong Kong, Hong Kong, andrewng@baf.msmail.cuhk.edu.hk.

The direct use of the definition above to calculate matrix exponentials turns out to be inefficient, even if A is a scalar. When A is a matrix, convengence is very slow when A contains both positive and negative elements, as in the case when $A = \mathbf{\Lambda}$ is the generator of a continuous-time Markov chain or a phase-type distribution. Estimation of the model parameters therefore can be very time consuming.

In this discussion we propose four solutions to this problem. The first solution is an analytic and exact method, and the others are numerical appproximations to the matrix exponential of $x\mathbf{\Lambda}$. The strengths and shortcomings for each of the proposed solutions also will be discussed.

1. AN ANALYTIC AND EXACT METHOD

Let the survival function at age x for a particular cohort be $S(x)$, where

$$S(x) = \boldsymbol{\alpha} e^{x\mathbf{\Lambda}} \boldsymbol{e}.$$

In the expression above $\boldsymbol{\alpha}$ is the first unit row vector, $\boldsymbol{e}$ is a column vector of ones, and

$$\mathbf{\Lambda} = \begin{bmatrix} -(\lambda_1 + q_1) & \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & -(\lambda_2 + q_2) & \lambda_2 & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -(\lambda_{n-1} + q_{n-1}) & \lambda_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & -(\lambda_n + q_n) \end{bmatrix},$$

where $\lambda_n = 0$.

To derive a simple expression of $S(x)$, let $\boldsymbol{g}(x) = \boldsymbol{\alpha} e^{x\mathbf{\Lambda}}$ and denote by $g_i(x)$ the ith component of the row vector $\boldsymbol{g}(x)$, that is,

$$\boldsymbol{g}(x) = [g_1(x), g_2(x), \ldots, g_n(x)].$$

Since

$$\frac{d}{dx} \boldsymbol{g}(x) = \boldsymbol{\alpha} e^{x\mathbf{\Lambda}} \mathbf{\Lambda} = \boldsymbol{g}(x)\mathbf{\Lambda},$$

we have

$$g_i'(x) = \begin{cases} -(\lambda_i + q_i) g_i(x) & \text{when } i = 1 \\ \lambda_{i-1} g_{i-1}(x) - (\lambda_i + q_i) g_i(x) & \text{when } i = 2, \ldots, n. \end{cases}$$

The initial condition of the system of first-order differential equations is $\boldsymbol{g}(0) = \boldsymbol{\alpha}$. It is noteworthy that this exact and analytic method has an interesting probabilistic interpretation. We can consider in more detail $g_i(x)$, the probability for a newborn to be alive at chronological age x, at which he or she is at physiological age i; then the system of equations can be seen as a set of Kolmogorov's forward equations (see Ross 2003, Section 6.4). Owing to the special structure of $\mathbf{\Lambda}$, this system of forward equations can be solved without using the matrix exponential function.

The solution of the first differential equation is obviously

$$g_1(x) = e^{-(\lambda_1 + q_1)x}.$$

To solve the ith ($i \neq 1$) differential equation, we multiply the integrating factor $e^{(\lambda_i + q_i)x}$ on both sides. This yields

$$\frac{d}{dx} [e^{(\lambda_i + q_i)x} g_i(x)] = e^{(\lambda_i + q_i)x} \lambda_{i-1} g_{i-1}(x),$$

from which we obtain

$$g_i(x) = e^{-(\lambda_i+q_i)x}\lambda_{i-1}\int_0^x e^{(\lambda_i+q_i)y}g_{i-1}(y)\,dy.$$

Assume that

$$g_i(x) = \sum_{j=1}^{i} C_{ij}e^{-(\lambda_j+q_j)x}.$$

Then

$$\begin{aligned} g_{i+1}(x) &= e^{-(\lambda_{i+1}+q_{i+1})x}\lambda_i\sum_{j=1}^{i} C_{ij}\int_0^x e^{(\lambda_{i+1}+q_{i+1}-\lambda_j-q_j)y}\,dy \\ &= \lambda_i\sum_{j=1}^{i} C_{ij}\frac{e^{-(\lambda_j+q_j)x}-e^{-(\lambda_{i+1}+q_{i+1})x}}{\lambda_{i+1}+q_{i+1}-\lambda_j-q_j} \\ &= \lambda_i\sum_{j=1}^{i}\frac{C_{ij}e^{-(\lambda_j+q_j)x}}{\lambda_{i+1}+q_{i+1}-\lambda_j-q_j}-\left(\lambda_i\sum_{j=1}^{i}\frac{C_{ij}}{\lambda_{i+1}+q_{i+1}-\lambda_j-q_j}\right)e^{-(\lambda_{i+1}+q_{i+1})x}, \end{aligned}$$

