

**SIMULATING RANDOM VARIATES  
FROM MAKEHAM'S DISTRIBUTION  
AND FROM OTHERS WITH EXACT  
OR NEARLY LOG-CONCAVE DENSITIES**

**DAVID P.M. SCOLLNIK**

**ABSTRACT**

This paper describes how Markov chain Monte Carlo and related methods recently presented in the statistical literature can be used to quickly and efficiently simulate random draws from distributions with log-concave or nearly log-concave densities. These methods should be useful in many actuarial computer simulations, because a number of distributions in actuarial contexts have log-concave or nearly log-concave densities. To illustrate their application, the paper examines how they can be used to simulate realizations of life contingent functions under Makeham's law and certain other patterns of mortality when the distribution of the future lifetime random variable either has a log-concave density or possesses one that is nearly so. These simulation methods allow a variety of previously inaccessible inferences to be made routinely and easily. Several examples are included.

---

**1. INTRODUCTION**

The analytical laws of mortality devised by Gompertz [16] and Makeham [24] are well-known to actuaries and are also known to be very good representations of the mortality process at adult ages. Unfortunately, neither law lends itself to convenient mathematical analysis. As noted by London [23, p. 18–19] and Anderson [1, p. 124], among others, even evaluating characteristics like the mean or variance of their respective survival distributions is somewhat difficult. Computing the expected value of the life contingent functions defining annuities or insurances under either Gompertz's or Makeham's law is even more burdensome, as is the calculation of any of the associated higher moments or tail probabilities. Fortunately, a number of authors, including Mereu [25], Moore [28], and Carriere [6], have studied and facilitated the application of these laws, thus making them far more palatable to the practicing actuary. Nevertheless, some of these analytical results have been rather limited and yet have typically involved reasonably sophisticated mathematical analyses, often requiring the numerical

evaluation of the incomplete or left-truncated gamma function, for instance, Moore [28] and Carrier [6]. Further, none of these references has considered simulating random variates from either Gompertz's or Makeham's survival distribution.

In this paper I take a different approach and describe how Markov chain Monte Carlo (MCMC) methods, related to some of those recently appearing in the *Transactions* [5], can be used to quickly and efficiently simulate realizations of life contingent functions under Makeham's law and certain other patterns of mortality when the distribution of the future lifetime random variable either has a log-concave density or possesses one that is nearly so. For this purpose, a density function  $f(t)$  is said to be log-concave if the logarithm of this function is concave with respect to the argument  $t$  (that is, the first derivative of  $\ln[f(t)]$  is a monotonically decreasing function of  $t$ ). If the second derivative of  $\ln[f(t)]$  exists, then the density  $f(t)$  will be log-concave, provided that

$$\frac{d^2}{dt^2} \ln [f(t)] \leq 0$$

for all values of  $t$  in the domain  $D = \{t | f(t) > 0\}$  of the density function. The MCMC methodology described allows a variety of previously inaccessible inferences to be made routinely and easily, for example, in studying the distribution of a prospective loss random variable associated with some insurance contract to determine the necessary reserve under one of the patterns of mortality described above. However, the MCMC methods discussed here are very general and actuarial practitioners should find many other uses for them (see Section 5). The mortality laws of Gompertz and Makeham provide a context for the main examples.

## 2. THE MORTALITY LAWS OF GOMPERTZ AND MAKEHAM

The force of mortality for a life age  $x$  under Gompertz's law is given by

$$\mu_x = BC^x, B > 0, C \geq 1, x \geq 0,$$

and by

$$\mu_x = A + BC^x, A > -B, B > 0, C \geq 1, x \geq 0$$

under Makeham's (London [23] and Bowers et al. [3]). Gompertz's law is obviously a special case of Makeham's. Makeham intended the extra parameter  $A$  to capture that part of the hazard that is independent of age

[23, p. 19]. Although negative values of the parameter  $A$  may have little practical application, the case  $-B < A < 0$  is included for completeness. Henceforth, reference is made only to Makeham's law, with the understanding that Gompertz's law is a special instance of it.

As in Chapter 3 of Bowers et al. [3], let  $X$  denote the age-at-death random variable for a newborn life, let the symbol  $(x)$  denote a life age  $x$ , and let  $T(x)$  denote the future lifetime random variable for  $(x)$ . Then the survival function for a newborn life under Makeham's law is given by

$$\begin{aligned} s(x) &= \Pr[X > x] \\ &= \exp\left[-Ax - \frac{B}{\ln C} (C^x - 1)\right], x \geq 0; \end{aligned}$$

the survival function for  $(x)$  is given by

$${}_t p_x = \Pr[T(x) > t] = \frac{s(x+t)}{s(x)}, t \geq 0;$$

and the probability density function of the random variable  $T(x)$  is given by

$$\begin{aligned} f(t) &= {}_t p_x \mu_{x+t} \\ &= \exp\left[-At - \frac{BC^x}{\ln C} (C^t - 1)\right] (A + BC^{x+t}), t \geq 0. \end{aligned}$$

The most interesting observation is that

$$\frac{d^2}{dt^2} \ln [f(t)] = -BC^{x+t} \left[1 - \frac{A \ln C}{(A + BC^{x+t})^2}\right] \ln C.$$

When this expression is non-positive for all values of  $t$ , the continuous density  $f(t)$  is log-concave with respect to its argument. Inspection of the expression above indicates that log-concavity will always be obtained provided that  $-B < A \leq 0$ . On the other hand, log-concavity may not be obtained when  $A > 0$ . However, the Makeham parameters are usually confined to the ranges

$$0.001 < A < 0.003, \quad 0.000001 < B < 0.001, \quad 1.08 < C < 1.12$$

when describing human mortality [21, p. 24], and so the value of the second derivative of the log-density will still be very close to zero in this case even when  $A > 0$ . Consequently, the distribution of the future lifetime random variable  $T(x)$  has a density function that is either concave on the logarithmic

scale (that is, log-concave) or else very nearly so in the case of human mortality under Makeham's law. The significance of this result becomes apparent in the following sections.

### 3. A NUMBER OF USEFUL MCMC-RELATED METHODS

Recently, Carlin [5] showed that a random draw from a complicated  $k$ -variate distribution with density  $f(x_1, x_2, \dots, x_k)$  can be generated by iteratively and repeatedly sampling from each of the associated univariate full conditional distributions  $f(x_1|x_2, \dots, x_k)$ ,  $f(x_2|x_1, x_3, \dots, x_k)$ , ...,  $f(x_k|x_1, \dots, x_{k-1})$  in turn. This algorithm describes a very simple implementation of the so-called Gibbs sampler, a type of MCMC method. A MCMC method is a sampling-based procedure that is used to generate a sequence of dependent random draws from a distribution of interest on a fast computer. Over the last six or seven years, MCMC methods have become very popular within the statistical sciences (for example, Gelfand and Smith [11], Smith and Roberts [33], Tanner [34], and Tierney [35]), and a great deal of effort has been expended to develop efficient and general-purpose random number generators to implement the necessary draws from the full conditional distributions described above (for example, Wakefield et al. [36], Gilks and Wild [13], Gilks [12], and Wild and Gilks [37]). MCMC methods have also recently begun to be applied to actuarial problems. The Gibbs sampler was utilized by Carlin [4] to conduct the Bayesian state space modeling of non-standard actuarial time series, by Carlin [5] and Klugman and Carlin [22] in the context of Bayesian graduation, by Scollnik [31] to implement a Bayesian analysis of a simultaneous equations model for insurance rate-making, and by Scollnik [32] to conduct the Bayesian analysis of some generalized Poisson models. The Gibbs sampler is not used in this paper, but two methods that have recently appeared in the related MCMC literature are used: adaptive rejection sampling (ARS) and adaptive rejection Metropolis sampling (ARMS). These methods are introduced below along with some background material.

#### A. *The Rejection Sampling Method*

The rejection sampling method allows one to obtain a random draw from a distribution with a continuous density  $f(x)$ . It requires that a proposal distribution with density  $g(x)$  be available such that random draws from this distribution are easily obtained and also such that there exists a finite constant  $M > 0$  with  $f(x) \leq Mg(x)$  for all values of  $x$  in the domain of  $f(x)$ . Then

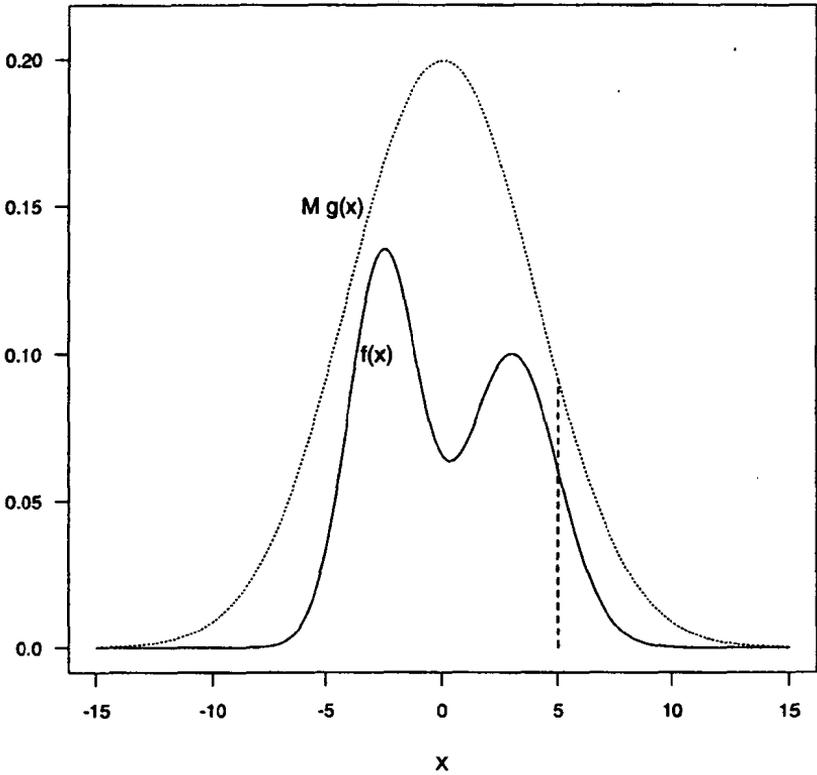
the following algorithm is known to yield a random draw from the distribution with density  $f(x)$ :

- Step 1.* Generate a uniform random variate  $U$  on the unit interval.  
*Step 2.* Generate a random variate  $X$  from the distribution with density  $g(x)$ .  
*Step 3.* Determine whether  $U < f(X)/Mg(X)$ . When this inequality is *false*, return to *Step 1*. When this inequality is *true*, accept the current value of  $X$  and stop the algorithm.

The value of  $X$  finally accepted will be a random draw from the distribution with density  $f(x)$ . The rejection sampling method may be applied as many times as necessary in the case that a random sample of size greater than one is required from the distribution with density  $f(x)$ . The number of iterations of this algorithm required in order to generate each accepted draw is known to be geometrically distributed with a mean of  $M$ , so this algorithm may be inefficient unless a good proposal distribution is available so that the value of  $M$  is close to 1. Rejection sampling and its properties are discussed in more detail by Devroye [10, p. 40–43] and Ross [30, p. 478–83]. These authors also present proofs that the algorithm performs as claimed. The rejection sampling method is included in the course of reading for Associateship Course 130, Operations Research [18, p. 871–2].

The rejection sampling method can also be described pictorially as in Figure 1. In this figure, the bimodal distribution with density  $f(x)$  is the one from which we desire a random draw. The density  $f(x)$  is plotted with a solid (—) line. The proposal distribution has a density  $g(x)$ , and some constant  $M$  is presumed to be known such that  $f(x) \leq Mg(x)$  for all values of  $x$  in the domain of  $f(x)$ . The curve  $Mg(x)$  is plotted with a dashed (---) line. We assume that a method is available that permits us to implement random draws from the proposal distribution. Then any value  $X$  sampled from the proposal distribution will be accepted with a probability equal to the ratio of the height  $f(X)$  to the height  $Mg(X)$ . These heights are indicated on the figure for the value  $X=5$ . Rejected values are discarded. This procedure can be repeated as many times as necessary in the case in which a random sample of size greater than one is required from the distribution with density  $f(x)$ , and the final collection of accepted values constitutes a random sample from the distribution with density  $f(x)$ .

