Evaluating risk in insurance and finance by stochastic simulation

Erik Bølviken
University of Oslo and the Norwegian Computing Centre

December 4, 2000

Contents

1 Introduction

Part 1 Method
1 The general idea
2 Dynamic models
3 Building models

Part 2 Applications
4 General insurance
5 Life insurance
6 Financial portfofoles
7 Simulating business
8 Option pricing
9 References

Appendix

Distributions

Summary

The enormous power of modern computers has made stochastic simulation a practical tool for evaluating actuarial and financial risk. An elementary, state-of-the-art introduction is presented, geared towards skills necessary to solve practical problems. It is argued that the computational method influences mathematical modeling and empirical estimation; how is an integral part of the discussion. Small algorithms is pseudo-code are given throughout. It is shown by examples how simple modifications of existing programs can produce solutions of different (but related) problems, and how software that has been developed can be combined to handle problems of increasing complexity.

1 Introduction

Stochastic simulation is well established as a tool for evaluating economic risk. Yet I doubt whether all practitioners are aware of the possibilities offered. Simulation (or, interchangeably sampling or Monte Carlo) has often been portrayed as a mean of last resort to solve tricky computational problems other methods are unable to cope with. With the steady advance in computational power (and software) this view has today become obsolete. Simulation is slow compared to other ways of doing numerical integration, but what is the significance of a computation taking 0.5 seconds instead of 0.01 seconds? Of course, if the computations are to be repeated a large number of times (say thousands or ten thousands) it does matter, but if not, other features are the important ones, notably generality, versatility and speed of implementation.

The Monte Carlo method is general, because it can be used with almost all types of problems. This saves the amount of technical know-how needed and opens for general, unified ways of organizing computations. The method is versatile in the sense that an existing program with minor modifications can be used to solve a related problem, and it also permits the utilization of hierarchies to handle increasing levels of model complexity. Simulation programs in actuarial and financial risk can often be made quite small and is
easy to develop once some basic skill is mastered.

A course in stochastic simulation must not fall into the trap of dealing solely with sampling techniques. Of course one needs examples and cases to run methods on, but there is another, deeper point in that the access to simple, computational power also influences the modeling itself and the way available data is utilized. No longer, or at least to a much less extent, should real problems be simplified to cope with computational obstacles, and one could (and should) be much more conscious than traditional teaching dictates in keeping purely mathematical assumptions to a minimum. Actually, working with more general models may also simplify the fitting process. Several examples will be given.

This tutorial introduction to Monte Carlo evaluation of risk will confronts all aspects mentioned. The presentation is organized hierarchically, starting at the top level and working downwards, ending with applications. The general issues of sampling are dealt with in Section 2, for example how simulations are transferred to estimates of means or Values-at-Risk, how simulation uncertainty is judged, how the number of simulations are determined, how sampling routines form hierarchies and what programming platforms are suitable.

The next level of the hierarchy is the dynamics and how it is described mathematically. Typically financial risk of a future period is defined by the outcome of a random time series. These could be accounts fluctuating according to random input of cost and income or stochastic models of stock and securities. An important model in practice is the random walk. We shall use this as a tool to discuss the simulation approach. The stock market requires the vectorial version with many processes running in parallel, and the extension to autoregressive processes will also be considered to portray oscillations in interest rates.

The lowest level of the hierarchy is the family of distributions describing single variables. The most important ones for financial and actuarial risk will be summarized in an appendix, including simple algorithms for sampling. Although software packages like SAS and Splus have built-in routines, there is, at times, convenient to escape limitations set by designers of these programs and operate completely independently. Emphasis here is on sampling techniques that are as easy as possible to implement.

Another, equally important task is how historical data is utilized. This is discussed in Section 4. One of the contentions of the present article is that sampling based implementations yield flexibility in this area and that the work of the analyst may actually be shortened and simplified at the expense of more computation. The steady decrease of computational cost makes this is a good bargain.

The second part is on applications. A number of cases from insurance and finance is discussed to give the reader a broad picture as to how simulation can be put to work.

2 The general idea

2.1 Why sampling?

Let $X$ be a random variable measuring, in some sense, economic risk, and suppose $X$ can be described by a probability density function $f(x)$. If $f(x)$ is known and in simple form, it is easy to compute the relevant financial quantities associated with it, for example its expectation or its percentiles (which defines Value-at-Risk or VaR). However, in practice the density is often far from simple, sometimes almost intractable to derive and compute. Consider the following two examples:

**Example 1 Aggregated claims** In insurance many numerical evaluations
concern random sums of the form
\[ X = \sum_{j=1}^{N} Z_j, \]  
(2.1)

where \( Z_1, Z_2, \ldots \) is a sequence of independently distributed random variables. The distribution of \( Z_j \) will usually vary with \( j \) to a greater or lesser extent, and we shall consider such situations later, but for now it is assumed that all individual claims \( Z_j \) follow the same distribution.

The number of terms \( N \) can be the number of policies in the portfolio (and hence a known, deterministic quantity), but in a situation of identical policies (as we are presently assuming) it may be more natural to define \( N \) as the number of claims within a time period, and hence as a random variable. The standard model for \( N \) is the Poisson distribution, and it also common to let \( N \) be stochastically independent of the sequence \( Z_1, Z_2, \ldots \).

In spite of all these simplifications the resulting distribution of \( X \) is cumbersome to find, unless we are dealing with distributions like the gamma or the normal inverse Gaussian density (NIG) which are closed under convolutions; see appendix. Simulation, however, is easy.

**Example 2 Price of an option** Let \( Y_1, \ldots, Y_J \) be normally distributed, correlated random variables. If we are dealing with a so-called European call option with strike price \( K \) of a basket portfolio in the Black and Scholes framework (see Section 9 below), we are interested in the risk variable
\[ X = \max(V - K, 0), \]  
(2.2)

where
\[ V = \sum_{j=1}^{J} w_j \exp(Y_j), \]  
(2.3)

and where \( w_1, \ldots, w_J \) are given weights assigned the \( r \) securities in the basket. Here \( V \) is the value of the portfolio at the date the option can be exercised and \( K \) the agreed price if the buyer chooses to buy. The option price is then \( E_Q(X) \), under a carefully constructed normal model \( Q \) for \( Y_1, \ldots, Y_J \) (see Section 9). Again we cannot simply find the density function \( f(x) \), unless we put in considerable effort and again simulation rescues us; how this is detailed in 2.4 below.

**2.2 Using the samples**

Suppose we somehow, in Examples 1 or 2 or in some other situation, have been able to draw \( m \) samples \( X_1^*, \ldots, X_m^* \) of \( X \). Throughout Monte Carlo variables obtained in the computer will be marked with a * to distinguish them from other variables. The natural estimate of the mean \( \mu = E(X) \) from the simulations is then
\[ \bar{X}^* = m^{-1} \sum_{j=1}^{m} X_j^*, \]  
(2.4)

just as ordinary means of real data are routinely used as estimates of population means in elementary statistics. Of course \( \bar{X}^* \) will not be exactly equal to \( E(X) \), but it will not be far away, and we can make it as close as we want by letting \( m \to \infty \).

How does one tell whether a selected value of \( m \) is large enough? By using statistics! The standard error of a mean is the well-known \( \sigma/\sqrt{m} \), where \( \sigma \) is the standard deviation of the simulations. Of course we do not know \( \sigma \) any more than we know \( \mu \), but again elementary results from statistics help us along. Simply compute
\[ s^* = \{(m - 1)^{-1} \sum_{j=1}^{m} (X_j^* - \bar{X}^*)^2\}^{1/2} \]
as an estimate of \( \sigma \) and use \( s^*/\sqrt{m} \) as an assessment of the simulation standard error of \( \bar{X}^* \). The normal region of uncertainty of (2.4) is two or three times this estimate. If you are too lazy to do this, a useful alternative is to replicate the whole experiment with the same \( m \) a few times, and judge whether
the observed variation is satisfactory small; see below.

The estimate (2.4) will be used with option pricing in Example 2. In the other example a simple formula is available for $\mu^1$, but much interest also involves worst case scenarios, i.e., upper percentiles or Value-at-Risk, for which simple formulas do not exist for aggregated claims. Let $x_\epsilon$ be the upper $\epsilon$ percentile of the distribution of $X$. To compute this quantity from the random samples, we must invoke sorting and rank the simulations in descending order as $X^*_1 \geq X^*_2 \geq \ldots \geq X^*_m$. Then

$$x^*_\epsilon = X^*_\lfloor \epsilon m \rfloor,$$  \hspace{1cm} (2.5)

ignoring the slight inaccuracy that $m\epsilon$ may not be exactly an integer.

The simulation (or Monte Carlo) variation of $x^*_\epsilon$ can be judged by repeating the computations. An analytical formula is also available, although less elementary than in Example 1 and only justified as a large-sample approximation as $m \to \infty$. However, large $m$ is precisely what is likely in practice, so it seems worthwhile to present the formula.

Indeed, it can be proved that the standard error of (2.5) is approximately

$$\gamma^{-1}(\epsilon(1-\epsilon)/m)^{1/2}, \quad \gamma = f(x_\epsilon).$$

Here $\gamma$ is unknown and must be estimated from the Monte Carlo samples.

There is, to this end, an enormous body of techniques in the literature on statistical density estimation; see, for example, Scott (1992). A standard method is based on Gaussian kernels. A smoothing parameter $h > 0$ is selected, and the estimate of $\gamma$ becomes

$$\gamma^* = (hm)^{-1} \sum_{j=1}^{m} \varphi\{(x^*_\epsilon - X^*_j)/h\} \quad (2.6)$$

where $\varphi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ is the standardized Gaussian density. The estimate of the standard error of $x^*_\epsilon$ is then

$$\gamma^{*-1}(\epsilon(1-\epsilon)/m)^{1/2}.$$

---

<table>
<thead>
<tr>
<th>$m$</th>
<th>Repeated experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>16.5  12.6  16.9  15.9  10.6</td>
</tr>
<tr>
<td>1000</td>
<td>14.1  13.6  14.3  14.0  13.3</td>
</tr>
</tbody>
</table>

Table 1: Value-at-Risk (5%) for Example 1.

The choice of $h$ in (2.6) has received much attention in specialist literature. A pragmatic choice for values of $m$ likely to arise in practice could be $h = 0.1$ or 0.2.

2.3 Examples

Suppose it is from experience known that $N$ in Example 1 follows a Poisson model with mean claim frequency $\lambda = 5$ and that each claim is Pareto distributed according to the density $\alpha/(1+z)^{1+\alpha}$ for $z > 0$ with parameter $\alpha = 2$. The money unit does not matter, but let us say that it is one million US$. For an outline of the Pareto distribution, see A.9.