which shows that C_{ij}'s can be obtained from $C_{11} = 1$ and the recursive relations

$$C_{ij} = \begin{cases} \lambda_{i-1}C_{i-1,j}/(\lambda_i+q_i-\lambda_j-q_j) & \text{when } j < i \\ -\lambda_{i-1}\Sigma_{j=1}^{i-1}C_{i-1,j}/(\lambda_i+q_i-\lambda_j-q_j) & \text{when } j = i \\ 0 & \text{when } j > i \end{cases}$$

for $i > 1$. Notice also that the sum of C_{ij}'s over j is always 0 for $i > 1$. Finally,

$$\boldsymbol{g}(x) = [e^{-(\lambda_1+q_1)x}, e^{-(\lambda_2+q_2)x}, \ldots, e^{-(\lambda_n+q_n)x}]\boldsymbol{C}'$$

and hence

$$S(x) = [e^{-(\lambda_1+q_1)x}, e^{-(\lambda_2+q_2)x}, \ldots, e^{-(\lambda_n+q_n)x}]\boldsymbol{C}'\boldsymbol{e}.$$

The formula above gives a quick method to calculate the survival function for all ages 1, 2, . . . , m simultaneously when the number of states is small. Suppose we want to calculate $[S(1), S(2), \ldots, S(m)]'$, then we may first compute

$$\boldsymbol{X} = \begin{bmatrix} e^{-(\lambda_1+q_1)} & e^{-(\lambda_2+q_2)} & \cdots & e^{-(\lambda_n+q_n)} \\ e^{-2(\lambda_1+q_1)} & e^{-2(\lambda_2+q_2)} & \cdots & e^{-2(\lambda_n+q_n)} \\ \vdots & & \ddots & \vdots \\ e^{-m(\lambda_1+q_1)} & e^{-m(\lambda_2+q_2)} & \cdots & e^{-m(\lambda_n+q_n)} \end{bmatrix}.$$

The survival function at various ages then can be evaluated by

$$[S(1), S(2), \ldots, S(m)]' = \boldsymbol{X}\boldsymbol{C}'\boldsymbol{e}.$$

The analytic method avoids matrix inversion and can calculate the survival function for many ages by using one multiplication. It is very efficient if the number of states is small. But when the number of states is large and λ_i and q_i are very close, C_{ij} may oscillate and even blow up.

2. THE PADÉ APPROXIMATION

The Padé approximation of a function f is the best approximation of f in the form of a rational function of given orders in the numerator and denominator. This method is used intensively in computer languages because it often gives better approximation of the function than truncating its Taylor series, and it may still work where the Taylor series does not converge. As explained in Section 11 of Moler and Van Loan (2003), the matrix exponential function (expm) in MATLAB uses the Padé approximation together with scaling and squaring to approximate e^A. Although the source code of expm is not distributed, the expm1 function in the M-file demos directory shows that e^A is computed in MATLAB by first calculating the numerator matrix $\boldsymbol{E}$ and the denominator matrix $\boldsymbol{D}$ in the Padé approximation, followed by the operation $\boldsymbol{D}^{-1}\boldsymbol{E}$.

The Padé approximation suffers from the following problems:

1. For $\boldsymbol{D}$ to be nonsingular, the order of the numerator and denominator of the rational function used in Padé approximation has to be large.
2. Even if $\boldsymbol{D}$ is nonsingular, it may be very close to singularity.

3. DIAGONALIZATION

Diagonalization is a standard method to evaluate powers of matrices, and this method also can be applied to evaluate matrix exponentials. First, we consider the simplest case when $\boldsymbol{A} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix. Since $\boldsymbol{A}^k = \text{diag}(\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k)$,

$$e^{\boldsymbol{A}} = \text{diag}(e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n}).$$

If $\boldsymbol{A}$ has distinct eigenvalues, then it is similar to a diagonal matrix: that is, there exists a diagonal matrix $\boldsymbol{F} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and a nonsingular matrix $\boldsymbol{Q}$ such that $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{F}\boldsymbol{Q}^{-1}$. Since $\boldsymbol{A}^k = \boldsymbol{Q}\boldsymbol{F}^k\boldsymbol{Q}^{-1}$,

$$e^{\boldsymbol{A}} = \boldsymbol{Q}\,\text{diag}(e^{\lambda_1}, e^{\lambda_2}, \ldots, e^{\lambda_n})\boldsymbol{Q}^{-1}.$$