FIGURE 1  
AN ILLUSTRATION OF THE REJECTION SAMPLING METHOD



**B. Numerical Example 1**

The following simple example illustrates the rejection sampling method. Imagine that we want to generate a random variate from the beta distribution (with parameters 3 and 5) having the density function

$$f(x) = 105 x^2 (1 - x)^4, 0 < x < 1.$$

Since the random variable is concentrated on the unit interval, we use the rejection sampling method together with the uniform proposal distribution

$$g(x) = 1, 0 < x < 1.$$

To determine the constant  $M$  such that  $f(x) \leq Mg(x)$ , we need to determine the maximum value of the ratio

$$\frac{f(x)}{g(x)} = 105 x^2 (1 - x)^4,$$

when  $0 < x < 1$ . Differentiation of this ratio yields

$$\frac{d}{dx} \left[ \frac{f(x)}{g(x)} \right] = 210 x (1 - x)^4 - 420 x^2 (1 - x)^3.$$

Setting this expression equal to zero and solving for  $x$  shows that the maximal value of the ratio is obtained when  $x = 1/3$ , so that

$$\frac{f(x)}{g(x)} \leq 105 \left( \frac{1}{3} \right)^2 \left( \frac{2}{3} \right)^4 = \frac{560}{243} \equiv M$$

and

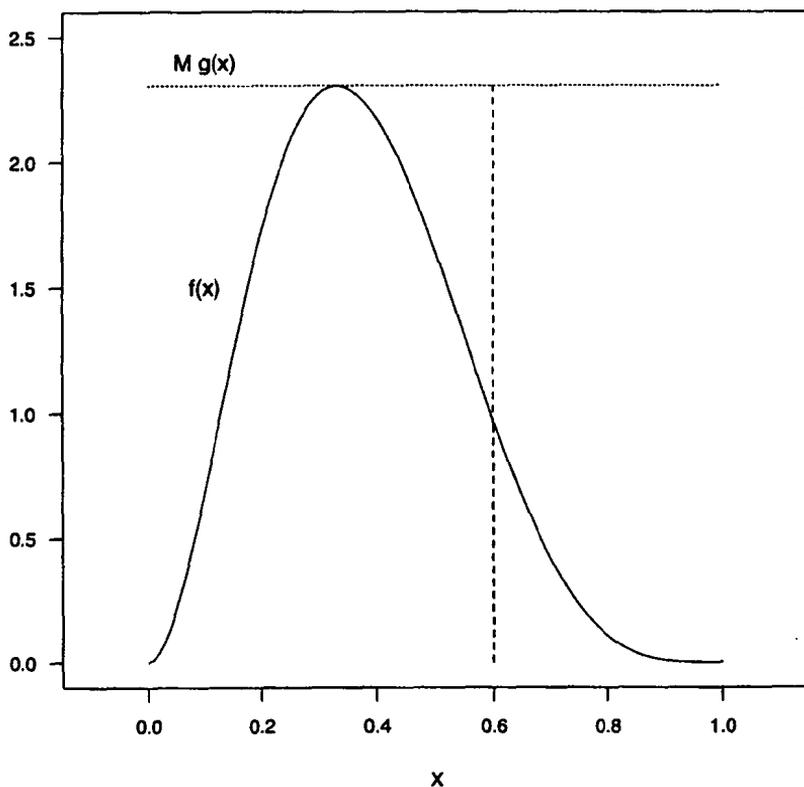
$$\frac{f(x)}{Mg(x)} = \frac{729}{16} x^2 (1 - x)^4.$$

When this last expression is substituted into the rejection sampling algorithm, we find that to generate a random variate from the original beta distribution  $f(x)$  of interest, we should perform the following procedure:

- Step 1.* Generate a uniform random variate  $U$  on the unit interval.
- Step 2.* Generate a random variate  $X$  from the uniform proposal distribution with density  $g(x)$ .
- Step 3.* Determine whether  $U < f(X)/Mg(X) = (729/16) X^2 (1-X)^4$ . If this inequality is *false*, return to *Step 1* and start over. If this inequality is *true*, then accept the current value of  $X$  and stop the algorithm.

The value of  $X$  finally accepted will be a random draw from the desired beta distribution with density  $f(x)$ . This example is described pictorially in Figure 2. As mentioned previously, the number of iterations required to generate an accepted draw is known to be geometrically distributed with a mean of  $M$ , so for this simple example approximately 2.3 iterations of the algorithm are required. In a sense this is an atypical result, since many applications of the rejection sampling method require dozens or even hundreds of iterations on average to generate a single accepted draw.

FIGURE 2  
A SECOND ILLUSTRATION OF THE REJECTION SAMPLING METHOD



### ***C. The Adaptive Rejection Sampling Method***

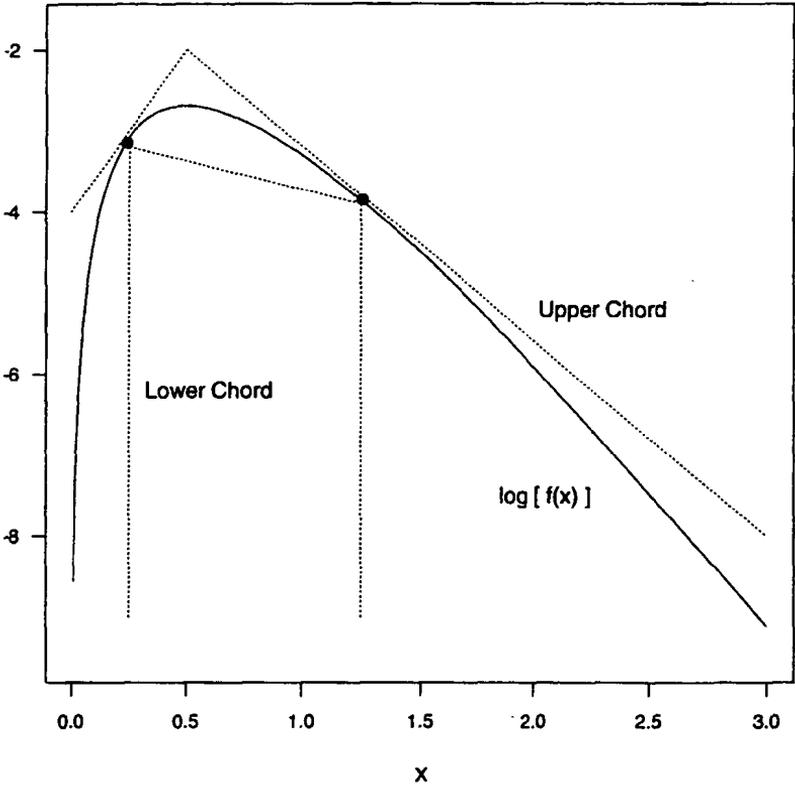
Two difficulties are associated with this rejection sampling method. First, the practitioner must be able to specify a proposal distribution with a density  $g(x)$  satisfying the relation  $f(x) \leq Mg(x)$  for some constant  $M$ . Second, since  $M$  is the average number of iterations required to generate an accepted draw, the practitioner must be able to deduce a small value of  $M$ , so that the implementation of the rejection sampling method will be efficient. Gilks and Wild [13] and Gilks [12] have designed ARS methods that remove the difficulties associated with, and improve upon, straightforward rejection

sampling when the random draw is required from a distribution with a log-concave density function  $f(x)$ . Specifically, their ARS methods construct the proposal distribution automatically and cleverly update it periodically, so that the probability of rejecting a candidate draw from the proposal distribution decreases monotonically from iteration to iteration. ARS can be used to obtain an independent random sample of any size from a distribution with a log-concave density. The following paragraphs describe the main idea of ARS, specify the formal algorithm, clearly state the conditions necessary for its successful utilization, and indicate how programming code implementing ARS can be obtained.

Briefly, ARS proceeds by using the fact that a concave function can be squeezed between two bounds or hulls with piece-wise linear structures. The original formulation of ARS given by Gilks and Wild [13] constructs the upper hull by joining tangent lines evaluated at points along the function's domain and the lower hull by joining chords between these same points. This notion of bounding a log-concave function is illustrated in Figure 3. When this construction is applied to the logarithm of a log-concave density, the exponentiated upper hull defines a distribution with a piece-wise exponential form. This is the proposal distribution from which random draws are initially generated. A random draw from the proposal distribution is then accepted with a probability related to how closely the true density is squeezed between the upper and lower hulls at the value of the random draw. Furthermore, if the draw fails a certain test condition, then the upper and lower hulls are updated to incorporate the value of this draw. As the hulls are updated, the proposal distribution comes to more closely resemble the distribution from which random draws are sought, and the probability of rejecting a candidate draw from the proposal distribution monotonically decreases. By carefully tailoring the rejection step and test condition, accepted draws are independently generated realizations from the distribution of interest with the original log-concave density function.

The ARS method of Gilks and Wild [13] can be thought of as a black-box technique for sampling from any univariate log-concave probability density function  $f(x)$ . The density need only be specified up to a constant of integration, so that rejection sampling can be performed by using  $g(x)$  instead of  $f(x)$ , where  $g(x) = cf(x)$  for some possibly unknown value of  $c$  (this is not the same  $g(x)$  as in the previous section). This is useful when the form of the density is available, but the normalizing constant  $c = \int g(x)dx$  is difficult to compute. To use ARS to sample from a density function  $f(x)$ , it is necessary to check that the domain  $D = \{x | f(x) > 0\}$  is a connected set, that  $g(x)$

FIGURE 3  
 BONDING A LOG-CONCAVE DENSITY FUNCTION



is continuous and differentiable everywhere in  $D$ , and that  $h(x)=\ln[g(x)]$  is concave everywhere in  $D$  (that is, the first derivative of  $h(x)$  decreases monotonically with increasing  $x$  in  $D$ ). If the second derivative of  $\ln[f(x)]$  exists, then it suffices to check that

$$\frac{d^2}{dx^2} \ln [ f(x) ] \leq 0,$$

for all values of  $x$  in the domain of the density function, to ensure that  $h(x)$  is concave as required.

A formal description of the ARS algorithm of Gilks and Wild [13] begins by supposing that  $h(x)$  and  $h'(x)$  have been evaluated at  $k$  points  $x_1 \leq x_2 \leq \dots \leq x_k$  in the domain of  $f(x)$ . Let  $T_k = \{x_i; i=1, \dots, k\}$  and let  $u_k(x)$  and  $l_k(x)$  denote the piece-wise linear upper hull formed by connecting the tangents to  $h(x)$  at the points in  $T_k$ , and the piece-wise linear lower hull formed by connecting the chords between the adjacent points in  $T_k$ , respectively (in the spirit of Figure 3). Define the sampling density

$$S_k(x) = \frac{\exp[u_k(x)]}{\int \exp[u_k(x)] dx}$$

The following ARS algorithm allows as many independent random draws as desired to be made from the distribution with density proportional to  $f(x)$ .

- Step 1. Select and order the values in  $T_k$ . If the domain of  $f(x)$  is unbounded on the left, then select  $x_1$  so that  $h'(x_1) > 0$ . If the domain of  $f(x)$  is unbounded on the right, then select  $x_k$  so that  $h'(x_k) < 0$ .
- Step 2. Generate a uniform random variate  $U$  on the unit interval.
- Step 3. Generate a random variate  $X$  from the sampling density  $S_k(x)$ .
- Step 4. Determine whether  $U < \exp[l_k(X) - u_k(X)]$ . When this inequality is *true*, accept the current value of  $X$ . When this inequality is *false*, determine whether  $U < \exp[h(X) - u_k(X)]$ . If this second inequality is *true*, then accept the current value of  $X$ ; otherwise reject it.
- Step 5. If the first inequality in Step 4 is *false*, then insert  $X$  into the set  $T_k$ , increment  $k$ , relabel and reorder the values in  $T_k$ , and redefine the functions  $u_k(x)$ ,  $l_k(x)$ , and  $S_k(x)$ .
- Step 6. Return to Step 2 and iterate this procedure until as many values of  $X$  have been accepted as required. The final collection of accepted values constitutes a random sample from the distribution with density proportional to  $f(x)$ .