The expectation of (2.1) is $E(X) = \lambda/(\alpha - 1)$ (which equals 5 under the parameters selected), but the fat tails of the Pareto distribution makes the Value-at-Risk much larger, as shown in Table 1 for $\alpha = 5\%$ (5 replicates). The estimates for $m = 100$ simulations are too variable, but $m = 1000$ seems sufficient for what is going to be used merely a rough guideline for the amount of money needed to be reserved. The estimate of simulation error obtained from the formula in 2.2 above was 1.3 for $m = 100$ and 0.7 for $m = 1000$. The former is a clear underestimate, which means that the large sample approximation does not work for $m$ this small.

The experiment for Example 2 was run with two correlated securities with means 0 and standard deviation (or volatility) 0.5 and 1.0. The correlation was 0.5 and the basket gave equal weight to both securities, i.e. $w_1 = w_2 = 1/2$. Strike price was $K = 1.274$. The results for $m = 1000$ and $m = 10000$ simulations are shown in Table 2 (four replications).

Suppose the money unit is million US$. The variation for $m = 1000$ is then uncomfortably large, but $m = 10000$ could, per-

---

1It is proved in elementary textbooks that $\mu = E(N)E(Z_1)$ in Example 1.
modify Algorithm 2.1. appropriately insert between lines 4 and 5 the two command lines
\[
Z^* \leftarrow \min(Z^*, l_u)
\]
\[
Z^* \leftarrow 0.5 \max(Z^* - l_t, 0).
\]
Now only the values affecting the re-insurance are incremented in line 5. This technique of hierarchically adding complexity to an existing program is extremely important. We shall meet it again repeatedly.

The second example is different in that the vector \((Y_1, \ldots, Y_J)\) in (2.3) is a dependent Gaussian vector, so that a subprogram generating such vectors must be available; see A.5. On that assumption, the sampling of \(X\) in (2.2) is organized as follows:

**Algorithm 2.2 Example 2**
1. Draw \(Y_1^*, \ldots, Y_J^* \sim\) multi-normal.
2. \(V^* \leftarrow 0\).
3. Repeat for \(j = 1, \ldots, J\)
4. \(V^* \leftarrow V^* + w_j \exp(Y_j^*)\)
5. \(X^* \leftarrow \max(V^* - K, 0)\).
6. Return \(X^*\).

Again the algorithm is to be repeated to generate many samples of \(X\).

### 2.5 Speed: The name of the game

Computational speed is, and will always remain, of primary importance for Monte Carlo as a methodology. Although there are many simple problems that can be solved without speed presenting any difficulty at all and that the set of those problems expand consistently as computer power grows, it is easy to write down obstacles it would take a week ( or a year!) to overcome.

What are the determining factors? Apart from the problem itself, there is the computer, the software and the implementation. It would scarcely have been worthwhile to recommend sampling as a versatile technique of broad applicability, if it had required an expensive, advanced computer, but nowa-
days standard computers found in any business office is powerful enough. The software matters a great deal. Excel, for example, is definitely too slow, but it is possible to carry out part of the computations outside (in C, for example) and interface with Excel. Splus and Matlab are feasible, but do not expect overwhelming speed. SAS is better, but the best way is C or Fortran programs, if you know how to use those.

It is also possible to enhance speed by statistical tricks implemented during the coding of the program. One way is to utilize that it is advantageous to force simulations to be negatively correlated. This is most easily understood in terms of the Monte Carlo mean (2.4). Suppose, for simplicity that \( m = 2 \). Then

\[
\text{var}(\bar{X}) = \sigma^2(1 + \tau)/2
\]

where \( \tau = \text{corr}(X_1^*, X_2^*) \), and the variance is reduced if \( \tau < 0 \).

One way to obtain negative correlation is to note that if \( U \) is uniform \((0, 1)\), then \( V = 1 - U \) is as well, and negatively correlated with \( U \). All Monte Carlo drawings can be reduced to operations on a set \( U_1^*, U_2^*, \ldots \) of uniform random variables. One way to generate negative correlation in the output is then to define \( X_j^* = 1 - U_j^* \), \( j = 1, 2, \ldots \), and obtain simulations of \( X \) as pairs, one from the \( U \)-sequence the other from that from \( V \).

The method is more complicated to implement, especially since the required number of uniforms may vary from one simulation to another. Even more sophisticated techniques are importance sampling (Ripely, 1987) and quasi Monte Carlo algorithms (Press et al 1986).

3 Dynamic models

Nearly all actuarial and financial computations involve time, at least indirectly. The next step in our course in simulations is to introduce simple stochastic processes as tools to describe how the account of an insurance portfolio or the price of some stock evolve in time. Such a random process will be denoted \( \{Y_k\} \), where \( k = 1, 2, \ldots, n \) are the values of the process at certain (discrete) time points \( t_k = k\delta \). The time increment \( \delta \) may be everything from seconds to years. In what follows the indexing variable \( k \) will not necessarily be time at all. The variable \( X \) of the preceding section will often be some function of a stochastic process \( \{Y_k\} \), for example its terminal value \( Y_n \).

3.1 Random walk

A very useful model in insurance and finance is the random walk, defined though

\[
Y_k = Y_{k-1} + \varepsilon_k^*, \quad (3.1)
\]

where \( \varepsilon_1, \varepsilon_2, \ldots \) are independently distributed random variables. Usually in probabilistic literature it is also required that the sequence \( \{\varepsilon_k\} \) consists of identically distributed variables, but we are not assuming this. An algorithm for simulating (3.1) runs as follows:

Algorithm 3.1 The random walk

0. Input \( Y_0^* \).
1. Repeat for \( k = 1, 2, \ldots, n \)
2. Draw \( \varepsilon_k^* \) \( \% \) Problem-dependent.
3. \( Y_k^* = Y_{k-1}^* + \varepsilon_k^* \).

After having specified the starting value \( Y_0^* \), the simulation proceeds by sampling the random terms in (3.1) successively. Note that their distribution may depend on \( k \). If the terminal value \( Y_n^* \) is the one of interest, then \( X^* = Y_n^* \) would be the simulation analyzed in Section 2.

3.2 Insurance applications

We shall in this subsection design simulation algorithms for aggregated claims against an insurance portfolio of \( n \) policies, not necessarily identical. The details on each policy are stored on file, for example age, claim deductibles and so forth. The viewpoint is
general and applies to an arbitrary period, say a year or a month. Discounting is not included.

**Pure endowment** Suppose the portfolio is one of pure endowment, payable upon death of the insured. Let \( p_j \) be the probability that individual \( j \) dies during the period in question, in which case the company (without any time lag) covers the amount \( \zeta_j \). The algorithm runs as follows.

**Algorithm 3.2 Pure endowment**

1. \( Y^* \leftarrow 0 \)
2. Repeat for \( j = 1, 2, \ldots, n \)
3. Draw \( U^* \sim \) uniform \((0, 1)\)
4. If \( U^* < p_j \) then \( Y^* \leftarrow Y^* + \zeta_j \).
5. Return \( Y^* \).

The computer searches though the portfolio and determines from a given mortality table who survive and who do not and add the claims of the latter.

**Life annuity.** A pension fund is handled almost in the same manner, only now is the aggregation in line 4 over those who survive. The only change required is thus to turn line 4 into

If \( U^* \geq p_j \) then \( Y^* \leftarrow Y^* + \zeta_j \).

However, on recalling that the program is to be run many times, it is quicker to initialize by the sum of all pensions if there are no deaths at all and subtract those that occur, i.e. as follows:

**Algorithm 3.3 Pension fund**

1. \( Y^* \leftarrow \zeta_1 + \ldots + \zeta_n \)
2. Repeat for \( j = 1, 2, \ldots, n \)
3. Draw \( U^* \sim \) uniform \((0, 1)\)
4. If \( U^* < p_j \) then \( Y^* \leftarrow Y^* - \zeta_j \).
5. Return \( Y^* \).

Step 1 is only called once on multiple runs.

**General insurance** The only difference from the pure endowment algorithm is that the fixed claim \( \zeta_j \) is replaced by a stochastic one \( Z_j \). Thus, algorithm 3.2 can be used as it is, except for line 4 being replaced by

If \( U^* < p_j \) then

- Draw \( Z_j \sim \) relevant distribution.
- \( Y^* \leftarrow Y^* + Z_j \).

If there can be several claims from the same policy within a single period, this must be reflected in how \( Z_j \) is drawn.

**Re-insurance** One of the advantages of the Monte Carlo method is that it is so easy to modify the preceding algorithms to cater for re-insurance coverage taken out by the cedent. The programming depends on whether the re-insurance treaty applies to single claims or sums of claims. In the former case the re-insurance influences how the drawing of \( Z_j \) is carried out in the last example, in the latter is the final output \( Y^* \) that is modified by the re-insurance in vigor; more on this in section 5.

### 3.3 Financial applications

In the preceding section the random walk was merely a question of accounting. The model is also a candidate for financial data, but now the validity becomes much less obvious. Consider, for example, the series plotted in Figure 1, which are the monthly share indexes of the stock exchanges in Oslo, Tokyo and New York in the period from October 1985 to September 2000, all together 180 notations. A model describing the dynamics of series of this type is needed, both to investigate future risk of investing in the stock market and to price options. These applications will be considered in Sections 7, 8 and 9, and our concern here is only the model.

It seems obvious that a realistic description should be stochastic, and it is also plau-
possible that randomness only has deterministic effect further on. If there were correlation between different epochs, smart investors could bet on it and beat the market, but they would in the process destroy the basis for their own investment strategy, and drive the correlations they utilized back to zero. If this reasoning holds, it is at a given moment impossible to estimate where in the business cycle the economy is, and utilize such knowledge for investment purposes.

This notion of a perfect market can be put in mathematical form in more ways than one, but by experience it turns out that a random walk model in terms of the logarithms of the actual prices is often a good description. Thus if $S_{jk}$ is the price of index or stock $j$ in period $k$, the random walk applies to $Y_{jk} = \log(S_{jk})$. An empirical indication of this is given in Figure 2, where the auto-correlation function of

$$\varepsilon_{jk} = Y_{jk} - Y_{jk-1},$$

has been plotted.\(^2\) Suppose $\{Y_{jk}\}$ follows the random walk model (3.1). The correlation between $\varepsilon_{jk}$ and $\varepsilon_{jk-l}$ should then be zero for any lag $l > 0$, and so it seems in Figure 2. All correlations are small, and almost all are within the dashed lines marking statistical significance from zero.