Finally, for the general case, let $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{J}\boldsymbol{Q}^{-1}$ where $\boldsymbol{Q}$ is a nonsingular matrix and $\boldsymbol{J} = \text{diag}(\boldsymbol{J}_1, \boldsymbol{J}_2, \ldots, \boldsymbol{J}_m)$ is the Jordan canonical form of $\boldsymbol{A}$. Here each $\boldsymbol{J}_k$ is an $n_k \times n_k$ square matrix of the form

$$\boldsymbol{J}_k = \lambda_k \boldsymbol{I}_{n_k} + \boldsymbol{E}_k,$$

where all the elements in the main diagonal and the lower triangular part of $\boldsymbol{E}_k$ are zero. Also, $\sum_{k=1}^{m} n_k = n$. Then by definition,

$$e^{\boldsymbol{A}} = \boldsymbol{Q}\,\text{diag}(e^{\boldsymbol{J}_1}, e^{\boldsymbol{J}_2}, \ldots, e^{\boldsymbol{J}_m})\boldsymbol{Q}^{-1},$$

where

$$e^{\boldsymbol{J}_k} = e^{\lambda_k \boldsymbol{I}_{n_k}} e^{\boldsymbol{E}_k} = e^{\lambda_k} \sum_{i=0}^{\infty} \frac{[\boldsymbol{E}_k]^i}{i!} = e^{\lambda_k} \sum_{i=0}^{n_k-1} \frac{[\boldsymbol{E}_k]^i}{i!}$$

since $\boldsymbol{E}_k$ is nilpotent.

Diagonalization is theoretically attractive but may be practically inapplicable for the following two reasons:

1. To obtain the Jordan canonical form, one needs to compute eigenvalues. A single rounding error in the computer can alter the structure of $\boldsymbol{Q}$ and $\boldsymbol{J}$.
2. Although $\boldsymbol{Q}$ is nonsingular, it can be very close to a singularity.

4. AN APPROXIMATION BASED ON A TRANSFORMATION OF $x\Lambda$

As one can see from the above, for Padé approximation and diagonalization, even if $\boldsymbol{A}$ is well behaved, $\boldsymbol{D}$ and $\boldsymbol{Q}$ can be so poorly conditioned that the matrix inversion blows up due to rounding errors.

Ross (2003, Section 6.8) makes use of the fact that

$$e^{x\boldsymbol{\Lambda}} = \lim_{n\to\infty} \left(\boldsymbol{I} + \frac{x}{n}\boldsymbol{\Lambda}\right)^n$$

to approximate $e^{x\Lambda}$. Let $n = 2^k$ and define

$$M = I + \frac{x}{2^k}\Lambda.$$

Then we can calculate M^2 by $M \times M$, M^4 by $M^2 \times M^2$, until

$$e^{x\Lambda} \approx M^{2^k} = M^{2^{k-1}} \times M^{2^{k-1}}.$$

In this way $e^{x\Lambda}$ can be approximated by using only k multiplications.

It is desirable to select a value of k such that the the diagonal elements of M (and hence all elements of M) are nonnegative, so that convergence can be achieved more easily. But in some cases the required value of k would be very large. A remedy of this problem is to consider

$$e^{x\Lambda} = \lim_{n\to\infty}\left[\left(I - \frac{x}{n}\Lambda\right)^{-1}\right]^n,$$

let

$$M = \left(I - \frac{x}{2^k}\Lambda\right)^{-1},$$

and calculate M^{2^k} by k multiplications. It can be shown easily that all elements of M are nonnegative.

The methods we discussed are applicable to not only Lin and Liu's mortality model, but also other Markovian mortality/morbidity models including, for example, the model for genetics of breast and ovarian cancer (MacDonald et al. 2003a,b).

REFERENCES

MACDONALD, ANGUS S., HOWARD R. WATERS, AND CHESSMAN T. WEKWETE. 2003a. The Genetics of Breast and Ovarian Cancer I: A Model of Family History. *Scandinavian Actuarial Journal*, 1–27.

———. 2003b. The Genetics of Breast and Ovarian Cancer II: A Model of Critical Illness Insurance. *Scandinavian Actuarial Journal*, 28–50.

MOLER, CLEVE, AND CHARLES VAN LOAN. 2003. Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-five Years Later. *SIAM Review* 45(1): 3–49.

ROSS, SHELDON M. 2003. *Introduction to Probability Models*. 8th ed. San Diego, CA: Academic Press.