Although the ARS algorithm of Gilks and Wild [13] is fairly simply stated, designing and then programming an efficient implementation of it might prove to be a daunting task. Fortunately, this is not required. Wild and Gilks [37] discuss an implementation of the ARS method, which appears as Algorithm AS 287 in *Applied Statistics* [37], and Fortran code implementing ARS is freely available on the computer Internet in the StatLib archive site\* maintained at Carnegie Mellon University. Gilks [12] describes

\*Send an electronic mail message containing only the two words *send index* to statlib@lib.stat.cmu.edu to receive a message describing the contents of this archive site and how to

a *derivative-free* variant of ARS, which neither assumes continuity in the derivatives of  $f(x)$  nor requires the evaluation of these derivatives.

TABLE 1  
DENSITIES USEFUL IN ACTUARIAL APPLICATIONS  
AND THEIR LOG-CONCAVITY PROPERTIES

Name of Density	Parameters	Log-Concave wrt:
Exponential	$\lambda$	$x, \ln x, \lambda$
Gamma	$\alpha, \beta$	$\ln x, x$ (if $\alpha \geq 1$ ), $\alpha, \beta$
Lognormal	$\mu, \sigma^2$	$\ln x, \mu, 1/\sigma, \log \sigma$
Normal	$\mu, \sigma^2$	$x, \mu, 1/\sigma, \log \sigma$
Pareto	$\alpha, \lambda$	$\ln x, \alpha$
Weibull	$c, \tau$	$x$ (if $\tau \geq 1$ ), $\ln x, c, \tau$
Bernoulli	$p$	$p, \text{logit } p$
Binomial	$n, p$	$p, \text{logit } p$
Poisson	$\lambda$	$\lambda$
Negative Binomial	$r, p$	$p, \text{logit } p$

Random variate generation techniques for distributions with log-concave density functions should be of interest to the actuarial practitioner precisely because so many of the distributions in practical actuarial applications share this property. A number of these densities, along with their log-concavity properties, are listed in Table 1. These densities are parameterized as in Hogg and Klugman [20]. When a density is indicated as being concave on the logarithmic scale with respect to a transformation of its random variable (for example,  $\ln x$ ) in Table 1, this means that the density for the transformed random variable is log-concave with respect to its argument. As shown in Table 1, a number of these densities are also log-concave with respect to one or more of their continuous parameters, which can prove useful in the context of a Bayesian analysis (for an example, see Gilks and Wild [13]). Of course, many distributions besides those listed in Table 1 also possess log-concave densities. In fact, for the general location-scale family of distributions with densities of form

$$f(x|\theta, \sigma) \propto \frac{1}{\sigma} \phi \left( \frac{x - \theta}{\sigma} \right)$$

retrieve selections from it. Send an electronic mail message containing only the four words *send index from apstat* to [statlib@lib.stat.cmu.edu](mailto:statlib@lib.stat.cmu.edu) to obtain an index of all the *Applied Statistics* algorithms currently archived in StatLib. Finally, send an electronic mail message containing only the four words *send 287 from apstat* to [statlib@lib.stat.cmu.edu](mailto:statlib@lib.stat.cmu.edu) to obtain Fortran code implementing Algorithm AS 287.

the logarithm of the density will be concave with respect to  $x$ ,  $\theta$ , and  $\tau=1/\sigma$ , provided that the function  $\phi(z)$  is log-concave with respect to  $z$  ([13, p. 343]).

#### ***D. The Metropolis-Hastings Algorithm***

Rejection sampling and ARS each allow generation of a sequence of independent draws from a distribution of interest. On the other hand, a MCMC method generates a sequence of dependent draws from some distribution. Good discussions of these methods are provided by Smith and Roberts [33], Tanner [34], and Tierney [35]. The Metropolis algorithm (Metropolis et al. [26]) is a form of MCMC method. Hastings [17] proposed a generalization of this method, which has become known as the Metropolis-Hastings (MH) algorithm. The MH algorithm allows a dependent sequence of random draws to be sampled from a distribution with density proportional to  $f(x)$ , provided that a proposal distribution with density  $g(x|z)$  is available such that a random draw is easily obtained from it for any value of  $z$  in the domain of  $f(x)$ . Then the MH algorithm proceeds as follows:

- Step 1.* Select a starting value  $X_0$  and set  $i$  equal to 0.
- Step 2.* Generate a uniform random variate  $U$  on the unit interval.
- Step 3.* Generate a random variate  $X$  from  $g(x|X_i)$ .
- Step 4.* Determine whether

$$U < \min \left( 1, \frac{f(X) g(X_i|X)}{f(X_i) g(X|X_i)} \right).$$

When this inequality is *true*, accept the current value of  $X$  and set  $X_{i+1}$  equal to  $X$ . When this inequality is *false*, reject the current value of  $X$  and set  $X_{i+1}$  equal to  $X_i$ .

- Step 5.* Increment  $i$  to  $i+1$  and return to Step 2.

After a suitably large number of iterations, relevant asymptotic theory (for example, Smith and Roberts [33] and Tierney [35]) states that the sequence  $X_1, X_2, \dots$ , can be considered to be a dependent random sample from  $f(x)$  in the sense that

$$X_t \xrightarrow{d} X \sim f(x) \text{ as } t \rightarrow \infty$$

and

$$\frac{1}{t} \sum_{i=1}^t h(X_i) \rightarrow E_f[h(X)] \text{ as } t \rightarrow \infty,$$

almost surely.

In plain English, the first result says that as  $t$  becomes moderately large, the value  $X_t$  is very nearly a random draw from  $f(x)$ . In practice, a value of  $t \approx 10$  to 15 is typically more than sufficient. This result also allows an approximately independent random sample to be generated from the distribution with density  $f(x)$  by using only every  $k$ -th value appearing in the sequence. The value of  $k$  should be taken to be large enough so that the sample autocorrelation function coefficients for the values appearing in the subsequence are reminiscent of those for a purely random process or stochastically independent sequence, that is, until there are no significant autocorrelations at non-zero lags. An illustration is provided below. Autocorrelation functions are covered in some depth in the course of reading for Associateship Course 120, Applied Statistical Methods (also see Miller and Wichern [27, 333–7, 356–65]). The second result says that if  $h$  is an arbitrary  $f$ -integrable function of  $X$ , then the average of the function  $h$  taken over the sampled values  $X_t$  (the ergodic average of the function) converges (almost surely, as  $t \rightarrow \infty$ ) to its expected value under the distribution with density  $f(x)$ .

Note that the proposal distribution  $g(x|z)$  must satisfy some mild regularity conditions to guarantee that the MH algorithm converges in the sense described above. Basically, if  $u$  and  $v$  are two values in the domain of  $f(x)$ , then the proposal distribution utilized must permit it to be possible to move from  $u$  to  $v$  in a finite number of iterations of the MH algorithm with non-zero probability. Further, the number of iterations required to move from  $u$  to  $v$  should not have to be a multiple of some number. These conditions are usually satisfied if  $g(x|z)$  has a positive density on the domain of  $f(x)$ , or else has a positive density over a restricted domain (for example,  $g(x|z)$  corresponds to a uniform distribution around  $z$  with finite width) [8], [9]. In short, the practitioner has considerable freedom when selecting the proposal distribution. A popular choice is for the proposal distribution  $g(x|z)$  to correspond to a normal distribution with mean  $z$ , or perhaps a heavier tailed student  $t$  distribution.

Another common selection is to let  $g(x|z)$  correspond to a fixed distribution, independent of  $z$ , that is,  $g(x|z) \equiv g(x)$ . An example of this is a normal distribution with zero mean. In practice, the proposal distribution should be selected so that the autocorrelations in the sequence of simulated values

generated by the MH algorithm are not too high, and this is often the case if the proposal distribution is selected in such a way that the average number of times the acceptance check in Step 4 of the MH algorithm rejects a candidate draw is about 50% [8], [9].

### ***E. Numerical Example 2***

The MH algorithm is illustrated with a trivial yet illuminating example. Consider the distribution with density

$$f(x) = 2x, 0 \leq x \leq 1.$$

The MH algorithm is applied with a uniform proposal distribution on the unit interval, so that

$$g(x|z) \equiv g(x) = 1, 0 \leq x \leq 1.$$

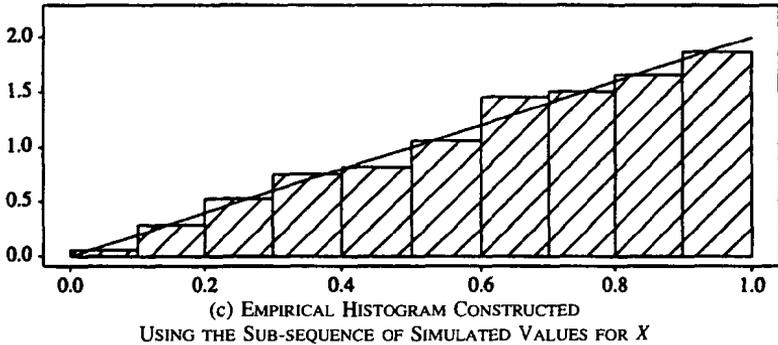
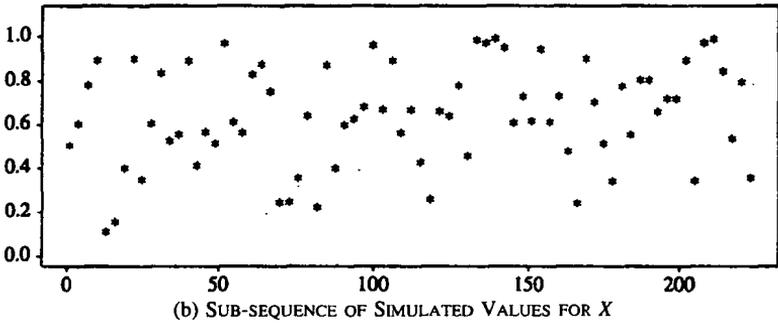
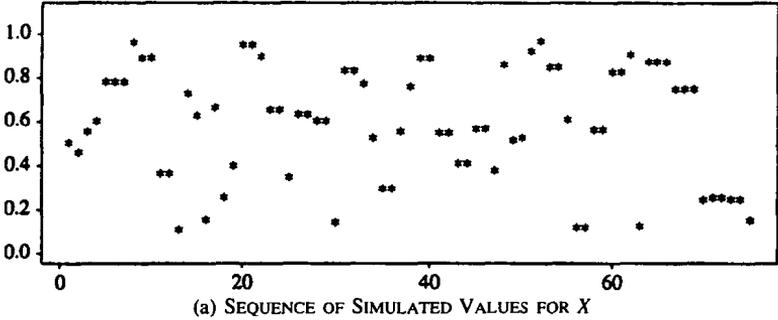
In this case, the acceptance probability appearing in Step 4 of the MH algorithm becomes

$$\min\left(1, \frac{f(X) g(X_i|X)}{f(X_i) g(X|X_i)}\right) \equiv \min\left(1, \frac{X}{X_i}\right).$$

To initialize the MH algorithm, the starting value  $X_0$  was arbitrarily set equal to 0.5. The  $(i+1)$ -st iteration was entered with the value  $X_i$ ; a candidate value  $X$  was sampled from the uniform proposal distribution on the unit interval; and the iteration was exited with the value of  $X_{i+1}$  set equal to  $X$  with a probability equal to  $\min(1, X/X_i)$ . Otherwise,  $X_{i+1}$  was assigned the iteration's entering value  $X_i$ . A total of 3,000 iterations of this algorithm was performed, and the first 75 values of  $X$  are plotted in Figure 4a. The dependent nature of this sequence is immediately perceptible.