Note that it is likely to be correlation between different series the same month. All stock exchanges in our global world interact, and the model for three financial series must be conceived as a vectorial one where the triples $(\varepsilon_{1k}, \varepsilon_{2k}, \varepsilon_{3k})$ come from some joint distribution.

A check on the model is also provided by simulating it in the computer. This is actually a simple and often valuable exercise. We then need the distribution of $(\varepsilon_{1k}, \varepsilon_{2k}, \varepsilon_{3k})$, assumed to be the same at all $k$. Our experiment here is based on adopting the normal distribution with parameters estimated from the real series. To give an indication of realistic values in financial series, the estimated means and standard deviations are shown in Table 3. Note the huge values of the local variation ('volatility') compared to the means, which accounts for the systematic drift upwards of the series. Correlations between the series were in the range 0.3 to 0.6. A simulation algorithm for $J$ correlated series are detailed in Algorithm 3.4.

**Algorithm 3.4 Random walk on log-scale**

0. Input: $(Y_{10}, \ldots, Y_{J0})$ %Arbitrary!.
1. Repeat for $k = 1, \ldots, n$

\(^2\)It is then assumed that all random terms $\varepsilon_{jk}$

<table>
<thead>
<tr>
<th>Norway</th>
<th>Japan</th>
<th>New York</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0065</td>
<td>0.0070</td>
</tr>
<tr>
<td>SD</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 3: Estimated means and standard deviation for the differences between logarithms for the three series in Figure 1.
2. Draw \((\varepsilon_1^*, \ldots, \varepsilon_J^*)\) jointly.
3. Repeat for \(j = 1, \ldots, J\)
4. \(Y_{jk}^* \leftarrow Y_{jk-1}^* + \varepsilon_j^*\)
5. \(S_{jk}^* \leftarrow \exp(Y_{jk}^*)\)
6. Return \(S_{jk}^*\) all \(j \) and \(k\).

A single simulation from Algorithm 3.4 is shown in Figure 3. The general behavior of the real series seem to be reasonably captured; see, however, Section 8.

### 3.4 Autoregressive models

Random walk models can definitely not be used with all financial variables. Consider interest rates. Now there is no longer any systematic growth in the long run, but it may not be good enough to replace the positive means in Table 3 by zero. A random walk without drift is still known to make very long excursions away from the main line in a way interest rates do not.

Interest rates are influenced by complicated political factors that make them difficult to describe mathematically, but if such a description is needed for a financial risk analysis, a useful candidate is the class of autoregressive models; see Shumway (1988) and Priestley (1981). We shall only discuss the simplest first order case. The defining equation is

\[
Y_k = \mu + a(Y_{k-1} - \mu) + \varepsilon_k, \quad (3.2)
\]

where \(\mu, a\) are parameters and \(\{\varepsilon_k\}\) is a process of stochastically independent random terms, which themselves may contain other parameters. We are back to the random walk (3.1) if we take \(\mu = 0\) and \(a = 1\). Simulation of (3.2) is carried out as follows:

#### Algorithm 3.5 Auto-regression

0. Input: \(Y_0^*\)
1. Repeat for \(k = 1, \ldots, n\)
2. Draw \(\varepsilon^*\) any distribution
3. \(Y_k^* \leftarrow \mu + a(Y_{k-1}^* - \mu) + \varepsilon^*\)
4. Return \((Y_1^*, \ldots, Y_n^*)\).

There are two types of applications within the scope of this article. The first is an extension of the insurance applications in Section 3.2. Take \(\mu = 0\) so that (3.2) now becomes

\[
Y_k = aY_{k-1} + \varepsilon_k.
\]
4 Building models

The preceding section dealt with dynamic models, and the most common distributions in insurance and finance are summarized in an appendix. The purpose of this section is to discuss some important issues on how these tools are used for model building.

4.1 What is the scale?

Scale is immediate when dealing with accounts, but less obvious with interest rates. For example, suppose the autoregressive model (3.2) is imposed, with \( Y_k \) being the interest rate in period \( k \). The series \( \{Y_k\} \) then fluctuates around a mean \( \mu \), but it is nothing in the formulation that prevents negative rates. This is not necessarily damning for the model, for its parameters can make such events remote, but it is, at least, an issue that must be faced.

Suppose we take the view that interest is unlikely to fall below certain low values, but that it could climb a good deal higher up from its average. It is two ways to impose such a view on autoregressive models. One possibility is to make the distribution of the error terms in \( \{\varepsilon_k\} \) skewed towards the right. In the preceding section several alternatives were described. Another approach is to change the scale. Now \( Y_k \), instead of being the interest rate \( r_k \) itself, is some transformation \( Y_k = g(r_k) \). A standard choice would be \( g(r) = \log(r) \), but a more flexible possibility is the Box-Cox transformation

\[
g(r) = (r^\gamma - 1)/\gamma, \tag{4.1}
\]

where \( \gamma \geq 0 \) and \( g_0(r) = \log(r) \), the limit of the right hand side of (4.1) as \( \gamma \to 0 \). The parameter \( \gamma \) can even be fitted empirically.

4.2 Statistical estimation

Fitting by maximum likelihood is, as always, a strong possibility for parametric models. It would be beyond our scope to go into this well-known approach in any detail, and we
shall primarily focus on the elementary situation of stochastically independent and identically distributed observations, say \( z_1, \ldots, z_n \). Among the examples cited the data might be claims in a non-life portfolio or increments in a random walk model for securities.

The distribution of \( z_1, \ldots, z_n \) may be determined by fitting one of the models in the appendix. An alternative is the so-called bootstrap. In its simplest and most common form the estimated distribution assigns a probability \( 1/n \) to each observation \( z_i \). A simulation \( Z^* \) from this distribution is then generated according to

\[
P(Z^* = z_i) = 1/n, \quad i = 1, \ldots, n. \tag{4.2}
\]

Note that although we may envisage the true distribution to be continuous, we may still use a discrete estimate. A useful feature of the approach is that it is equally convenient for vectorial data, as, for example, when many securities are considered jointly. The pros and cons of the technique will be examined in Section 4.3 below.

With the autoregressive model (3.2) the general parametric approach is to impose one of the models in the appendix on the error terms, derive the likelihood function and maximize with respect to unknown parameters through some numerical procedure. We might even include the transformation (4.2) in the set-up and find the best-fitting \( \gamma \). The details are too complicated to be included, and we shall instead consider a simpler procedure. The standard estimates of \( \mu \) and \( a \) is the ordinary mean and the lag one autocorrelation coefficient respectively, i.e.

\[
\hat{\mu} = n^{-1} \sum_{k=1}^{n} y_k \\
\hat{s}^2 = n^{-1} \sum_{k=1}^{n} (y_k - \hat{\mu})^2 \\
c_1 = n^{-1} \sum_{k=2}^{n} (y_k - \hat{\mu})(y_{k-1} - \hat{\mu}) \\
\hat{a} = c_1 / \hat{s}^2.
\]

These are least-squares estimates and close to likelihood estimates under Gaussian models. The distribution of the error terms \( \varepsilon_k \) in (3.2) could be estimated as in (4.2), but now with \( z_i = y_i - \hat{\mu} - \hat{a}(y_{i-1} - \hat{\mu}) \).

### 4.3 Parametric or not?

The practical choice of a parametric or a non-parametric approach is influenced by three factors. The first is model accuracy, which is, in turn, determined by the amount of empirical data available. If the relevant part of the time series is long enough, it is tempting to stay clear of parametric assumptions (usually lacking in conviction) and proceed non-parametrically through the bootstrap. The disadvantage is that random errors in such estimated models are larger than for the parametric counterparts. Excellent results with the bootstrap have been reported even with quite few observations in other application areas (down to \( n = 20 \) or \( 10 \)), but in finance there is the powerful countargument against such a practice that the large, rare changes that are so important might not be represented in the material at all. If so, automatic use of the bootstrap (4.2) would underestimate risk.

On the surface, a parametric approach does better. Suppose (for example) a Pareto distribution is fitted. Extremes might then have a considerable chance to materialize. However, they would so due to the fit in the central region of variation of the distribution, and it is uncertain (to say the least) whether they actually captured the real risk at the right tail. It might in such cases be equally sensible to add a heavy-tailed distribution to the estimated, non-parametric one, specifying that some small fraction of the observations belonged to it.

The second factor is speed of implementation. Here the non-parametric approach holds a distinct advantage. During fitting, no numerical optimization is involved and no expert knowledge regarding special families of distributions. Simply apply (4.2). And the bootstrap is also simple to implement in the computer; see below.

The bootstrap is also good in terms computational speed, the third factor. Its sampling algorithm runs as follows:
Algorithm 4.1 Bootstrap sampling
1 Draw $U^* \sim \text{uniform}(0,1)$.
2 $j^* \leftarrow \text{int}(1 + nU^*)$.
3 Return $Z^* = z_{j^*}$.

In the second row $\text{int}(1 + nU^*)$ is the integer value of $1 + nU^*$ (i.e. the largest integer smaller than or equal to it), and the index $j^*$ selected is always between 1 and $n$, since $0 < U^* < 1$. This is much faster than generating drawings from the distributions in the appendix. The advantage is especially great when the observations are vectors.

However, families of distributions possessing the convolution property (i.e. normal, gamma and NIG) recovers speed when simulating the random walk (autoregression also), because we can then write down the distribution of the sum analytically.

4.4 Aspects of time

In practice the models will be influenced by time, both by time resolution and time horizon. The concept of resolution is a central one in engineering and other technical disciplines and signifies how densely data are recorded. In applications in the finance industry this might vary from seconds to years. The shorter is the resolution, the heavier the tails are likely to be. Among the distributions in the appendix the NIG family (see Section A6) might be appropriate for daily notations, whereas monthly ones would be closer to the normal, and yearly data even more.

It is easy to understand this if we impose the random walk (3.1), say for daily data. Then, for monthly ones

$$ Y_k - Y_{k-30} = \sum_{j=0}^{29} \varepsilon_{k-j}, $$

which will be closer to the normal than the daily ones, because of the central limit theorem.

The time horizon, i.e. the time we want to look ahead, is similar. Again many smaller contributions would average out and approach the normal as the lag ahead expands. To approximate the distribution of a prediction it is then not necessary to do more than estimating mean and standard error and use the normal approximation.

4.5 Influence of error

There are two basic sources of errors, either random error due to the future being uncertain and the data experience being limited and specification error, caused by the model in the real world, being infinitely more complex, than the simplification we work with. In a certain, restricted sense it is possible to investigate the adequacy of a model by empirical study, but here there is a very clear limit, and much of it will remain conjecture. This has repercussions for how we should perceive uncertainty estimates. We should regard them as underestimates, since they ignores one source of uncertainty.