This observation can be confirmed by examining the sample autocorrelation function for this sequence of 3,000 simulated values. This sample autocorrelation function appears as Figure 5a. The heights of the 20 different spikes in this plot represent the values of the sample autocorrelation coefficients at lags 0 through 19. If the sequence of 3,000 sampled values is truly independent, then all the sample autocorrelations at non-zero lags should be close to zero. Spikes crossing either of the two horizontal dashed lines identify autocorrelation coefficients that are significantly different from zero (at the 95% level of significance). For the sequence of 3,000 sampled values, significant autocorrelations are identified at the non-zero lags 1 through 4, reinforcing the original impression that this sequence was

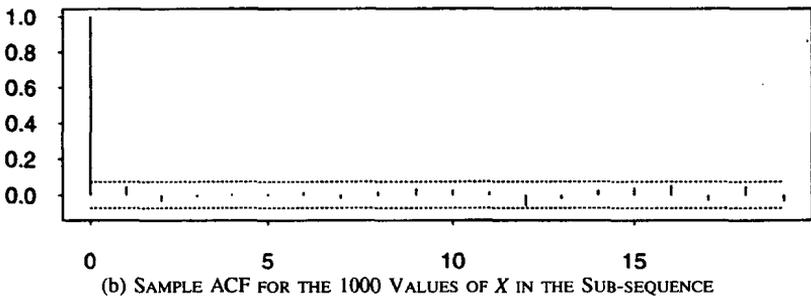
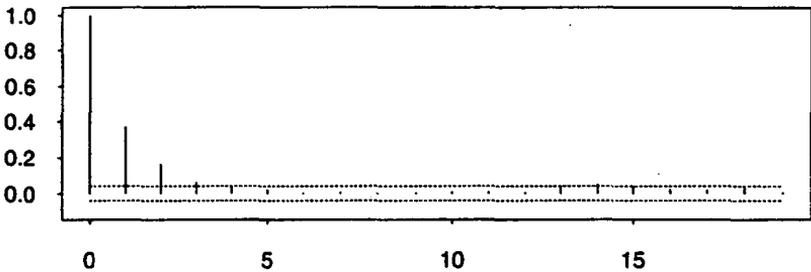
FIGURE 4  
PLOTS FOR NUMERICAL EXAMPLE 2



dependent. Even so, the empirical mean (0.66121) and variance (0.05520) of these 3,000 sampled values are each very close to the exact theoretical values associated with the random variable with distribution  $f(x)$  (0.66667 and 0.05556, respectively).

FIGURE 5

SAMPLE AUTOCORRELATION FUNCTION PLOTS FOR NUMERICAL EXAMPLE 2



Incidentally, the implementation of the MH algorithm for this example was monitored as it ran on a fast computer, and the acceptance check in Step 4 of the algorithm rejected 978 of the 3,000 candidate draws from the proposal distribution. Thus, the observed rejection rate was 32.6%. The rejection rate is sometimes referred to as the “staying rate,” since when a candidate draw  $X$  from the proposal distribution is rejected in Step 4 of the MH algorithm, the value of  $X_{i+1}$  stays the same as that of  $X_i$ . Obviously, this accounts for the dependence observed in the sequence of 3,000 simulated values.

To recover an approximately independent random sample from this sequence, the sub-sequence of 1,000 values was formed by keeping only every third value in the original sequence of 3,000 (that is,  $k=3$  in the discussion above). The first 75 values of this sub-sequence constitute Figure 4b. The sample autocorrelation function for this sub-sequence of 1,000 values appears as Figure 5b. This sample autocorrelation function is reminiscent of what we would expect for a purely random process, since none of the autocorrelations at non-zero lags is significantly different from zero. An empirical histogram constructed using the 1,000 simulated values of the sub-sequence appears as Figure 4c, along with the exact density function for the distribution of interest. We note that the correspondence between the empirical histogram and the exact density is very good, indicating that the sub-sequence does appear to form a random sample from the desired distribution.

#### ***F. The Adaptive Rejection Metropolis Sampling Method***

As its name suggests, ARMS is a hybrid method combining ARS with the MH algorithm. ARMS generalizes ARS to deal with distributions possessing non-log-concave densities by appending a MH acceptance step (like Step 4 in the description of the MH algorithm above) to the ARS algorithm. In general, ARMS is much more convenient to use than MH alone, since ARMS automatically constructs the required proposal distribution using ARS. ARMS also operates such that when the distribution of interest does possess a log-concave density, the MH step always accepts and ARMS simply reduces to ARS. This means that there is no additional overhead in using ARMS when just ARS would have sufficed. When the distribution of interest does not possess a log-concave density, the ARMS algorithm defines a valid MCMC method and generates a dependent sequence of random draws from the distribution of interest. A detailed description of ARMS can be found in Gilks et al. [14], [15]. Those two papers also announce the free availability of C code implementing ARMS, utilizing a derivative-free version of ARS, and indicate that it may be obtained from the authors on request (e-mail wally.gilks@mrc-bsu.cam.ac.uk). This C code was utilized to perform the simulations discussed in the next section. I will gladly provide interested readers with the main calling program used in conjunction with the ARMS code to implement the random draws from the Makeham survival distribution described in the next section.

#### 4. APPLYING ARS AND ARMS IN THE MAKEHAM CONTEXT

When operating under Makeham's law, as noted at the end of Section 2, the distribution of the future lifetime random variable  $T(x)$  either has a log-concave density or possesses one that is nearly so for all practical selections of the Makeham constants. The discussion in Section 3 states that a very efficient and fast random number generator now exists, which can be used to simulate a sequence of (possibly dependent) random draws from the distribution of  $T(x)$ , in these cases. By utilizing well-known Monte Carlo simulation methods (for example, Ross [29], Tanner [34]), it will also be possible to easily infer a variety of characteristics related to the distribution of various life contingent functions under the assumption of Makeham's law. The remainder of this section provides a few examples illustrating some of the applications of this method.

##### A. Numerical Example 3

Consider a life age 50 whose future lifetime random variable  $T(50)$  is assumed to follow the Makeham distribution, as in Section 2, with parameters

$$A = 0.001, B = 0.0000070848535, C = 1.1194379,$$

as found in Mereu [25] and Moore [28]. ARMS was utilized to generate 5,000 dependent realizations of  $T(50)$  from this Makeham distribution. This simulation took a second or two to run on a desktop UNIX workstation (that is, a SUN Sparcstation LX operating at 50 MHz).† The MH acceptance step in the ARMS algorithm rejected none of the 5,000 candidate draws from the proposal distribution, indicating that the ARS part of ARMS was doing a very good job at constructing and updating an efficient proposal distribution. This comes as no great surprise, since in Section 2 the density of the Makeham survival distribution was very nearly log-concave, so that the ARS algorithm should be able to construct a proposal distribution that closely matches the Makeham survival distribution of interest. As in Example 2, the sample autocorrelation function was examined for the dependent sequence of 5,000 simulated values, and it was reminiscent of that for a purely random process (that is, none of the autocorrelations at non-zero

†This simulation time, along with those reported below, should be comparable to that using a fast PC (for example, 486 or Pentium operating at 50 Mhz or faster). In fact, since my Sparcstation is circa 1992 with no upgrades, I would expect even better simulation times using a fast Pentium-based PC.

lags was significantly different from zero). Thus, the 5,000 dependent realizations were treated as independent draws for all practical intents and purposes. The empirical histogram for these sampled values is plotted against the exact density for the random variable  $T(50)$  in Figure 6a to demonstrate that they do appear to constitute a random sample from the desired distribution. Recall that the exact density function  $f(t)$  for the future lifetime random variable  $T(x)$  under Makeham's law was previously derived in Section 2.

For this same life, consider a newly issued unit whole life insurance with benefit payable at the moment of death, so that the present value of the benefit is described by the random variable

$$Z = v^{T(50)} = \exp[-\delta T(50)], \quad T(50) \geq 0,$$

in which  $\delta \geq 0$  denotes the force of interest. Also, consider a continuous whole life annuity with the random variable

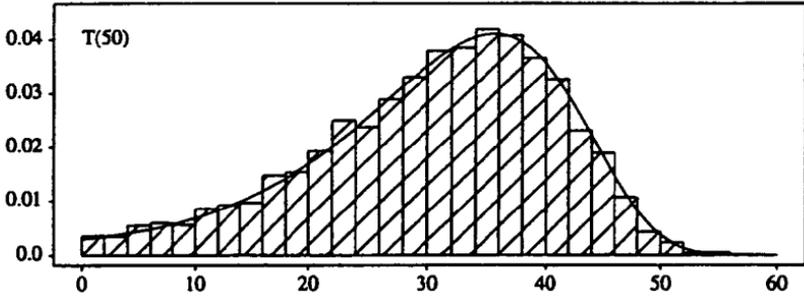
$$Y = \bar{a}_{\overline{T(50)|}} = \frac{1 - \exp[-\delta T(50)]}{\delta}, \quad T(50) \geq 0,$$

describing the present value of the annuity payments. Both of these random variables are typical of the sort found in Bowers et al. [3]. We applied these two transformations to the 5,000 previously sampled values of  $T(50)$  to generate (effectively independent) random samples from the distributions of  $Z$  and  $Y$ , taking the force of interest to be equal to  $\delta = \ln(1.025)$  for illustrative purposes. In Figures 6b and 6c the empirical histograms of these transformed values have been plotted against the exact density functions for  $Z$  and  $Y$  to demonstrate that they essentially constitute random samples from the distributions we claim. The exact density functions for the random variables  $Z$  and  $Y$  were obtained by applying the standard transformation of variable technique to the density function of  $T(50)$  (for example, Hogg and Craig [19, p. 132-3]).

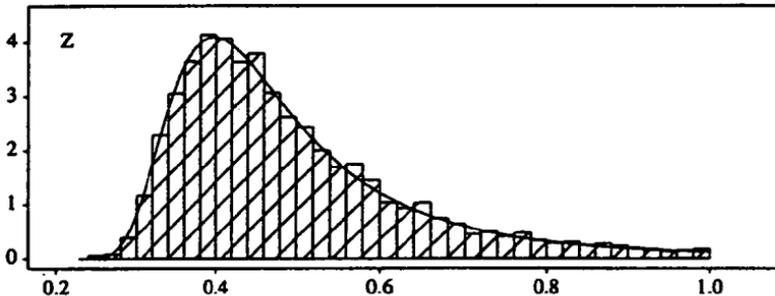
At this time ARMS once again was utilized to generate 250,000 realizations of  $T(50)$  from the same Makeham distribution as before; these sampled values were also transformed into realizations of  $Z$  and  $Y$ . This simulation took less than 2 minutes of real time on the same desktop SUN Sparcstation LX. The empirical mean, variance, skewness, kurtosis, and 95th percentile of these sampled values were used to estimate the corresponding population quantities. These empirical values are presented in the third column of Table 2. Estimated approximate Monte Carlo standard errors [34, p. 30] are also

FIGURE 6

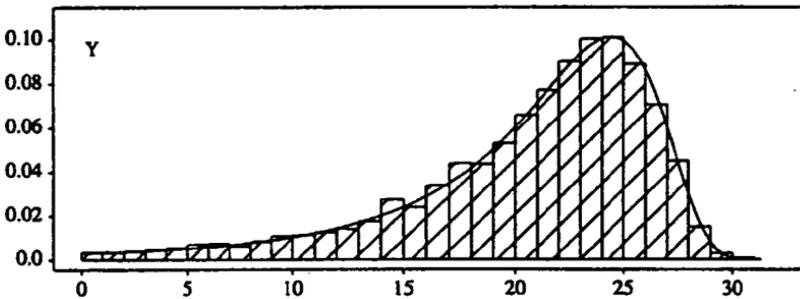
EXACT DENSITIES AND EMPIRICAL HISTOGRAMS FOR NUMERICAL EXAMPLE 3



(a) PLOTS FOR THE FUTURE LIFETIME RANDOM VARIABLE  $T(50)$



(b) PLOTS FOR THE INSURANCE BENEFIT PRESENT VALUE RANDOM VARIABLE  $Z$



(c) PLOTS FOR THE ANNUITY PRESENT VALUE RANDOM VARIABLE  $Y$

included for the three empirical means in brackets beside these entries in the third column. Recall that if  $n$  is the Monte Carlo sample size and  $x_1, \dots, x_n$  denote the independently sampled values, then the estimated Monte Carlo standard error associated with the sample mean is given by

$$\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}$$

This time the MH acceptance step in the ARMS algorithm rejected a single one of the 250,000 candidate draws from the proposal distribution, indicating once again that the ARS part of ARMS was doing a very good job at constructing and updating an efficient proposal distribution. As before, the sample autocorrelation function for the dependent sequence of 250,000 simulated values was examined and no autocorrelations at non-zero lags were significantly different from zero. This means that the 250,000 dependent realizations can be treated as independent draws for all practical intents and purposes, and so we are comfortable using the definition for the estimated Monte Carlo standard error associated with the sample mean given above.

Finally, the values in the second column of Table 2 are numerical approximations to the exact theoretical values and were determined with the

TABLE 2  
NUMERICAL APPROXIMATIONS AND EMPIRICAL VALUES  
FOR NUMERICAL EXAMPLE 3

Quantity of Interest	Numerical Approximation	Empirical Value
E [T(50)]	30.81125	30.79607 (0.02086)
Variance [T(50)]	108.87118	108.80472
Skewness [T(50)]	-0.60911	-0.61111
Kurtosis [T(50)]	2.96685	2.96872
95th Percentile (T(50))	45.39895	45.35621
E [ $v^{T(50)}$ ]	0.48388	0.48406 (0.00027)
Variance [ $v^{T(50)}$ ]	0.01851	0.01852
Skewness [ $v^{T(50)}$ ]	1.26008	1.26134
Kurtosis [ $v^{T(50)}$ ]	4.50669	4.50692
95th Percentile ( $v^{T(50)}$ )	0.77004	0.77025
E [ $\bar{a}_{T(50)}$ ]	20.90160	20.89457 (0.01102)
Variance [ $\bar{a}_{T(50)}$ ]	30.36526	30.38141
Skewness [ $\bar{a}_{T(50)}$ ]	-1.26008	-1.26134
Kurtosis [ $\bar{a}_{T(50)}$ ]	4.50669	4.50692
95th Percentile ( $\bar{a}_{T(50)}$ )	27.29774	27.2838

aid of a symbolic mathematics computer package (namely, Maple V Release 3). A symbolic mathematics package represents another means by which simple characteristics like moments or tail probabilities associated with the Makeham survival distribution can be determined. Unfortunately, these packages are not always well-suited for more sophisticated problems (for example, like numerical example 4 below), and they are not as commonplace as Fortran or C compilers (Fortran and C being two languages for which programming code implementing ARS and ARMS is freely and readily available). Further, symbolic mathematics packages are usually not designed for simulation applications, so that they may not be very convenient computing environments in which to implement simulation methods such as ARS or ARMS.

### ***B. Numerical Example 4***

Now consider a whole life insurance issued to a life age 30 with a 10,000 benefit payable at the moment of death and with a premium of an amount equal to  $\pi$  payable at the beginning of each year. For this policy, the random variable describing the present value of the loss faced by the insurer is defined by

$$L = 10,000 v^{T(30)} - \pi \ddot{a}_{\overline{K(30)+1}|}, \quad T(30) \geq 0,$$

where  $K(x)$  is used to denote the curtate future lifetime random variable for  $(x)$ . The value of  $\pi$  is determined according to some premium principle as in Chapter 6 of Bowers et al. [3]. The equivalence principle, for example, selects  $\pi$  such that the insurer's expected loss is equal to zero (that is,  $E[L]=0$ ). For the present numerical illustration, we set  $\pi$  as small as possible such that the insurer faces a positive loss on this single insurance contract with a probability no greater than 5% (that is,  $Pr[L>0] \leq 0.05$ ). Unfortunately, the distribution of  $L$  is somewhat awkward to study analytically since  $L$  is a function of the continuous random variable  $T(30)$  along with its discrete valued part  $K(30)$ . This would also complicate the application of a symbolic mathematics package. Nevertheless, we may easily estimate the value of  $\pi$  on the basis of a Monte Carlo simulation.

Specifically, if the future lifetime random variable  $T(30)$  is assumed to follow a Makeham distribution as in Section 2, then it is a simple matter to simulate a large number of realizations of  $T(30)$  using ARMS and, for a given value of  $\pi$ , then transform these sampled values into realizations of  $L$ . On the basis of just such a Monte Carlo simulation, assuming that the

Makeham parameters remained as in numerical example 3, taking the force of interest  $\delta$  to be equal to  $\ln(1.075)$ , and utilizing 250,000 sampled values of  $T(30)$ , we determined that for a value of  $\pi$  equal to 143.20, the probability  $Pr[L > 0]$  is approximately equal to 0.05. Incidentally, during this simulation the MH acceptance step in the ARMS algorithm rejected 1,796 of the 250,000 candidate draws from the proposal distribution, thus yielding a rejection rate of 0.7184%. This rejection rate is very small and indicates that the ARS part of ARMS is doing a good job at constructing and updating an efficient proposal distribution. Note that this rejection rate is considerably larger than the one encountered in the previous example, possibly indicating that the Makeham survival distribution for a life age 30 is not quite as log-concave as it is for a life age 50.

To test the value of  $\pi$  determined above, another 250,000 realizations of  $T(30)$  were sampled. The empirical histogram of these values is presented in Figure 7a along with a plot of the exact density curve. Taking  $\pi$  equal to 143.20, we transformed these values into realizations of  $L$ . The proportion of these realizations exceeding zero was 5.082% (that is, the observed value of  $L$  exceeded zero 12,705 times out of 250,000). The empirical histogram for these sampled values of  $L$  is also presented in Figure 7b. Note that this second histogram is plotted on a square root scale, so that the proportion of realizations of  $L$  observed between  $-2,000$  and  $-1,500$  is approximately equal to  $500 \times 0.039 \times 0.039 = 0.7605$ , for example. A square root scale was adopted for the vertical axis of this histogram to better illustrate the long tail of the distribution on the right-hand side. Table 3 presents the empirical mean, variance, skewness, kurtosis, and 95th percentile of the sampled values of  $L$ .

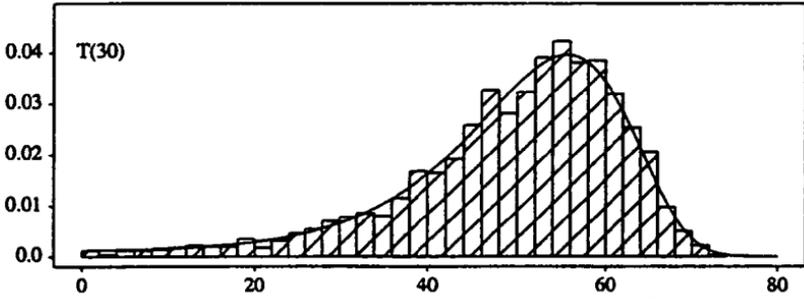
Finally, assume that the life under study survives to age 50 and that we are interested in studying the adequacy of the reserve at that time. Using ideas found in Chapter 7 of Bowers et al. [3], the insurer's prospective loss random variable at that time is defined by

$${}_{20}L = 10,000 v^{T(50)} - 143.20 \ddot{a}_{\overline{K(50)+1}|}$$

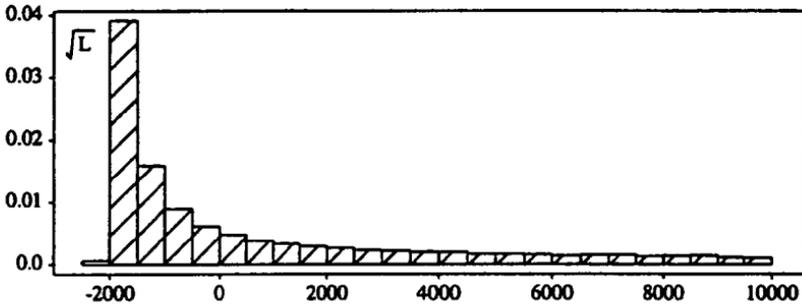
By now, it should be apparent that we can simulate random draws from the distribution of  ${}_{20}L$  by sampling values of  $T(50)$  using ARMS and then applying the appropriate transformation to them. We simulated 250,000 realizations of  ${}_{20}L$  in this manner and plotted the empirical histogram for the sampled values in Figure 7c. As before and for the same reason, this histogram is plotted on a square root scale. Table 3 presents the empirical mean,

FIGURE 7

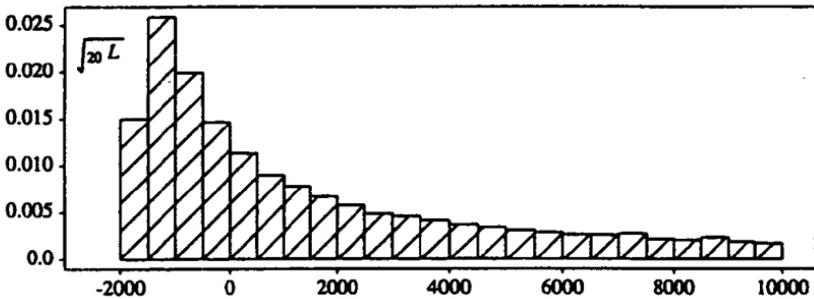
EXACT DENSITY AND EMPIRICAL HISTOGRAMS FOR NUMERICAL EXAMPLE 4



(a) PLOTS FOR THE FUTURE LIFETIME RANDOM VARIABLE  $T(30)$



(b) PLOTS FOR THE LOSS RANDOM VARIABLE  $L$



(c) PLOTS FOR THE LOSS RANDOM VARIABLE  ${}_{20}L$

TABLE 3  
EMPIRICAL VALUES FOR NUMERICAL EXAMPLE 4

Quantity of Interest	Empirical Value
$E[L]$	-1,444.17600 (2.31665)
Variance $[L]$	1,341,721
Skewness $[L]$	5.49276
Kurtosis $[L]$	39.11419
95th Percentile $[L]$	12.89028
$E[{}_{20}L]$	-267.6755 (3.57699)
Variance $[{}_{20}L]$	3,198,709
Skewness $[{}_{20}L]$	2.68140
Kurtosis $[{}_{20}L]$	11.31819
95th Percentile $[{}_{20}L]$	3,529.356

variance, skewness, kurtosis, and 95th percentile of the sampled values of  ${}_{20}L$ .

### 5. CLOSING REMARKS

This paper discusses the application of MCMC-related methods to simulate life contingent functions under Makeham's law and certain other patterns of mortality, when the distribution of the future lifetime random variable either has a log-concave density or possesses one that is nearly so. Specifically, the application of recently formulated ARS and ARMS methods is discussed, and several illustrations are provided. As noted in the main text, ARS and ARMS can be used to simulate random draws from a wide variety of other distributions commonly used by actuarial practitioners. These include, but are not restricted to, the gamma, lognormal, normal, Pareto, and Weibull distributions. Many other problems in actuarial science can be expected to benefit from an application of MCMC and related methods. One possible application may be to use MCMC methods to simulate from arbitrary aggregate claims distributions. I expect to report upon this application in the near future.

A variety of very timely research papers concerned with MCMC methods are currently available on the computer Internet via the MCMC Preprint Service located at <http://www.statslab.cam.ac.uk/>. Those papers in the reference list with an asterisk were available through this service when this paper was being prepared.