Random error is easier. An interesting question is how much a prediction is influenced. Much of the uncertainty of a prediction resides in the uncertainty of the future, but there is also a part than can be attributed to estimation errors in the parameters in the model. Can these errors be disregarded? If there is a fair amount of data behind them, often, but not if the data material is limited. We shall examine this issue in several of the examples later.

5 General Insurance

5.1 Aggregated claims

Our first case study is a classical one. Based on historical records of claims the problem is to compute Value-at-Risk of the portfolio for the coming year. This is both an estimation problem (what is the underlying stochastic model?) and a computational one (how to find the upper percentile of the aggregated claim distribution when the model is given).
An extremely simple approach, solving both issues in one stroke, is to apply the bootstrap. Suppose \( n \) claims \( z_1, \ldots, z_n \) have been reported over a total risk exposure \( A \), measured in years. The estimated claim frequency is then,
\[
\hat{\lambda} = \frac{n}{A}
\]
per policy per year, and the fitted distribution for the number of claims in a portfolio with \( J \) policies might be Poisson with parameter \( J\hat{\lambda} \). Consider the following algorithm:

**Algorithm 5.1 Aggregated claims**
1. Draw \( N^* \sim \text{Poisson}(J\hat{\lambda}) 
2. \( X^* \leftarrow 0 \)
3. Repeat \( N^* \) times
4. Draw \( U^* \sim \text{uniform}(0,1) \)
5. \( j^* \leftarrow \text{int}(1 + nU^*) \)
6. \( X^* \leftarrow X^* + z_{j^*} \)
7. Return \( X^* \).

On output \( X^* \) is a simulation of the sum of claims against the portfolio, based on the available experience. Note the generality of the method. The selection of the index \( j^* \) in line 5 ensures that all preceding claims have the same chance of being chosen. If deductibles and details of the contracts vary to the extent that they should be taken into account, we must also draw the policy from which originates claim \( z_{j^*} \) and modify the payment of the company appropriately.

The method has been tested on simulated data, where a Pareto model with parameters \( \alpha = 3 \) and \( \theta = 2 \) were chosen as the true model. This enables us to compare Value-at-Risk estimates with those obtained when knowing the actual parameters (unrealistically) and when fitting the Pareto model to the data. The results in Table 4 are for \( J = 10000 \) policies with individual accident frequency \( \lambda = 0.01 \), and the experience was either 1000 or 100 claims. The experiment was replicated five times, each time with different data. This accounts for the VaR estimates varying along the horizontal lines.

<table>
<thead>
<tr>
<th>Experience: 100 claims.</th>
<th>Method</th>
<th>Replications</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>126</td>
<td>126</td>
</tr>
<tr>
<td>Fitted</td>
<td>142</td>
<td>116</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>138</td>
<td>113</td>
</tr>
</tbody>
</table>

**Table 4: Estimated 5% VaR under the true and fitted Pareto model and by bootstrapping.**

<table>
<thead>
<tr>
<th>Experience: 1000 claims.</th>
<th>Method</th>
<th>Replications</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>123</td>
<td>139</td>
</tr>
<tr>
<td>Fitted</td>
<td>120</td>
<td>139</td>
</tr>
</tbody>
</table>

Table 5: Damage (in million US$) from twelve accidents on oil rigs.

<table>
<thead>
<tr>
<th></th>
<th>6.74</th>
<th>1.46</th>
<th>5.31</th>
<th>5.64</th>
<th>9.90</th>
<th>0.60</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.17</td>
<td>5.02</td>
<td>0.91</td>
<td>2.47</td>
<td>1.93</td>
<td>3.13</td>
</tr>
</tbody>
</table>

Table 5: Damage (in million US$) from twelve accidents on oil rigs.

except when the true model was used, where Monte Carlo variability is the only source of uncertainty. The number of simulations (10000) is clearly sufficient to keep the effect of simulation error down.

In summary, the simple bootstrap method comes out well. We might guess that it is slightly more variable than the Pareto method, but then the latter was used at home turf, coinciding with the true, stochastic model. In practice we would not know this from an experience of 100 claims.

### 5.2 Premium under uncertainty

Our next example deals with the insurance of big industrial constructions, say oil platforms. Suppose a record of \( n = 12 \) accidents that has occurred in the past is available, as shown in Table 5. The problem is to draw on this limited experience to estimate the pure premium per year per rig, and judge the uncertainty of the estimate. Clearly the margin for error is considerable, and this aspect must be highly relevant when fixing the premium.

In reality the data were simulated from the Pareto distribution with shape parameter \( \alpha = 3 \) and parameter of scale \( \theta = 8 \).
million US$. The mean damage is then 4 million, see Section A.4. Imagine that the twelve accidents have occurred after watching 200 cases of yearly operation of oil platforms. This means that the yearly frequency of accidents is \( \hat{\lambda} = 12/200 = 0.06 \), and a natural estimate of the pure premium \( \pi \) of a single oil rig is

\[
\hat{\pi} = \hat{\lambda} \hat{\mu}
\]

(5.2)

where \( \hat{\mu} \) is the estimate of the mean claim \( \mu \) of each accident.

The claim is influenced not only by the damage, of course, but also by the re-insurance treaty at work. We assume a \( l_d \times l_u \) treaty where \( l_d \) is the deductible and \( l_u \) the maximum retention limit, i.e. the maximum amount covered under the treaty. In this artificial example we do know that damage follows a Pareto distribution. In practice that would be impossible to ascertain from such a limited experience, and we have therefore in Table 6 compared calculations, obtained in four different ways. The first used the true Pareto distribution as model for the damage. The second took the fitted Pareto distribution, the third employed a fitted Gamma distribution and the fourth the mean claim of the observations of Table 5. The deductible \( l_d \) was 1 million US$ and the retention limit \( l_u \) either 10 or 25.

To obtain the estimated expected claim \( \hat{\mu} \) from the fitted (or true) model under the re-insurance treaty requires some tedious (but elementary) mathematical details, which is natural to skip in a course on simulation. The sampling procedure then becomes:

**Algorithm 5.2 Pure premium**

1. Draw \( N^* \sim \text{Poisson}(n) \) \% \( n = 12 \)
2. \( \hat{\lambda}^* \leftarrow N^*/B \) \% \( B = 200 \)
3. Repeat for \( j = 1, \ldots, n \)
4. \( \text{Draw } z_j^* \) % Several alternatives!
5. If \( z_j^* > l_u \), then \( z_j^* \leftarrow l_u \)
6. \( z_j^* \leftarrow \max(z_j^* - l_d, 0) \) % Re-insurance!
7. Estimate \( \hat{\mu}^* \) from \( z_1^*, \ldots, z_n^* \)
8. Return \( \hat{\pi}^* \leftarrow \hat{\lambda}^* \hat{\mu}^* \).

The two sources of uncertainty in \( \hat{\pi} \) is reflected in the algorithm. First the frequency \( \hat{\lambda}^* \) of accidents is sampled and then the claims \( z_1^*, \ldots, z_n^* \) are generated and passed to an estimate \( \hat{\mu}^* \) in line 7. Note that the number of accidents \( n \) should not be sampled in this part of the algorithm.

The results are shown in Table 6, where estimates and 90% confidence limits are compared for the four methods used to estimate the expected claim \( \hat{\mu} \). The number of simulations was \( m = 999 \). Note the effect of the estimation error on the upper confidence limit, which is much higher when the true model is used.

<table>
<thead>
<tr>
<th>Method</th>
<th>( l_u = 10 )</th>
<th>( l_u = 25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>.04 .142 .34</td>
<td>.05 .176 .49</td>
</tr>
<tr>
<td>Pareto</td>
<td>.06 .154 .23</td>
<td>.06 .169 .27</td>
</tr>
<tr>
<td>Gamma</td>
<td>.07 .160 .26</td>
<td>.07 .165 .27</td>
</tr>
<tr>
<td>Boot.</td>
<td>.07 .169 .30</td>
<td>.07 .164 .30</td>
</tr>
</tbody>
</table>

Table 6: Estimated pure premium with lower (lo) and upper (up) 90% confidence limits around estimate (in the middle).

This is not accidental. We have in Table 7 reported a larger simulation experiment based on 1000 data sets of the same type as in Table 5. The confidence limits were computed for each data set and it was verified whether the the true pure premium \( \pi \) was located below or above the two limits. In this way we could estimate the true confidence level, as opposed to the targeted 5% at each side. The picture emerging is that uncertainty upwards is grossly undervalued, and this might be a factor when pricing the re-insurance.
\[ l_u = 10 \quad l_u = 25 \]

<table>
<thead>
<tr>
<th>Method</th>
<th>Lo</th>
<th>Up</th>
<th>Lo</th>
<th>Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto</td>
<td>.03</td>
<td>.09</td>
<td>.02</td>
<td>.13</td>
</tr>
<tr>
<td>Gamma</td>
<td>.08</td>
<td>.12</td>
<td>.04</td>
<td>.19</td>
</tr>
<tr>
<td>Boot.</td>
<td>.04</td>
<td>.12</td>
<td>.03</td>
<td>.19</td>
</tr>
</tbody>
</table>

Table 7: Estimated confidence levels based on 1000 replications. Nominal value: 0.05 at either side.

6 Life insurance

6.1 Introduction

General life insurance is based on a multi-state stochastic process, which runs randomly through a given set of states, such as active, disabled or dead, influencing payments. It is not difficult to write down simulation algorithms for such systems, utilizing that the sequence of events occurs recursively and randomly on each other. Non-Markovian models are no more difficult than Markov ones.

The general case will be deferred to Section 6.5 below. In order to keep things simple and transparent most of the discussion deal with two-state Markov models. The stochastic part is then determined by the mortality intensity, say \( \nu_a \) of dying immediately upon having the reached age \( a \). We shall need the probability \( h_p_a \) of somebody \( a \) years old surviving a time period of length \( h \), linked to the intensity through the well-known relation

\[
h_p_a = \exp\left(-\int_0^h \nu_{a+t} dt\right). \tag{6.1}
\]

For the experiments reported we have imposed the Gompertz-Makeham model

\[
\nu_a = \alpha + \beta \gamma^a \tag{6.2}
\]

which yields

\[
h_p_a = \exp\left\{-\alpha h - \beta \gamma^{h} - \frac{1}{\log \gamma}\right\}. \tag{6.3}
\]

The coefficients were assigned the values \( \alpha = 0.0009, \beta = 0.0000044 \) and \( \gamma = 1.10154 \), which gives a reasonable description of real life mortality.

6.2 Portfolio description

Evaluation of the portfolio risk requires additional details about the policy holders and the contracts, notably

- age \( a \),
- premium \( \pi \), payable each time unit,
- sum insured/pension, both denoted \( \zeta \),
- age of retirement \( a_r \) (for pensions),
- age \( a_c \) of the when contract expires.

All these quantities depend on the policy in question, and we should imagine that a list of them has been stored in the computer. The fact that they depend on the individual policy \( j \) has been suppressed in the notation.

In general, each policy holder must be assigned a state or a label which changes over time. In our examples below the insured is active (alive and paying premium), retired (receiving pension) or dead. Note that these states are traversed in deterministic order (although the changes take place at random time), and for this reason we can do without an explicit state variable \( c \) in the scheme. But under other circumstances (for example when dealing with disability insurance) a state variable would be required; see 6.5 below.

6.3 Pure endowment

Consider first pure endowment insurance where an insured, initially at age \( a \), is followed \( n \) time units of length \( h \). The premium \( \pi \) is paid at the start of each period until the insurance expires at the age \( a_r \). The discount factor per period is \( v_t \). If the policy holder dies, his beneficiaries receive the amount \( \zeta \) at the end of the current period.

Algorithm 6.1 starts by setting the discount factor \( v \) and the initial age \( a \). The movements of the account is then tracked until, possibly, the policy holder dies. On return \( Y^* \) contains the change of the account during the the period.
6.4 Pension schemes

The second example is a simple pension plan, with advance payment of both premium $\pi$ and pension $\zeta$ at the start of each time period. The simulation algorithm 6.2 is constructed along lines as in 6.3, albeit with a few changes. As the policy holder reaches the retirement age $a_r$, the pay into the account is switched to withdrawals, and the scheme stops in case of death. Note the ease with which an endowment could have been worked in. The amount is simply withdrawn in Step 5 before going to 8. The pension lasts until death, but a time limit could have been included, as well.

Algorithm 6.2 Pension scheme

1. $Y^* \leftarrow 0, v \leftarrow 1, a \leftarrow \text{initial age}$
2. Repeat for $k = 1, \ldots, n$ while $a \leq a_e$
3. $Y* \leftarrow Y* - \pi v$
4. Draw $U^* \sim \text{uniform} \ (0,1)$
5. If $(U^* > \lambda p_a)$ then \%Death!
   $Y^* \leftarrow Y^* + \zeta v$, goto 7
6. $v \leftarrow v_1v, a \leftarrow a + h$ \% Update.
7. Return $Y^*$

Again the algorithm applies to a single policy and has to be run $J$ times for a portfolio consisting of $J$ policies. The output $Y^*$ from each run is then added for the portfolio output.

We have tested the algorithm on a small portfolio of 1000 policies. The contracts were identical and drafted when each individual was 30 years. The pension period starts 35 years later at 65. A pension of 15000 Nkr is then paid out at the start of each month. The premium (Nkr 1625, equivalence determined at annual interest 4%) was paid monthly in advance as were the withdrawals from the account. The age distribution of the members of the pension scheme was laid out so that it matched exactly the Gompertz-Makeham law, specified in 6.1. That meant that 764 among the 1000 policies had reached retirement age and received monthly payments.

<table>
<thead>
<tr>
<th>Size</th>
<th>Replications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>12.3 12.3 12.2 12.3 11.4</td>
</tr>
<tr>
<td>10000</td>
<td>80.5 80.3 80.3 81.2 81.3</td>
</tr>
</tbody>
</table>

Table 8: 5% VaR estimates for the yearly account of a pure endowment portfolio with 1000 and 10000 policies.

Algorithm 6.1 Pure endowment

1. $Y^* \leftarrow 0, v \leftarrow 1, a \leftarrow \text{initial age}$
2. Repeat for $k = 1, \ldots, n$ while $a \leq a_e$
3. $Y* \leftarrow Y* - \pi v$
4. Draw $U^* \sim \text{uniform} \ (0,1)$
5. If $(U^* > \lambda p_a)$ then \%Death!
   $Y^* \leftarrow Y^* + \zeta v$, goto 7
6. $v \leftarrow v_1v, a \leftarrow a + h$ \% Update.
7. Return $Y^*$

The algorithm applies to a single policy holder. To simulate a portfolio of $J$ policies steps 1-7 are repeated $J$ times, taking initial age, premium and so forth from a stored list. The output from these $J$ runs are then added for the total account of the portfolio.

An example is shown in Table 8. There were $J = 1000$ or $J = 10000$ policies, all having been drawn up at the age of 30 years and expiring 35 years later. Under the contract a claim of one million Nkr is presented the company upon death of the insured, payable at the close of the current half-year. The premium is paid in advance two times a year. With interest 4% p.a. the equivalence premium is Nkr 2620. This amount has been used for $\pi$ below. The age distribution of the portfolio followed exactly the Gompertz-Makeham probabilities given above.

The results in Table 8 are based on 1000 simulations. There are signs that some random error is present, but for planning purposes for the following year the accuracy seems sufficient.
Table 9: VaR (5\%) and median from simulations (99) of a pension scheme (in million Nkr)

<table>
<thead>
<tr>
<th>Replications</th>
<th>95%</th>
<th>126.1</th>
<th>126.8</th>
<th>126.6</th>
<th>126.5</th>
<th>126.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>50%</td>
<td>125.0</td>
<td>125.1</td>
<td>125.2</td>
<td>125.0</td>
<td>125.1</td>
<td></td>
</tr>
</tbody>
</table>

VaR estimates (5\%) of the net claim\(^3\) on the account of the pension scheme have for the following year been computed in Table 9 and compared to the median. Only 99 simulations have been used. The simulated values are stable even for this low number. This is a result of the underlying distribution being quite flat (only 1 - 1.5 million in difference between the median and the 95\% percentile). The total claim with no death occurring at all is around 135 million and the estimates actually obtained correspond to around 50 - 60 casualties. If the number of simulations was increased to 1000, the median was 125.05 and the VaR varied between 126.3 and 126.4 million for the repeated runs.

6.5 General life insurance

In the general case an individual at each point in time is in one of the states 0, 1, \ldots, L and receives or has to pay an amount \( \zeta(c) \) at state \( c \). For some states \( c \) premium is received by the company, in others a pension or a one-time endowment is payed back to the insured or his relatives. In the algorithm below this pay-off function must be identified from the details of the contract. As above, \( \zeta \) will vary between policies, without this having been explicitly included in the notation.

It is assumed below that a policy during its lifetime evolves according to a Markov process \( \{C_t\} \). The initial state is known, but the subsequent development can take many different courses, governed by the system of probabilities

\[
p_a(j, l) = P(C_{a+h} = l | C_a = j),
\]

which defines the likelihood that an individual is in state \( j \) at age \( a + h \) having been in state \( l \) at age \( a \). These quantities are assumed known.

The extension of algorithms 6.1 and 6.2 is detailed in Algorithm 6.3.

Algorithm 6.3 General

1. \( Y^* \leftarrow 0, v \leftarrow 1, a \leftarrow \text{initial age}, \)
2. \( C^* \leftarrow \text{initial state.} \)
3. \( Y^* \leftarrow Y^* + \zeta(C^*) v \quad \% \text{Payoff!} \)
4. \( q \leftarrow 0 \quad \% \text{New state} \)
5. Repeat for \( c = 0, 1, \ldots, L \) \% selected
6. \( q \leftarrow q + p_a(C^*, c) \quad \% \text{here} \)
7. Draw \( U^* \sim \text{uniform}(0, 1) \)
8. If \( U^* < q \) then
   \( C^* \leftarrow c, \text{goto 9} \)
9. \( v \leftarrow v_1, a \leftarrow a + h \quad \% \text{Update v and a}. \)
10. Return \( Y^* \).

The insured enters the scheme at age \( a \) which is incremented at step 9. The age defines the probabilities of switching states, which is carried out in steps 4 to 8 and ensures that the new state is picked out with the right probability. The general run is similar to the earlier algorithms. Note that all movements in and out of the account are advance payments.

7 Financial application

We shall in this section try to predict the future value of an investment basket in the three monthly stock indices in Figure 1 in Section 3. Both the mean and the Value-at-Risk will be used as indicators. An important issue is what model to use. This is discussed first.

7.1 Empirical analysis

The data in Figure 1 is a 15 year extract of a series that are about twice that long. One

\(^3\text{net claim means the amount paid out minus premium income.}\)
possible model is the multivariate random walk on log-returns. The simulations in Figure 2, obtained by sampling a multivariate Gaussian distribution (parameter estimates obtained from the whole series), showed that such a model is for phenomena exhibiting large random variation. The same picture is conveyed by Figure 5 where several simulations series of one of the indexes (that from New York) have been plotted jointly for 15 years ahead.

It may be difficult to believe that 2 of the 5 series should produce a growth of more than 10 times the original index value, but this does occur in spite of the standard autocorrelation plot in Figure 2 showing little sign of any deviation from the random walk model. The explanation is non-linear effects that are not captured by autocorrelation plots. The assertion is substantiated in Figure 6, where the autocorrelation of the absolute values of the log-series have been plotted. The patterns there would hardly materialize without some dependence in the series, actually at quite long range. The effect could be non-linear and caused by investors after long periods of growth cashing in or after decline betting on renewed growth.

We shall ignore this effect and impose a simple random walk. The model then overstates variability in the long run, which suggests that the Value-at-Risk estimates may be conservative, exaggerating the risk. It will emerge, however, in Section 7.2 that the picture is not necessarily that simple.

We shall apply the bootstrap algorithm 4.1 to sample the error terms of the random walk, using the empirical distribution of the differences (3.2), as calculated from the observed series. There is a possible trap in this. The empirical distribution function (and any other fitted distribution too) contains random error. Does this matter? Look at Table 3. The drifts upwards were from 0.006 to 0.008 with standard deviation around 0.05. The standard error of the drift estimates based on 367 values is then $0.05 / \sqrt{367} \approx 0.004$, and this is a considerable value, suggesting that random error in the parameters can not be ignored when evaluating future risk.

### 7.2 Evaluation of risk

The basket had equal shares (1/3) of the three stock indexes. The value at termination of the actual quotations were taken as the initial value of the portfolio and its development simulated by Algorithm 3.4 with the random vectors drawn by means of the boot-

---

**Figure 5:** Simulations (five replications) of the New York index 15 years ahead.

**Figure 6:** Estimates of the correlation and cross-correlation functions for the three series in Figure 1, when transformed to absolute values of differences between logarithms.
strap algorithm 4.1. The relative increase in value over 1, 5 and 10 years were noted and repeated 100000 times for assessments little influenced by random error. The mean in Table 10 is the average of the relative increase of these runs and the 5% Value-at-Risk obtained as the value of the 5000th smallest among the simulations.