## REFERENCES

1. ANDERSON, A.W. *Pension Mathematics for Actuaries*. 2nd ed. Wellesley, Ma.: Arthur W. Anderson, 1990.
2. BECKER, R.A., CHAMBERS, J.M., AND WILKS, A.R. *The New S Language*. Pacific Grove, Calif.: Wadsworth & Brooks, 1988.
3. BOWERS, N.L., GERBER, H.U., HICKMAN, J.C., JONES, D.A., AND NESBITT, C.J. *Actuarial Mathematics*. Itasca, Ill.: Society of Actuaries, 1986.
4. CARLIN, B.P. "State Space Modeling of Non-Standard Actuarial Time Series," *Insurance: Mathematics and Economics* 11 (1992): 209-22.
5. CARLIN, B.P. "A Simple Monte Carlo Approach to Bayesian Graduation," *TSA XLIV* (1992): 55-76.
6. CARRIERE, J.F. "An Investigation of the Gompertz Law of Mortality," *ARCH* 1994.2 (1994): 161-77.
7. CASELLA, G., AND GEORGE, E.I. "Explaining the Gibbs Sampler," *The American Statistician* 46 (1992): 167-74.
- \*8. CHIB, S., AND GREENBERG, E. "Understanding the Metropolis-Hastings Algorithm," Olin School of Business, University of Washington, 1994. Also available via anonymous ftp from [wublib.wustl.edu](ftp://wublib.wustl.edu/pub/chib/cg_mh.ps) in `pub/chib/cg_mh.ps`.
9. CHIB, S., AND GREENBERG, E. "Understanding the Metropolis-Hastings Algorithm," *The American Statistician* 49 (1995): 327-35.
10. DEVROYE, L. *Non-Uniform Random Variate Generation*. New York: Springer-Verlag, 1986.
11. GELFAND, A.E., AND SMITH, A.F.M. "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association* 85 (1990): 398-409.
12. GILKS, W.R. "Derivative-Free Adaptive Rejection Sampling for Gibbs Sampling," in *Bayesian Statistics 4*, ed. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith. Oxford: University Press, 1992, 641-9.
13. GILKS, W.R., AND WILD, P. "Adaptive Rejection Sampling for Gibbs Sampling," *Applied Statistics* 41, no. 2 (1992): 337-48.
14. GILKS, W.R., BEST, N.G., AND TAN, K.K.C. "Adaptive Rejection Metropolis Sampling for Gibbs Sampling," Technical Report, Medical Research Council Biostatistics Unit, Cambridge, U.K., 1992.
- \*15. GILKS, W.R., BEST, N.G., AND TAN, K.K.C. "Adaptive Rejection Metropolis Sampling within Gibbs Sampling," Technical Report, Medical Research Council Biostatistics Unit, Cambridge, U.K., 1994.
16. GOMPERTZ, B. "On the Nature of a Function Expressive of the Law of Human Mortality, and On a New Mode of Determining the Value of Life Contingencies," *Philosophical Transactions of the Royal Society* 115 (1825): 513-85.

17. HASTINGS, W.K. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika* 57 (1970): 97-109.
18. HILLIER, F.S., AND LIEBERMAN, G.J. *Introduction to Operations Research*. 5th ed. New York: McGraw Hill Publishing Company, 1990.
19. HOGG, R.V., AND CRAIG, A.T. *Introduction to Mathematical Statistics*. 4th ed. New York: Macmillan Publishing Co., Inc., 1978.
20. HOGG, R.V., AND KLUGMAN, S.A. *Loss Distributions*. New York: John Wiley & Sons, 1984.
21. JORDAN, C.W. *Life Contingencies*. 2nd ed. Chicago, Ill.: Society of Actuaries, 1982.
22. KLUGMAN, S.A., AND CARLIN, B.P. "Hierarchical Bayesian Whittaker Graduation," *Scandinavian Actuarial Journal* 2 (1993): 183-96.
23. LONDON, D. *Survival Models and Their Estimation*. 2nd ed. Winsted and Avon, Conn.: Actex Publications, 1988.
24. MAKEHAM, W.M. "On the Law of Mortality, and the Construction of Annuity Tables," *The Assurance Magazine and Journal of the Institute of Actuaries* VIII (1860): 301-10.
25. MEREU, J.A. "Annuity Values Directly from Makeham Constants," *TSA* XIV (1962): 269-86.
26. METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H., AND TELLER, E. "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics* 21 (1953): 1087-92.
27. MILLER, R.B., AND WICHERN, D.W. *Intermediate Business Statistics: Analysis of Variance, Regression, and Time Series*. New York: Holt, Rinehart and Winston, 1977.
28. MOORE, M.H. "Derivatives of the Annuity Function Assuming Makeham-Gompertz Mortality; and the N-Ages Method," *Communications in Statistics: Computation and Simulation* (1981): 359-67.
29. ROSS, S.M. *A Course in Simulation*. New York: Macmillan Publishing Company, 1990.
30. ROSS, S.M. *Introduction to Probability Models*. 4th ed. San Diego: Academic Press Inc., 1989.
31. SCOLLNIK, D.P.M. "A Bayesian Analysis of a Simultaneous Equations Model for Insurance Rate-making," *Insurance: Mathematics and Economics* 12 (1993): 265-86.
32. SCOLLNIK, D.P.M. "The Bayesian Analysis of Generalized Poisson Models for Claim Frequency Data Utilising Markov Chain Monte Carlo Methods," *ARCH* 1995.1, 339-56.
33. SMITH, A.F.M., AND ROBERTS, G.O. "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society Series B* 55 (1993): 3-23.

34. TANNER, M.A. *Tools for Statistical Inference, Methods for the Exploration of Posterior Distributions and Likelihood Functions*. 2nd ed. New York: Springer-Verlag, 1993.
35. TIERNEY, L. "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics* 22, no. 4 (1994): 1701–28.
36. WAKEFIELD, J.C., GELFAND, A.E., AND SMITH, A.F.M. "Efficient Generation of Random Variates via the Ratio-of-uniforms Method," *Statistics and Computing* 1 (1991): 129–33.
37. WILD, P., AND GILKS, W.R. "Adaptive Rejection Sampling from Log-concave Density Functions," *Applied Statistics* 42 (1993): 701–9.



## DISCUSSION OF PRECEDING PAPER

JACQUES F. CARRIERE:

Dr. Scollnik's paper presents MCMC-related methods to simulate random variables from densities that are nearly or exactly log-concave. The purpose of this discussion is to argue that these simulations can be done with simple and standard methods.

Let us suppose that we have generated a sequence of independent and identically distributed random variables, denoted as  $X_1, \dots, X_n$ , from a common cumulative distribution function (cdf), denoted as  $F(x)$ . It is well-known that this sequence can be used to construct an empirical function

$$\hat{F}(x) = \frac{1}{n} \sum_{k=1}^n 1(X_k \leq x),$$

which can be used to approximate any property of  $F(x)$ . The paper states that these "simulation methods allow a variety of previously inaccessible inferences to be made routinely and easily." To suggest that this paper solves problems that were "previously inaccessible" is too strong a statement, because these problems are easily solved with standard techniques. Moreover, it is not the simulation methods that allow these inferences to be made easily; rather, it is the empirical function  $\hat{F}(x)$ . For example, to approximate  $\bar{A}_{50}$ , we would use the formula

$$\int_0^\infty v^x d\hat{F}(x) = \frac{1}{n} \sum_{k=1}^n v^{X_k},$$

assuming that  $T(50) \sim F(x)$ . Again, the more complicated Numerical Example 4 is easily solved, as long as  $\hat{F}(x)$  is known.

The paper also states that "none of these references has considered simulating random variates from either Gompertz's or Makeham's survival distribution," possibly implying that these simulations are difficult to implement, thus requiring a sophisticated MCMC-related method. Actually, generating random variables from these distributions is quite simple. Let  $U$  be a random variable from a uniform distribution on the interval  $(0,1)$ . Also, let  $F^{-1}(u) = \inf\{x \in \mathcal{R} : F(x) \geq u\}$  be an inverse function of  $F(x)$ . It is well-known that if  $F(x)$  is a continuous function in  $x$ , then  $F^{-1}(U)$  is a random variable with a cdf equal to  $F(x)$ . The key to using this result is calculating the inverse function  $F^{-1}(u)$ . This is a trivial exercise for the Gompertz

distribution. For Makeham's law, no explicit expression exists for the inverse function, but the problem of finding the value  $X$  such that  $F(X)=U$  is easy. In this case the Newton-Raphson method is used, a technique that all actuaries know.

The paper also states that the MCMC-related methods "can be used to quickly and efficiently simulate random draws," possibly implying that these methods are quicker than others. As evidence, the paper states that 250,000 realizations from a Makeham distribution "took less than two minutes of real time" on a SUN Sparcstation LX. Using the GAUSS programming language and the Newton-Raphson method, it took me about six minutes to generate 250,000 observations from the Makeham distribution. These calculations were done on a 66 MHz Pentium processor, and so it is not clear to me that the MCMC-related methods are much faster.

In conclusion, the MCMC-related simulation methods are very interesting and they may be useful under certain situations, but I am not convinced that they are practical enough for simply simulating a Makeham random variable or any other univariate distribution that actuaries may use.

#### **GORDON E. KLEIN:**

Dr. Scollnik presents a method for simulating random draws from distributions such as Makeham's. The methods he presents are interesting, and they provide a nice addition to the acceptance-rejection method as covered in the syllabus for Examination 130 [18, p. 871–2]. (Note that my references are to Dr. Scollnik's bibliography.) In particular, the method as presented on the 130 Syllabus can be used only for random variables that are bounded on both sides. The rejection sampling method described by Dr. Scollnik overcomes this problem.

My criticisms of the paper can be stated as follows: (1) This is a method in search of an actuarial problem to solve. I can think of no real problem that is solved by this method, and this paper certainly does not provide one. The illustrations in the paper (particularly Numerical Example 4) are simply artificial problems designed to illustrate the method. I can think of no actuarial application for generating large numbers of random draws from Makeham's distribution. (2) Even if this were not the case, there are ways to simulate Makeham-distributed random variables that are more generally useful. I address the second criticism first.

### ***Other Methods for Simulating Draws from Makeham's Distribution***

What if one didn't have the methods of this paper and it were necessary to generate a large number of random draws for a future-lifetime random variable that had the Makeham distribution? One method that comes to mind is to invert the cdf of the future-lifetime random variable and to apply this to random draws from the uniform distribution. Actually, it is a little easier to do this with the survival distribution, of a life currently age ( $x$ ), instead of the cdf. For a life subject to Makeham's law of mortality, we have

$${}_t p_x = s^t g^{c^x(c^t-1)},$$

using the notation of [21], where

$$\ln s = -A,$$

and

$$\ln g = -\frac{B}{\ln c}.$$

Now, let's say that you have a random draw  $u$  from the uniform distribution on the unit interval. Setting  ${}_t p_x$  equal to  $u$  and solving for  $t$  will result in a random draw,  $t$ , from the Makeham-distributed random variable,  $T(x)$ .

The equation

$$u = {}_t p_x = s^t g^{c^x(c^t-1)}$$

cannot be solved explicitly for  $t$ , but this presents no real problem. Taking the natural logarithm of each side, we have:

$$\ln u = t \ln s + c^x(c^t - 1) \ln g,$$

which, in terms of the original parameters, is

$$\ln u = -At - \frac{Bc^x}{\ln c} (c^t - 1).$$

Moving all the terms to the left, we have a function whose root we need. Applying the Newton-Raphson Method, we find  $t$  as the limit of the quadratically convergent sequence (for an appropriate initial value)

$$t_{k+1} = t_k - \frac{At_k + \frac{Bc^x}{\ln c} (c^{t_k} - 1) + \ln u}{A + Bc^{x+t_k}}.$$

This routine required about seven minutes on a 486-66 computer using BASIC to generate 250,000 draws (and calculate the loss random variable for each draw). This compares with less than two minutes for Dr. Scollnik's routine using a SUN Sparcstation and a C-language routine. This does not seem to me to be a significant difference given the infrequency with which one needs to generate 250,000 random draws from a Makeham distribution.

Another method for generating random draws is to simply create a table of the cdf and to search it for each random draw from the uniform distribution on the unit interval. That is, create a table of  ${}_k|q_x$  for non-negative integral values of  $k$ , and then, for each random draw from the uniform distribution, find the integers that surround that number in the table. This method requires some type of interpolation between the surrounding integers (if you are interested in the exact time of death instead of the number of complete years lived) and an upper bound on the table (even though the Makeham distribution is unbounded). Neither of these problems is very serious. One advantage of this "table lookup" method is that it can be used for *any* distribution of mortality. I think that the method of Dr. Scollnik's paper is more complicated and less general than the method of this paragraph.

### ***A Solution in Search of an Actuarial Problem***

My other criticism is that I don't think that the methods of Dr. Scollnik's paper solve any actuarial problem. Numerical Example 4 purports to be a "more sophisticated problem" than can be handled by symbolic mathematics packages. This problem is a variation of one that is common on Examination 150 ("Actuarial Mathematics") and that is easily solved by hand. The random variable  $L$  is a continuous (except at the integers), monotonically decreasing function of the remaining-future-lifetime random variable,  $T(30)$ , which is subsequently referred to as  $T$ . Thus, if we find the number  $t^*$  such that  $\Pr[T < t^*] = 0.05$ , then the premium calculated as

$$\pi = \frac{10,000 v^{t^*}}{\ddot{a}_{\lfloor t^* \rfloor + 1}}$$

(where  $\lfloor t^* \rfloor$  indicates the greatest integer) will be such that

$$\Pr[L(\pi) > 0] = 0.05$$

(where we have indicated that the loss random variable is a function of the premium).

For the problem considered in Numerical Example 4,  $t^*$  is the solution of  ${}_tq_{30}=0.05$ . This is easily found to be 24.323001. Substituting this into the expression for the premium, we have

$$\pi = \frac{10,000 v^{24.323001}}{\ddot{a}_{\overline{25}|}} \doteq 143.7102.$$

This is exact to the number of places shown, but just to demonstrate, I generated 250,000 random draws of the loss random variable using this premium. Of these, 12,603 turned out negative, which is about 5.04%. (Note that this method does not require the generation of any random draws. This was done merely to parallel the "demonstration," actually also the estimation of the answer, in Dr. Scollnik's paper.)

I am interested in Dr. Scollnik's explanation of the discrepancy between this and his answer of 143.20. It appears to me that his method is much more complicated, *requiring* the generation of a huge number of random draws. Yet it gives an answer that is subject to a random error whose range is unbounded. That is, one cannot determine an interval in which the answer lies with probability 1 using Dr. Scollnik's method.

The second part of Numerical Example 4 likewise complicates the process of finding the distribution of the loss random variable 20 years after issue. For example, the 95th percentile of this random variable can easily be found by evaluating the random variable at the value of  $t^*$  where  ${}_tq_{50}=0.05$ . It is easily shown that  $t^*=10.58276$ , so that the 95th percentile of the loss random variable is

$${}_{20}L(10.58276) = 10,000 v^{10.58276} - 143.20 \ddot{a}_{\overline{11}|} \doteq 3,525.56,$$

(using Dr. Scollnik's premium for comparison). This is exact to the number of places shown. It compares with his "empirical value" of 3,529.356.

The estimation of the  $x$ -th percentile of a random variable using the Monte Carlo method (with a given degree of precision) requires a larger sample as  $x$  approaches 0 or 100. For example, to estimate the value of the loss random variable,  ${}_{20}L$ , that has a one-in-a-million chance of not being exceeded, we could use an estimate between the 10th and 11th order statistics out of a sample of 10 million. This estimate would be unbiased, but its variance would be large. A larger sample would reduce the variance of the estimate.

(The sample of 250,000 from the paper would be useless for this problem.) A better method, when it can be done, is to approach the problem directly. In this particular case, we can find that

$${}_{59.002996}P_{50} = 0.000001,$$

so that the 0.0001th percentile of the loss random variable is

$${}_{20}L(59.002996) = 10,000 v^{59.002996} - 143.20 \ddot{a}_{\overline{60}|} = -1,885.536.$$

### **Conclusion**

Despite my two criticisms of this paper, I find the methods interesting. It will be interesting to see what actuarial problems they can be applied to. I look forward to seeing Dr. Skollnik's work applying the methods to aggregate claims distributions.

### **JEFFREY S. PAI:**

Markov chain Monte Carlo methods have become very popular in recent years as a way of generating a sample from a complicated probability distribution. Dr. Scollnik has applied some of these techniques successfully on Makeham's distribution and other distributions. The author has also shown how to easily estimate the distribution of the present value of the benefit as well as the distribution of the present value of the annuity payments using sampling methods. The estimations can be made as accurate as desired by increasing the length of the MCMC simulations.

The problem of the computer generation of random variates with a given force of mortality can be done easily by applying the connection between the cumulative distribution function,  $F$ , and the cumulative force of mortality,  $M$ :

$$F(t) = 1 - \exp[-M(t)], \quad M(t) = \int_0^t \mu(s) ds.$$

For generating a random variate with cumulative mortality  $M$ , it suffices to invert an exponential random variate  $E$  [Scollnik's ref. 10, p. 260]. If the solution  $t$  of  $M(t)=E$  is not explicitly known, such as the Makeham's law, we can incorporate Newton-Raphson iterations, the thinning algorithm, or the composition method (see Pai [2]).

The Metropolis algorithm is a method of constructing a reversible Markov transition kernel with a specified invariant distribution. I would like to

suggest the applications of the Metropolis algorithm in the optimization procedure and the Bayesian analysis.

The Metropolis algorithm can be used in the simulated annealing method [1], which has attracted significant attention in finding a desired global extremum among many local extrema. Programs coded in C and Fortran are available from Press et al. [3].

Bayesian inference proceeds by obtaining marginal posterior distributions of the components of the model parameters as well as features of these distributions. For instance, suppose we use the data of size 200, say  $T=(t_1, \dots, t_{200})'$ , sampled from the Makeham's distribution with

$$\mu(t) = A + BC^{x+t}, \quad x = 50,$$

where

$$A = 0.001$$

$$B = 0.0000070848535$$

$$C = 1.1194379.$$

If we assume the priors of these parameters are

$$\pi(A) = 1/0.003, \quad 0 < A < 0.003,$$

$$\pi(B) = 1/0.001, \quad 0 < B < 0.001,$$

$$\pi(C) = 1/0.2, \quad 1 < C < 1.2,$$

then the posterior is proportional to the product of the priors and the likelihood function:

$$\pi(A, B, C|T) \propto \prod_{i=1}^n (A + BC^{x+t_i}) \exp \left[ -A \sum_{i=1}^n t_i - \frac{BC^x}{\ln C} \sum_{i=1}^n (C^{t_i} - 1) \right].$$

Using the uniform priors stated above, this is essentially the likelihood approach. The ARMS-within-Gibbs method or the Metropolis-within-Gibbs method can be utilized to sample from the full conditional densities. The results of using the Metropolis-within-Gibbs method and the estimated marginal posterior densities are shown in Figure 1 and Figure 2 with 2000 iterations.

As Dr. Scollnik stated, many problems in actuarial science can be expected to benefit from an application of MCMC methods. The attractiveness of the sampling methods is their conceptual simplicity and ease of implementation

FIGURE 1  
MCMC OUTPUT FROM THE METROPOLIS-WITHIN-GIBBS ALGORITHM

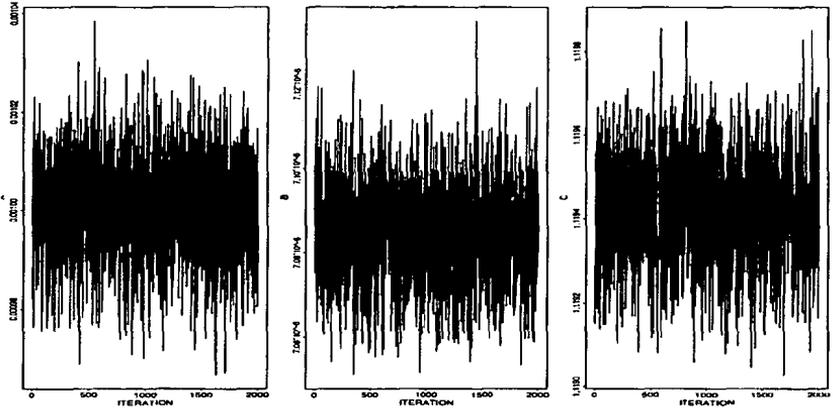
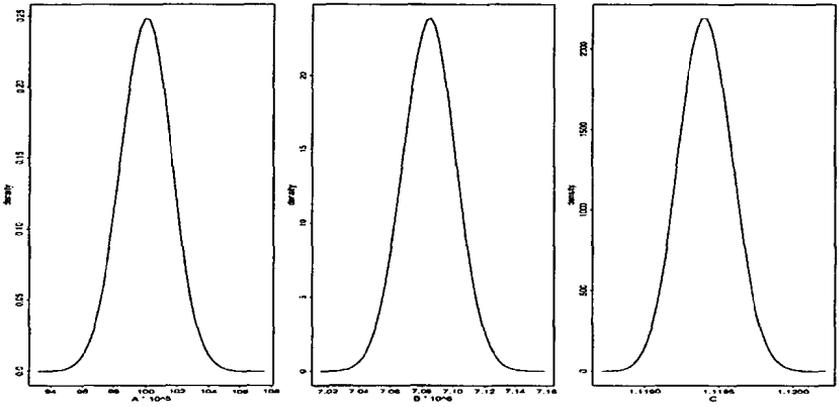


FIGURE 2  
ESTIMATED MARGINAL POSTERIOR DENSITIES



for users with available computing resources but without numerical analytic expertise. I believe that actuarial students will soon gain more insight and understanding in the classroom by using MCMC methods.

## REFERENCES

1. KIRKPATRICK, S. "Optimization by Simulated Annealing: Quantitative Studies," *Journal of Statistical Physics* 34 (1984): 975-86.
2. PAI, J.S. *Generating Random Variates with a Given Force of Mortality and Finding a Suitable Force of Mortality by Theoretical Quantile-Quantile Plots*. Manual. University of Manitoba, 1996.
3. PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T., AND FLANNERY, B.P. *Numerical Recipes*. Cambridge: Cambridge University Press, 1992.

## JOHN A. MEREU:

Dr. Scollnik has written an interesting paper showing how some new techniques for generating random variables can be used to efficiently solve problems in which mortality follows Makeham's law.

This discussion presents an alternative to Monte Carlo methods referred to as a *grid* approach. The *grid* approach also takes advantage of the computational power of today's personal computers. It calls for the generation of values of a probability density function (pdf) over a broad and dense spectrum of the independent variable.

Let  $T$  be the future lifetime random variable for a person aged  $x$ . The expected value of  $T$  is given by the integral

$$E[T] = \int_0^{\infty} t {}_t p_x \mu_{x+t} dt.$$

The integral is the limit of the following sum as  $n$  approaches infinity and  $h$  approaches zero.

$$E[T] = \lim_{\substack{n \rightarrow \infty \\ h \rightarrow 0}} \sum_{k=0}^n \int_{Kh}^{(K+1)h} t {}_t p_x \mu_{x+t} dt \doteq \sum_{K=0}^n Kh {}_{Kh} p_x \mu_{x+Kh}.$$

Setting  $n$  to 6000 and  $h$  to 0.01, I computed the following values given in Table 2 of the paper with an APL program using 1.87 seconds of CPU time:

Quantity of Interest <i>T</i> (50)	Numerical Approximation	Empirical Value (Monte Carlo)	Empirical Value (Grid Method)
Mean	30.81125	30.79607	30.80711
Variance	108.87118	108.80472	108.84535
Skewness	-0.60911	-0.61111	-0.60872
Kurtosis	2.96685	2.96872	2.96621
95th Percentile	45.39895	45.35621	45.39

Let *Z* be the discounted benefit random variable for 1 payable at the moment of death. The expected value of *Z* is given either of the two integrals, one based on the pdf of *Z* and the other based on the pdf of *T*. The grid method can be used to approximate the quantities of interest for either integral. The results are approximately the same and also conform to those given by the Dr. Scollnik in Table 2. If the pdf of *T* is used, the histograms require a mapping of the pdf values to histogram segments of *Z* and a totaling of the probabilities by segment. If the pdf of *Z* is used, the mapping is more straightforward, but some calculus is required to determine the pdf of *Z*, given by

$$\text{pdf}(Z) = \frac{1}{\delta Z} \left\{ (A + Bc^{x+s}) \exp \left[ -\frac{Bc^x}{\ln c} (c^s - 1) \right] \exp(-As) \right\}$$

where  $s = -(1/\delta) \ln Z$ . Similar comments apply to *Y*, the discounted benefit random variable for a continuous annuity to the moment of death.

In Table 3 and Figure 7 Dr. Scollnik provides an analysis of the semi-continuous loss function random variable, *L*, assuming the premium charged is large enough to reduce the probability of positive loss to 5%. Because it does not seem feasible to derive a probability density function for *L*, the *grid* method was applied using the pdf of *T*. The results obtained were consistent with those published by Dr. Scollnik. The premium itself can be obtained by solving  ${}_t p_{30} = 0.95$  using the *grid* method to get *t* and then solving  $L(t) = 0$  to find the premium. I found the premium to be \$143.76.