The worst case scenarios are below 1 when looking both 1 and 5 years ahead, signifying loss of money. We have to move 10 years into the future before the a positive return is expected even when taking a very pessimistic view. The positive drift upwards in Table 3 then prevails over bad luck.

What is the random error in these estimates? The simulation error is minor, as mentioned, but there is uncertainty in the historical record. To investigate the effect of the latter a double bootstrap experiment was run in the following manner. Suppose there are 15 years of monthly indexes at disposal, altogether 180 notations. Then these records are sampled with replacements exactly 180 times to produce an alternative set of 15 year records, from which we can compute forecasts exactly as in Table 10. When this process is repeated, we obtain a series of alternative estimates which convey the random variation due to uncertainty in the historical record. The technique is discussed at length in Efron and Tibshirani (1993).

The standard error estimates quoted in Table 11 are obtained as empirical standard deviations of 100 replicated estimates of the mean and the Value-at-Risk. It confirms that the uncertainty due to the limited historical records is huge.

<table>
<thead>
<tr>
<th>Data history</th>
<th>Period ahead (years)</th>
<th>1</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 years</td>
<td>Mean</td>
<td>1.13</td>
<td>1.85</td>
<td>3.44</td>
</tr>
<tr>
<td></td>
<td>VaR</td>
<td>0.82</td>
<td>0.90</td>
<td>1.22</td>
</tr>
<tr>
<td>30 years</td>
<td>Mean</td>
<td>1.13</td>
<td>1.86</td>
<td>3.46</td>
</tr>
<tr>
<td></td>
<td>VaR</td>
<td>0.85</td>
<td>0.94</td>
<td>1.24</td>
</tr>
</tbody>
</table>

Table 10: Mean and 5% Value-at-risk estimates for the relative increase of the equal weights basket.

<table>
<thead>
<tr>
<th>Data history</th>
<th>Period ahead (years)</th>
<th>1</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 years</td>
<td>Mean</td>
<td>0.06</td>
<td>0.48</td>
<td>1.73</td>
</tr>
<tr>
<td></td>
<td>VaR</td>
<td>0.05</td>
<td>0.23</td>
<td>0.59</td>
</tr>
<tr>
<td>30 years</td>
<td>Mean</td>
<td>0.03</td>
<td>0.30</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>VaR</td>
<td>0.03</td>
<td>0.16</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Table 11: Estimated standard error of the predictions in Table 10; based on 100 simulated data sets.

8 Simulating business

8.1 Model

Stochastic simulation is a well established tool to try to obtain insight into the joint effect of several uncertain factors; see, for example, Daykin, Pentikäinen and Pesonen (1993). We shall in this section examine an artificially created insurance portfolio of 100000 policies. The monthly claim frequency is by experience $\lambda = 0.004$ and there are on file 10000 earlier claims, adjusted for inflation. In this exercise this file has been created by simulating claims from the Pareto distribution with parameters $\theta = 0.2$ (scale) and $\alpha = 3$ (shape). The mean claim is then 0.1 (see Section A.9), measured, say in million Nkr. The monthly pure premium is then $\pi = 0.0004$.

The reason for working with months as time unit is that some of the accumulated capital will be invested in securities noted on the Oslo Stock Exchange. To assess the risk involved in this, we shall involve the monthly index data of the preceding section. The purpose of the exercise is to investigate how the business uncertainty is influenced by the investment strategy chosen. Let $Y_k$ be the value of the account after month $k$, with $Y_0 = y_0$ be the amount put in initially. A possible model for $\{Y_k\}$ is the scheme

$$Y_k = (1+r_k)Y_{k-1} + J_k\pi(1+\eta) - X_k,$$

(8.1)
where

- $r_k$ is rate of return on capital,
- $J_k$ is the portfolio size,
- $\eta$ represents safety loading,
- and $X_k$ is the sum of claims (month $k$).

Many simplifications will be made. The portfolio size $J_k$ is the same for all $k$. There is no time lag between the damage and the payment, and all policies are treated as identical. The most sophisticated sub-model is that for $r_k$, which is assumed to be composed of a fixed and random component, i.e.

$$r_k = (1 - q)r^b + qr^s_k,$$  \hspace{1cm} (8.2)

where $r^b$ is a monthly, risk-free and fixed bond interest whereas $r^s_k$ is the random return in the Norwegian stock market. It was in algorithm 8.1 drawn by means of the bootstrap. The investment basket is re-balanced each month (without cost) to maintain a constant share $q$ in volatile stock. Premium $\pi(1 + \eta)$ is payed at the end of each month. There is no specific term for administration costs, and we may imagine that this is included in $\eta$.

8.2 Algorithm and results

The simulation of the model defined by (8.1) and (8.2) $n$ time units (months) ahead can be summarized like this. First compute and store $I_p = J(1 + \eta)\pi$, and then:

Algorithm 8.1 Insurance with investment
1. $Y^* \leftarrow y_0$
2. Repeat for $k = 1, \ldots, n$
3. Draw $X^* \%$ By algorithm 2.1
4. Draw $(r^s)^* \%$ By algorithm 4.1
5. $r^* \leftarrow (1 - q)r^b + q(r^s)^*$
6. $Y^* \leftarrow (1 + r^*)Y^* + I_p - X^*$
7. Return $Y^*$.

The experiments reported were carried out on the data and portfolio described in 9.1. above. The safety loading $\eta$ was 1% in Figure 7 (and Table 12) and $= 0$ in Figure 8. The latter case means that the company does not earn money on the insurance at all, only on the investments. The initial capital reserved for the portfolio was 50 million Nkr.

The development of the portfolio account is simulated 100 times in Figure 7 and 8 and plotted jointly for 60 months ahead (i.e. 5 years). Note that investments are varied from none at all on the left, via fixed return (6%) in the middle and 30% placed in the stock market on the right. The effect of including capital gains is evident and the higher return when risky shares are included is marked. Note that we could take
Table 12: Value-at-Risk evaluations (in million Nkr) of the insurance portfolio when $\eta = 0.01$

<table>
<thead>
<tr>
<th>VaR</th>
<th>1 year ahead</th>
<th>5 years ahead</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No inter.</td>
<td>6% inter.</td>
</tr>
<tr>
<td>1%</td>
<td>21.5</td>
<td>23.6</td>
</tr>
<tr>
<td>50%</td>
<td>56.5</td>
<td>59.4</td>
</tr>
<tr>
<td>9%</td>
<td>86.2</td>
<td>90.3</td>
</tr>
</tbody>
</table>

out the sample space of the account at any given point in time by drawing a vertical line though the cloud, except for 100 simulations then not being sufficient. Results from 10000 simulations are shown in Table 12 under the same three investment scenarios. Note the effect of having a stock market part included, the prospect of higher earnings is considerable and the risk downwards is not increased.

9 Option pricing

9.1 Introduction

The pricing of options requires heavier mathematics than in earlier sections. Much of what is available in scientific literature is written in the language of stochastic analysis and is not too easily accessible to everybody. Very readable accounts are given in Wilmot, Howison and Dewynne (1995) and Björk (1998). The comments on the Black and Scholes theory I give below, make no attempt at all to capture the mathematics. My aim is to explain the main idea in a form that is as transparent as possible to people lacking background in formal stochastic analysis. The aim of these notes is stochastic simulation anyway, but to understand the computational issues involved in pricing options, it is necessary to spend more time with modeling aspects than in earlier sections.

9.2 The model

Option pricing deals with the fair premium one should pay for having the right, but not the obligation, to buy or sell some financial asset after having seen how the market develops. A central issue is the stochastic model describing the future fluctuations of the price of the asset, balanced against the return when the money is put in the bank. The common viewpoint is in continuous time, and if not for other reasons, this is the type of model we should introduce.

Suppose there is, at time $t$, two types of assets, either a risky security, priced as $S(t)$, or a risk-less bond, valued $B(t)$. The development of the latter evolves as

$$d \log \{B(t)\} = r dt,$$  \hspace{1cm} (9.1)

where $r$ is interest rate. The differential form expresses that the account grows by an amount $dB(t) = rB(t)dt$ in an interval of length $dt$.

The model for the security is, in its simplest form,

$$d \log \{S(t)\} = adt + \sigma dW(t),$$  \hspace{1cm} (9.2)

where $\{W(t)\}$ is the Wiener process (also called Brownian motion), that is a Gaussian process with so-called orthogonal increments. The latter means that $W(t+h) - W(t)$ and $W(t' + h') - W(t')$ are 

uncorrelated

if $t < t + h < t' < t' + h'$, i.e if the intervals do not overlap. Additional requirements are

$$E \{W(t+h) - W(t)\} = 0,$$

$$\text{var}\{W(t+h) - W(t)\} = h,$$

for all $t$ and $h > 0$.

To interpret the model (9.2) it is convenient to pass to finite increments. Let

$$Y(t) = \log S(t).$$

Then, by (9.2)

$$Y(t+h) - Y(t) = ah + \sigma \{W(t+h) - W(t)\},$$  \hspace{1cm} (9.3)

for all $h > 0$. The most useful result is

$$Y(T) = Y(0) + \int_0^T \{a + \sigma W(t)\} dt.$$
from which it emerges that the process \( Y_k = Y(hk) \) is an ordinary Gaussian random walk as \( h \) is varied and \( h \) is kept fixed. The model (9.2), known as geometric Brownian motion, is thus the one we used in Section 7, but now in continuous time.

The parameters \( a \) and \( \sigma \) capture two different aspects. Clearly \( a \) represents deterministic drift (upwards or downwards), whereas \( \sigma \) expresses volatility; i.e. the magnitude of the randomness. Neither of these quantities is necessarily constant. Indeed, the pricing theory below is valid even if we allow their values to be influenced by the current value of the security, i.e. if

\[
a = a(S(t), t), \quad \sigma = \sigma(S(t), t). \tag{9.4}
\]

Effects of this type are known to be present in real life.

### 9.3 Model for many securities

We need a vectorial version of (9.2) to handle investment baskets where many securities are involved. Let \( S(t) \) be a vector of \( J \) assets \( S_j(t) \) running in parallel with \( \log \{ S(t) \} \) being the vector of logarithms of the individual processes. Moreover, let \( \mathbf{a} = (a_1, \ldots, a_J)' \) contain \( J \) drift parameters (one per process) and \( \mathbf{W}(t) \) the vector of \( J \) independent Wiener processes \( W_j(t) \). The vector version of (9.2) then becomes

\[
\frac{d \log \{ S(t) \}}{dt} = \mathbf{a} d t + \mathbf{C} d \mathbf{W}(t), \tag{9.5}
\]

where \( \mathbf{C} = (C_{ji}) \) is a non-singular \( J \times J \) matrix, which replaces the scalar volatility \( \sigma \) in (9.2).