(AUTHOR'S REVIEW OF DISCUSSIONS)

DAVID P.M. SCOLLNIK:

I thank the discussants for their valuable comments and for the interest that they have shown in this paper.

One goal of this paper was to introduce *Transactions* readers to some of the new and very generally applicable techniques for random number generation and stochastic simulation that have attracted much interest in recent statistical literature. ARS was introduced as an example of an adaptive random variate generation method, and the MH and ARMS algorithms served as two examples of MCMC methods for generating sequences of dependent random variates from a target distribution of interest. These techniques complement, but do not entirely replace, the few tried and true simulation methods currently appearing in the Associateship syllabus.

Another goal of the paper was to demonstrate the application of these MCMC methods in a simple and easily understood actuarial context and to illustrate some of the characteristics of these methods. However, emphasizing the univariate Makeham distributional setting as I did may have obscured the fact that the methods have a much wider field of application. Klein goes so far as to suggest that in this paper I have described a "solution looking for an actuarial problem." Notwithstanding Klein's comment, at the start of Section 3 I did supply a short list of references to a number of problems in actuarial settings with MCMC-driven solutions that have recently appeared in the literature (that is, [4], [5], [22], [31], [32] in the paper). This may be an opportune time to add to this list.

Towards this end, note that: Rosenberg [4] used MCMC to implement the Bayesian analysis of a hierarchical model for the rate of nonacceptable inpatient hospital utilization; Scollnik [5] considers an application of the Gibbs sampler to three hierarchical credibility models for classification rate making and involves the Bayesian prediction of frequency counts in workers compensation insurance; Shephard and Pitt [6] apply MCMC methods to a number of problems arising in the context of parameter-driven exponential family models, notably including the estimation of actuarial death rates and the estimation of the parameters appearing in several stochastic volatility models for financial returns. Professor Pai describes several other applications of MCMC methods in his discussion of my paper. MCMC is certainly not the only tool for stochastic simulation present in the modern actuary's toolbox, but it is another tool that is now available and is also one that more actuaries should be made aware of.

Carriere's first comment addresses the fact that the expected value of a function of a random variable can be estimated by averaging that function with respect to the empirical distribution for an independent sample from the distribution of the random variable in question. This is simply a restatement of my own comments in the second paragraph of Section 3-D regarding

the ergodic average of an  $f$ -integrable function  $h(x)$ , in the less general and trivial instance that  $g(x|z)=f(x)$ . When this is the case, it is obvious that every draw is independent and no draws are rejected, thus yielding the context of Carriere's observation.

Both Carriere and Klein suggest that a brute force method of random variate generation, direct inversion either of the cumulative distribution function (cdf) or of the survival function using the Newton-Raphson method, can be used to generate draws of the Makeham-distributed future lifetime random variable  $T(x)$ . Whereas I agree that this is technically the case, it must be pointed out that the efficiency of this brute force method is extremely dependent upon the choice of the Makeham parameters, the assumed age  $x$  of the life in question, and the selection of the starting value plugged into the Newton-Raphson algorithm. In order to partially illustrate this fact, I implemented the Newton-Raphson algorithm exactly as described by Klein, taking the Makeham parameters equal to the values appearing in my Numerical Example 3 and arbitrarily setting  $t_0$  equal to 75. The algorithm was allowed to iterate until the absolute difference between  $t_k$  and  $t_{k+1}$  was no larger than a modest 0.01. In this manner, I generated 100 draws of  $T(x)$  for each of six different values of  $x$  (that is,  $x=0, 10, 20, 30, 40,$  and  $50$ ). The observed average number of iterations required to generate a single draw of  $T(x)$  for the six different values of  $x$  is tabulated below. Examining these numbers, it is evident that direct inversion of the cdf using Newton-Raphson's method is not always particularly efficient. In fact, for a newborn life ( $x=0$ ) and under the conditions described above, it required in excess of 200 iterations to generate a single realization of  $T(0)$  on seven of the 100 attempts. Please note that these simulation results are only illustrative and are certainly not definitive.

Age $x$	Average Number of Iterations
0	67.95
10	34.30
20	14.56
30	7.08
40	3.91
50	3.62

Klein also suggests that random draws from an arbitrary univariate distribution can be generated by simply creating a table of the cdf and then searching it for each random draw from the uniform distribution on the unit

interval. However, and as noted by Klein, this method requires some type of interpolation and the table must be bounded above and below, even though the distribution itself may be unbounded. Consequently, the sampled values are only approximately from the distribution desired. In Section 2.3.6 of [35], Tierney describes how the MH algorithm can be used to correct for this failing of Klein's suggested "table lookup and interpolation" procedure, so that the values generated actually form a dependent sequence from the exact distribution of interest. This is an excellent example of how a MCMC method can be used to augment a traditional random variate generation procedure. For the convenience of the reader, I have summarized the MCMC method described in [35] immediately below using the terminology and notation previously adopted in Section 3-D of my paper.

Assume that we want to generate a dependent sequence of random draws from a distribution with a density proportional to  $f(x)$  on the real line. Candidate draws will be generated from a proposed distribution by means of a two-step procedure, requiring the availability of another distribution  $h(x)$  on the real line from which random draws are easily accomplished. To begin with, let  $x_1^*, \dots, x_m^*$  be a fixed set of points, and let  $X_i$  be the value of the draw generated in the  $i$ -th iteration of the MH sampling algorithm. At the start of the next iteration, select a point  $X^*$  from  $x_1^*, \dots, x_m^*$  according to a discrete distribution that is proportional to the density values  $f(x_1^*), \dots, f(x_m^*)$  at these points. Then generate a value  $Z$  from  $h(x)$  and add this number to  $X^*$  in order to obtain  $X = X^* + Z$ . This process defines a proposal distribution with density

$$g(x|X_i) = \frac{\sum_{j=1}^m f(x_j^*)h(x - x_j^*)}{\sum_{j=1}^m f(x_j^*)}.$$

According to the MH algorithm, the value of  $X$  is accepted with a probability equal to

$$\min \left[ 1, \frac{f(X) \sum_{j=1}^m f(x_j^*)h(X_i - x_j^*)}{f(X_i) \sum_{j=1}^m f(x_j^*)h(X - x_j^*)} \right]. \quad (1)$$

In the case that  $X$  is accepted,  $X_{i+1}$  is set equal to  $X$ ; otherwise,  $X_{i+1}$  is set equal to  $X_i$ . This algorithm is iterated until the required number of random variates has been generated.

This algorithm was very simple to program using S-Plus, and I used it to simulate 500 dependent values of the Makeham-distributed future lifetime random variable  $T(50)$  for the life in Numerical Example 3. I took  $m=100$ , and set the fixed points  $t_1^*, \dots, t_{100}^*$  equal to the integers from 1 to 100. For  $h(t)$ , I simply used the density for a normal distribution with a mean of 0 and a standard deviation of 10. Although this permits an occasional candidate draw from the proposed distribution to take on a negative value, such draws are never accepted because the MH acceptance probability (1) is always equal to 0 in these instances by construction. The simulation results were as follows: the observed staying rate (Section 3-E) for this example was 25.4%; the 500 simulated values of  $T(50)$  had a sample mean equal to 30.90; and the empirical histogram of the simulated values appeared to be in agreement with the exact density function for  $T(50)$  appearing in Figure 6(a). Unlike the "table lookup and interpolation" procedure suggested by Klein, the MCMC method I described is exact, requires no artificial bounds to be imposed, and requires no interpolation.

Klein describes a simple and elegant exact alternative to my simulation-driven analysis in Numerical Examples 3 and 4, and it is one that I should have presented for comparison. Concerning the value of  $\pi$ , the discrepancy between Klein's answer and my own is almost certainly due to a combination of simulation error and numerical rounding error. In any case, the relative error inherent in my simulation-based answer is approximately equal to  $(143.2 - 143.7102)/143.7102 \approx -0.0035502$ , which is not excessive. To correct a minor technical point made by Klein, note that the range of the random error inherent in my estimate of  $\pi$  is not unbounded in the context of Numerical Example 3, because the value of the estimate is restricted to the interval between 0 and 10,000 by construction.

Professor Pai describes how MCMC methods can be used to implement Bayesian posterior parameter estimation when a vector of survival times,  $\mathbf{T}$ , has been observed from a Makeham distribution with parameters  $A$ ,  $B$ , and  $C$  (compare estimation of the Gompertz parameters in [1]). It is a simple matter to extend Pai's discussion to implement Bayesian predictive inference with respect to the future lifetime  $T(y)$  for a new life age  $y$  as well. Let the values  $A_i$ ,  $B_i$ , and  $C_i$ , for  $i=1, \dots, N$ , be the dependent sequence simulated from the posterior distribution  $p(A, B, C|\mathbf{T})$  using the method described by Pai. According to the discussion in the second paragraph of Section 3-D concerning the ergodic average of a function, we have

$$\begin{aligned}
 f(T(y)) &= t|T \\
 &= \iiint f(T(y) = t|A, B, C) p(A, B, C|T) dA dB dC \\
 &\approx \frac{1}{N} \sum_{i=1}^N f(T(y) = t|A_i, B_i, C_i).
 \end{aligned}$$

The last expression is an estimate of the predictive density at the point  $t$ . A sample from the predictive distribution can be obtained by simulating a single draw of  $T(y)$  from each Makeham distribution with density  $f(T(y)=t|A_i, B_i, C_i)$ , for  $i=1, \dots, N$ .

The recent text by Gelman, Carlin, Stern, and Rubin [3] provides a comprehensive treatment of the statistical analysis of data from a Bayesian perspective. In particular, a very readable account of posterior integration and Markov chain simulation strategies is given in Chapters 10 and 11.

In his discussion, Professor Pai also mentions that MCMC methods like MH and ARMS can be used in conjunction with Gibbs sampling, thus yielding the so-called MH-within-Gibbs and ARMS-within-Gibbs variants of MCMC. Although the references I cited for ARMS discuss these variants, it appears that I failed to mention this fact explicitly in my paper. See [14] and [15] in the original paper or Section 11.3 of [3] for details. I am thankful to Dr. Pai for bringing this omission to my attention.

I was a student of Professor Mereu's at the University of Western Ontario in 1986–87, and so it was a distinct pleasure to receive his discussion of my paper. His "grid" approach to integration describes a version of numerical quadrature, akin to some of these covered in the Course 135 "Numerical Methods" syllabus. Although these methods are competitive for evaluating univariate or low-dimensional integrals, they may not be quite as successful in high dimensions. Bayesian analyses of actuarial data typically require the evaluation of high-dimensional integrals to obtain posterior and predictive densities, means, variances, and so forth. As indicated at the start of Section 3 and earlier in this discussion, MCMC methods have proven to be very useful in these contexts.

Since my paper was written, Evans and Swartz [2] have developed a class of adaptive rejection algorithms for generating independent random variates for a wide assortment of families of densities. These generators depend on the concavity structure of a transformation of the density. Makeham's distribution is included as one of the examples. Interested readers can obtain a copy of this technical report from the authors.

## REFERENCES

1. AMANDA, M.M., DALPATADU, R.J., AND SINGH, A.K. "Estimating the Parameters of the Force of Mortality in Actuarial Studies," *ARCH* 1993.1:129–141.
2. EVANS, M., AND SWARTZ, T. "Random Variable Generation Using Concavity Properties of Transformed Densities," *Technical Report No. 9606*, Department of Statistics, University of Toronto, 1996.
3. GELMAN, A., CARLIN, J.B., STERN, H.S., AND RUBIN, D.B. *Bayesian Data Analysis*. London: Chapman and Hall, 1995.
4. ROSENBERG, M. "A Hierarchical Bayesian Model of the Rate of Non-Acceptable In-Patient Hospital Utilization," doctoral thesis, University of Michigan, 1994.
5. SCOLLNIK, D.P.M. "An Introduction to Markov Chain Monte Carlo Methods and Their Actuarial Applications," *Proceedings of the Casualty Actuarial Society*, in press.
6. SHEPHARD, N., AND PITT, M.K. "Parameter-Driven Exponential Family Models," Nuffield College, Oxford, 1995. Available on the World Wide Web via the MCMC Preprint Service located at <http://www.statslab.cam.ac.uk/>.