The fluctuations of the individual securities become correlated by virtue of \( \Sigma \). Note that the \( J \) processes are independent if \( \Sigma \) is a diagonal matrix. Dependence is, of course, highly frequent in practice. Local variations in drift and volatility are allowed even in the vector version of the model. The analogies to (9.4) are \( \mathbf{a} = \mathbf{a}\{S(t)\} \) and \( \mathbf{C} = \mathbf{C}\{S(t)\} \).

### 9.4 Hedging

Among the many types of options traded in the marketplace, we shall consider the simplest and most common one where the option has to be exercised at a certain time \( T \) (the 'strike'), if at all. These are known as European options. One example is that a buyer has the right to buy at an agreed price \( K \). He will do so if \( S(T) > K \) gaining \( S(T) - K \).

The profit by holding the option is thus

\[
X = \max\{S(T) - K, 0\} \tag{9.6}
\]

with expected profit \( E(X) \).

More generally, suppose the same type of option is in terms of \( J \) securities, forming a basket with weights \( w_1, \ldots, w_J \). The gain of the option buyer now becomes

\[
X = \max\{\sum_{j=1}^{J} w_j S_j(T) - K, 0\}, \tag{9.7}
\]

making the computation of the expected gain \( E(X) \) more demanding.

It is most convenient to outline the theory in terms of a completely general function \( \Phi \) so that the profit is \( X = \Phi\{S(T)\} \) with expectation \( E\{\Phi\{S(T)\}\} \). Does it appear reasonable that the option premium has something to do with \( E(X) \)? There are several complicating factors. The option is set up (or traded between two parties) at an earlier time, say at \( t = 0 \). This means that the gain should be discounted by the factor \( \exp(-rT) \). Also note that the probability distribution of \( S(T) \) must really be the conditional one given \( S(0) = s \) (which has been observed when the deal is set up. We shall reflect the conditioning in the mathematical notation by writing \( E^*\{\Phi\{S(T)\}\} \) for the expected profit at \( T \).

Should this value, properly discounted, be the option premium? Superficially it may seem so, but there is another, more subtle side of this. Such a price would ignore that the option can be traded. In a liquid market it would be possible for a financial institution or an individual bearing responsibility.
for the option to hedge by taking position in the opposite direction. For example, if the option is of the type in (9.6), the seller may buy \( \delta \) units of the underlying asset, setting up a hedging portfolio. Suppose \( V^*(t) \) is the market value of the option at time \( t \). The value of the hedging portfolio is then

\[
-V^*(t) + \delta S(t),
\]

and the second term counteracts the first, since \( V^*(t) \), whatever it is, must increase with \( S(t) \). This type of argument can be advanced even if the asset is not traded; see Björk (1998).

What is the most risk-reducing value for \( \delta \)? Clearly that would depend on how \( S(t) \) develops. Hence \( \delta = \delta(t) \) is time-dependent. It can actually be shown that by adjusting \( \delta(t) \) appropriately, all risk is removed! The same applies to a basket of many financial assets, and the assertion holds under condition (9.4) if some weak restriction is imposed.

Suppose people actually do this, continuously balancing and re-balancing risk. What would then be the expected value of \( \Phi\{S(T)\} \)? **Answer:** The one obtained under a different model!

### 9.5 Risk neutrality

The model defined by (9.2) and (9.4) is in the financial mathematics literature known as the \( P \)-model.\(^4\) There is an associated \( Q \)-model

\[
d \log \{S(t)\} = r dt + \sigma dW(t), \quad (9.8)
\]

which is identical to (9.2) except for the drift term now being determined by the risk-free bond interest \( r \).\(^5\) Note that the value of the security under this second model has an average growth rate that is identical to an ordinary account in a bank. This must be so, since the random term \( dW(t) \) in (9.8) has mean zero.

The option premium is under the Black and Scholes theory calculated as if the \( Q \)-model is the true one, i.e. by

\[
\pi^* = E^*_Q \{ \Phi\{S(T)\} \}, \quad (9.9)
\]

where the notation \( E^*_Q \) signifies the distribution of \( S(T) \) given \( S(0) = s \) under the model \( Q \).

Why is (9.9) a fair price? Because otherwise there would exist opportunities of arbitrage, that is risk-less possibilities to earn money from financial speculation. The assertion has been established by rigorous mathematics. Here is a bit of intuition. A seller, responsible for the delivery of the asset at time \( T \) if the buyer so wants, constructs the risk-less portfolio described in 9.4. Presumably market forces prevent the value of such a portfolio to grow differently from the risk-less interest rate, and this becomes the drift parameter under which the fair option premium should be calculated.

The model \( Q \) is known as the risk neutral one. Often financial mathematics literature refer to \( P \) and \( Q \) as measures.

### 9.6 Discussion

Perhaps the most striking feature of the solution is that the drift term \( a \) in (9.2) has disappeared so that it does not matter if the two parties hold different views on it. Neither is it of any significance if \( a \) is wrongly estimated from historical data. This is fortunate since we saw earlier that the drift term might be hard to estimate reliably. The absence of the term \( a(S(t), t) \) in the solution may seem surprising, but on second thought it must be like that in the absence of arbitrage opportunities.

It is usually impossible to express the solution (9.9) in simple form, and simulation
is therefore a natural, and much applied, approach. There is a notable exception when there is only one security and the volatility $\sigma$ is constant. Then the famous Black and Scholes formula reads

$$\pi^* = s \phi \{ d_1 (s) \} - \exp (-r T) K \phi \{ d_2 (s) \},$$

where $\phi$ is the distribution function of the normal $(0, 1)$ and

$$d_1 (s) = \frac{1}{\sigma \sqrt{T}} \{ \log (s) - \frac{1}{2} \sigma^2 T \},$$

$$d_2 (s) = d_1 (s) - \sigma \sqrt{T}.$$

To compute the solution we only need the volatility $\sigma$. Historical data on the price fluctuations $S(t)$ might be used. Another approach is the so-called implied volatility, where the actual trading of options are used to determine $\sigma$. As justification for this approach it has been advanced in the literature that volatility is not constant.

### 9.7 Computation

Practical evaluation of the option premium requires the continuous model to be made discrete. Recall (9.3) and let $Y^0 = Y^0 (kh)$. Since we are working under the $Q$-model, the drift term $ah$ should be replaced by $rh$. Hence (9.3) can be rewritten

$$Y^0 = Y^0 \varepsilon + \varepsilon_k,$$

where

$$\varepsilon_k = rh + \sigma \{ W(kh) - W(kh - h) \}.$$

Note that $\varepsilon_k$ is Gaussian with mean $rh$ and standard deviation $\sigma \sqrt{h}$. We seek the distribution of $Y^0_n$ for $n = T/h$ given $Y^0_0 = \log (s)$. Then $S_n = \exp (Y^0_n)$ is the price $S(T)$ at $t = T$, and can be plugged into the option formula (9.6) above.

Samples of $Y^0_n$ may be generated by running Algorithm 3.1 or 3.4. Note that the distribution of $\varepsilon_k$ may depend on $Y^0_k$ if the volatility $\sigma$ is linked to the price of the asset through (9.5). It is obvious how this can be accommodated in the algorithms once a suitable relationship (9.5) has been established.

If the volatility is constant, the distribution of $Y^0_n$ is Gaussian and can be written down immediately. In the vectorial case $Y^0_n$ has mean $\log (s) + n r^0$ and covariance matrix $n \Sigma = n C^C$. The example run in Section 2.3 above (see Table 2) corresponds to a basket of two shares with time unit one month. The two financial assets were initially equally valuable; i.e. $S_1 (0) = S_2 (0) = 1$. Monthly interest was $r = 0.005$, and $T = 12$ (which corresponds to one year ahead). The volatility matrix $\Sigma$ had standard deviations 0.0042 and 0.0083 for the two shares. Their correlation was 0.5.

### 10 References


Here $r$ is the vector where all elements equal $r$. 

---

24
A Distributions

The purpose of this section is to summarize, for quick reference, the most important families of distributions for applications in insurance and finance, including how they can be sampled. In choosing algorithms the emphasis is on quick implementation rather than optimal speed. A very good reference on sampling technique is still Devroye (1986).

A.1 Sampling by inversion

There are several general methods of converting drawings of uniform random variables into samples from a given distribution, see Devroye (1986). The conceptually simplest of them is the so-called inversion method. Let $F(z)$ be the distribution function of a random variable $Z$ and assume that we know how to compute the $u$-percentile $F^{-1}(u)$, i.e. to solve the equation $F(z) = u$ for all $u$ between 0 and 1. The following algorithm generates simulations of $Z$ with distribution $F$:

Algorithm A.1 Sampling by inversion
1. Draw $U^* \sim \text{uniform } (0,1)$.
2. $Z^* \leftarrow F^{-1}(U^*)$.
3. Return $Z^*$.

A.2 Poisson

The Poisson model for an integer-valued random variable $N$ has the discrete probability density function (p.d.f)

$$f(n) = \frac{\lambda^n}{n!} \exp(-\lambda),$$

for $n = 0, 1, \ldots$ and $\lambda > 0$. The mean and variance are equal, i.e.

$$E(N) = \lambda, \quad \text{var}(N) = \lambda.$$

Sampling is not entirely straightforward, but the random variable, $N^*$ returned by the following algorithm is Poisson distributed with parameter $\lambda$.

Proof: If $U^*$ is uniform $(0,1)$ and $F(z)$ is the strictly increasing distribution function of $Z$, then

$$P[Z^* \leq z] = P[F^{-1}(U^*) \leq F^{-1}(z)] = P[U^* \leq F(z)] = F(z).$$

Algorithm A.2 Sampling Poisson
1. $S^* \leftarrow 0$
2. Repeat for $j = 1, 2, \ldots$
3. Draw $U^* \sim \text{uniform } (0,1)$
4. $S^* \leftarrow S^* - \log(U^*)$
5. Until $S^* > \lambda$

A.3 Negative binomial

The negative binomial (or Pascal) distribution arises when the parameter $\lambda$ in a Poisson model is stochastic and drawn from a gamma distribution. Suppose $\lambda$ follows the p.d.f.

$$f(\lambda) = \frac{(\alpha/\xi)^\alpha}{\Gamma(\alpha)} \lambda^{\alpha - 1} \exp(-\alpha \lambda/\xi)$$

for $\lambda > 0$. Here $\alpha > 0$ and $\xi > 0$ are parameters; see A.8 below. An example is heterogeneous populations in insurance. Then $\lambda$ is the random risk parameter attached to each individual, and $N$ is conditionally Poisson distributed given $\lambda$. The unconditional distribution of $N$ is then

$$f(n) = \binom{n + \alpha - 1}{\alpha - 1} p^n (1 - p)^n,$$

for $n = 0, 1, \ldots$ and

$$p = \frac{\alpha}{\alpha + \xi}.$$

The mean and variance are

$$E(N) = \xi, \quad \text{var}(N) = \xi(1 + \xi/\alpha),$$

which suggests that the negative binomial distribution becomes Poisson in the limit as $\alpha \to \infty$. This is indeed the case, and is due to the interpretation of the two parameters $\alpha$ and $\xi$ of the gamma model; see A.8 below.

The definition of the negative binomial model as a mixture of two experiments tells us how it can be sampled:

Algorithm A.3 Sampling Pascal
1. Draw $\lambda^* \sim \text{gamma } (\xi, \alpha)$. %see Section A.8.
2. Draw $N^* \sim \text{Poisson } (\lambda^*)$. %see Section A.4
3. Return $N^*$.
A.4 Binomial

The binomial model is also sometimes applied in insurance and finance. The density function of the number of successes out of $M$ tries is the well-known

$$f(n) = \binom{M}{n} p^n (1-p)^{M-n},$$

for $n = 0, 1, \ldots, M$ and $0 < p < 1$. The mean and variance are

$$E(N) = Mp, \quad \text{var}(N) = Mp(1-p).$$

The interpretation of $p$ as the success probability provides a recipe for how sampling can be carried out:

Algorithm A.4 Sampling binomial

1. $N^* \leftarrow 0$
2. Repeat $M$ times
3. Draw $U^* \sim $ uniform $(0, 1)$
4. If $U^* < p$, then $N^* \leftarrow N^* + 1$
5. Return $N^*$.

A.5 Normal

The normal is the only model which will be extended to the vectorial case. The p.d.f of the univariate normal is well known

$$f(z) = (2\pi\sigma^2)^{-1/2} \exp\{- (z - \mu)^2 / (2\sigma^2)\},$$

for all $z$. The parameters $\mu$ and $\sigma$ have the interpretation

$$E(Z) = \mu, \quad \text{var}(Z) = \sigma^2.$$

The inverse of the distribution function of a normal distribution is not a simple function to compute, although very good approximations are available, which makes inversion feasible. Another possibility is the Box-Muller algorithm:

Algorithm A.5 Sampling normal

1. Draw $U^*, V^* \sim $ uniform $(0, 1)$
2. $E^* \leftarrow \sqrt{\log(1/U^*)}$
3. $Z_1^* \leftarrow E^* \cos(2\pi V^*)$
4. $Z_2^* \leftarrow E^* \sin(2\pi V^*)$
5. Return $Z_1^*$ and $Z_2^*$.

On return $Z_1^*$ and $Z_2^*$ are two independent normal variables with zero mean and unit standard deviation.

The vectorial version is most conveniently defined by linear transformations. Let $\mu = (\mu_1, \ldots, \mu_J)$ be the vector of means and $C$ a $J \times J$ matrix. Then

$$\mathbf{Y} = \mu + C\mathbf{Z},$$

for a vector $\mathbf{Z} = (Z_1, \ldots, Z_J)$ of independent, normal variables with zero mean and unit standard deviation. Then $\mathbf{Y}$ is a dependent normal vector with mean $\mu$ and covariance matrix $\Sigma = CC^\top$. Note that it is $\Sigma$ (and not $C$) that is estimated from data, but we can compute $C$ by the Cholesky decomposition of matrices (see Press et al., 1986) for given $\Sigma$ and sample a normal vector directly from the definition, as follows:

Algorithm A.6 Sampling normal vectors

1. Compute $C$ from $\Sigma$. % Cholesky here.
2. Sample $J$ independent normal variables $(0, 1)$ and store in $\mathbf{Z}^*$.
3. Return $\mathbf{Y}^* = \mu^* + C\mathbf{Z}^*$.

Step 1 is called only once with repeated drawings.

A.6 Normal inverse Gaussian

The normal inverse, Gaussian (NIG) family of distributions runs, like the ordinary normal, over the whole axis. The model is able to portray many different types of shapes, but still, remarkably, possesses the convolution property, i.e. that sums of independent NIG variables remain NIG distributed. This does not apply in full generality. It is true when the NIG variables have identical distribution.

The probability density function is the complicated expression

$$f(z) = \sigma^{-1} f_0 ((z - \mu) / \sigma)$$

where

$$f_0 (z) = \alpha \pi^{-1} \exp(\sqrt{\alpha^2 - \beta^2 + \beta z}) \times K_1 (\alpha \sqrt{1 + z^2}) / \sqrt{1 + z^2},$$

and where $K_1$ is the so-called Bessel function of the first kind, i.e.

$$K_1 (z) = z \int_1^\infty \exp(-zt) \sqrt{1 - t^2} dt.$$
similar ones for the normal, whereas the remaining two, which must satisfy $\alpha > 0$ and $|\beta| < \alpha$, captures many different shapes; for example the normal as $\alpha \to \infty$, heavy-tailed distributions as $\alpha \to 0$ and skewed ones as $\beta$ is raised from 0 in either direction. The mean and variance are

$$E(Z) = \mu + \sigma \frac{\alpha/\beta}{\{1 - (\beta/\alpha)^2\}^{1/2}},$$

$$\text{var}(Z) = \sigma^2 \frac{\alpha^{-1}}{\{1 - (\beta/\alpha)^2\}^{3/2}}.$$  

It is of little importance that the expression for the density is a horrible mathematical expression. What does matter are the properties of the model, and that it can be sampled comparatively easy.

Suppose $Z_1$ and $Z_2$ are independently NIG-distributed with parameters $\mu_1, \sigma_1, \alpha_1, \beta_1$ and $\mu_2, \sigma_2, \alpha_2, \beta_2$. If $\sigma_1/\alpha_1 = \sigma_2/\alpha_2$, then the sum $Z_1 + Z_2$ is also NIG distributed, now with parameters $\mu_1 + \mu_2, \sigma_1 + \sigma_2, \alpha_1 + \alpha_2$ and $\beta_1 + \beta_2$. This is a useful property when using the model to describe fluctuations in share prices on log-scale. Another point of interest is tail behavior, which is close to, but not entirely equal to, the decay of an ordinary exponential distribution.

Sampling is carried out as shown in Algorithm A.7.

Algorithm A.7 Sampling NIG
0. Input: $\eta = (\alpha - \beta)^{-1/2}$.
1. Draw $Y^*_\eta \sim \text{normal}(0, 1)$
2. $W^* \leftarrow (Y^*_\eta)^2$
3. $D^* \leftarrow \eta \{1 + (\eta W^*/2) - \{\eta W^* + (\eta W^*)^2/4\}^{1/2}\}$
4. $V^* \leftarrow \eta^2/D^*$
5. Draw $U^* \sim \text{uniform}(0, 1)$
6. If $U^* < \eta/(\eta + D^*)$, then $V^* \leftarrow D^*$
7. Draw $Y^*_\eta \sim \text{normal}(0, 1)$
8. $Z^* \leftarrow \mu + \sigma(\beta V^* + \sqrt{V^*} Y^*_\eta)$
9. Return $Z^*$.

A.7 Log-normal

The log-normal model is defined by its relationship to the normal; i.e. $Z$ is log-normal $(\mu, \sigma)$ if $\log(Z)$ is normal $(\mu, \sigma)$. The expression for the probability density function is

$$f(z) = \frac{2\pi \sigma^2}{-\log(2\pi\sigma^2)} \frac{1}{z} \exp \{-\frac{1}{2} \{Q(z)\}\},$$

$$Q(z) = (\log z - \mu)^2/(2\sigma^2).$$

The mean and variance are

$$E(Z) = \exp(\mu + \sigma^2/2),$$

$$\text{var}(Z) = \exp(2\mu + \sigma^2)\{\exp(\sigma^2) - 1\}.$$  

Sampling is easily carried from the definition:

Algorithm A.8 Sampling log-normal
1. Draw $Z^* \sim \text{normal}(\mu, \sigma)$
2. $Z^* \leftarrow \exp(Z^*)$
3. Return $Z^*$.

A.8 Gamma

The gamma distribution has density

$$f(z) = \frac{(\alpha/\xi)^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \exp(-z/\xi),$$

for $z > 0$ and where $\alpha$ and $\xi$ are positive parameters\(^8\). The mean and variance are

$$E(Z) = \xi, \quad \text{var}(Z) = \xi^2/\alpha,$$

and this yields an immediate interpretation of the two parameters (and, incidentally, explains why we have chosen to redefine the parameters compared to those usually found in the literature).

The gamma model is closed under convolution in a way analogous to the NIG distribution. Suppose $Z_1$ and $Z_2$ are independently gamma distributed with parameters $(\alpha_1, \xi_1)$, and $(\alpha_2, \xi_2)$ respectively and suppose $\xi_1/\alpha_1 = \xi_2/\alpha_2$. Then the sum $Z_1 + Z_2$ is also gamma distributed, but now with parameters $\xi_1 + \xi_2$ and $\alpha_1 + \alpha_2$. The general gamma distribution is somewhat intricate to sample, and we refer to Devroye (1986) for algorithms.

A.9 Pareto

The Pareto distribution is the most heavy-tailed of all the models mentioned. It is therefore a conservative choice when evaluating risk, and was for this reason chosen for many of the numerical experiments. Its density is

$$f(z) = \frac{(\alpha/\theta)/\{1 + (z/\theta)\}^{\alpha+1}}{z},$$

for $z > 0$ and positive parameters $\alpha$ and $\theta$. The mean and variance are

$$E(Z) = \theta/(\alpha - 1),$$

$$\text{var}(Z) = \theta^2 \frac{\alpha}{(\alpha - 1)(\alpha - 2)}.$$  

\(^8\)The function $\Gamma(\alpha)$ is the so-called gamma function.
which are finite only if $\alpha > 2$ ($\alpha > 1$ for the mean).

Sampling is easy by the inversion, which yields the following algorithm:

**Algorithm A.9 Sampling Pareto**
1. Draw $U^* \sim \text{uniform } (0, 1)$.
2. $Z^* \leftarrow \theta \{ \exp \{ -\log(U^*)/\alpha \} - 1 \}$.
3. Return $Z^*$. 