1 Monte Carlo thinking and technique

1.1 Introduction

This book sees simulation as a tool on par with mathematical modelling. To get started we need both, and this chapter introduces both. The modelling part, though sketchy, is enough to see us through a lot of problems in the next chapter. There is also a deliberate thought behind the manner in which models are presented. The emphasis is on how they are simulated in the computer, not on their probabilistic description. This is the **constructive** approach where mathematics is developed in the way it is being used. One of the advantages is that we can move quicker beyond the most elementary. There will be more on the probabilistic side of things in Parts II and III.

Why Monte Carlo is such an important problem-solving tool was indicated in Chapter 1. Here is the same argument phrased in a more abstract way. Typically a risk variable X is made up of several (or many) random contributions. If so, it is usually hard, or even practically impossible, to find its density function $f(x)$ or distribution function $F(x)$ through mathematical deductions. That applies even when the random mechanisms involved are simple to write down and fully known. Here is where Monte Carlo comes in. By generating simulations $X_1^*, \ldots, X_m^*$ in the computer the distribution of X is *approximated*. The first we must learn is how to pass from such a random sample to statements on X and what error that brings. These issues are completely detached from the concrete situation and are best discussed at a general level.

We start there (Section 2.2). Next comes construction and design of the simulation experiments themselves. At the bottom is the notion of **uniform** random variables, in this book designated by the letter U. Every value between 0 and 1 is then equally likely, meaning that the density function is a horizontal straight line over the interval $(0, 1)$ or (equivalently) that $\Pr(U \leq u) = u$. A Monte Carlo simulation X^* is a transformation of an independent, computer-generated sample U_1^*, U_2^*,... of such uniforms. In mathematically terms

$$X^* = H(U_1^*, U_2^*, \ldots) \tag{1.1}$$

where the function H is some mathematical expression or merely command lines in the computer. The number of U_i^* may be very large indeed, sometimes even random (then determined by development). Computer software contains procedures for generating uniform random variables, and we might skip how it is done. Still, the issue is *not* without practical relevance and sometimes leads to worthwhile gain in computer time. The generation of uniform random numbers is treated in Chapter 4.

But do we have to be so basic? Sampling from the Gaussian and many other distributions are available in software packages. Can't we ignore their theory and proceed directly to how they are used? A lot of work *can* be satsfactory carried out that way, yet sampling algorithms should be studied. Otherwise we would be at the mercy of what software vendors have chosen to implement. Consider large claims in property insurance. A popular model is the the Pareto distribution (you see why in Chapter 9), but sampling software isn't routinely available. We have to know how ourselves. Then there is computational speed. Software packages have a tendency to run slowly. By writing a program in, say the C language, you may easily enhance speed by a factor of ten and more and even very much more if you invoke **quasi-randomness** (Section 4.7). Advantages: Larger problems can be tackled. Money is saved if we can get around on one of the cheap compilers.

1.2 How simulations are used

Introduction
Let X be the risk variable of interest. Typical quantities sought are expectation, standard deviation and also percentiles and the probability density function. The objective of this section is to show how these quantities are deduced from simulations $X_1^*, \ldots, X_m^*$, what error that entails and how the sample size m is determined. We draw on statistics, using the *same* methods with the same error formulas as for historical data. The simulation experiments below have useful things to say about error in ordinary statistical estimation too.

Mean and standard deviation
Let $\xi = E(X)$ be expectation and $\sigma = \mathrm{sd}(X)$ the standard deviation (or **volatility** of X. Their Monte Carlo estimates are sample average and sample standard deviation

$$\bar{X}^* = \frac{1}{m}(X_1^* + \ldots + X_m^*) \qquad \text{and} \qquad s^* = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(X_i^* - \bar{X}^*)^2}. \tag{1.2}$$

The statistical properties of the sample mean are the well-known

$$E(\bar{X}^* - \xi) = 0 \qquad \text{and} \qquad \mathrm{sd}(\bar{X}^*) = \frac{\sigma}{\sqrt{m}}, \tag{1.3}$$

and Monte Carlo estimates of ξ are unbiased. In theory error may be pushed below any prescribed level by raising m. An estimate of $\mathrm{sd}(\bar{X}^*)$ is $s^*/\sqrt{m}$ where σ in (1.3) right has been replaced by s^*. This kind of uncertainty is often of minor importance compared to other sources of error; see Chapter 7. If $\bar{X}^*$ is a price, people may require high Monte Carlo accuracy nevertheless.

The statistical properties of s^* may be less elementary than those of the sample mean, yet they are simple enough. Approximately

$$E(s^* - \sigma) \doteq 0 \qquad \text{and} \qquad \mathrm{sd}(s^*) \doteq \frac{\sigma}{\sqrt{2m}}\sqrt{1 + \kappa/2}, \tag{1.4}$$

where κ is the **kurtosis** of X, see Exercise 2.2.6 (and also Appendix A) for the definition. This result will be useful in Chapter 5 when the volatility of financial variables is estimated from historical data. For normal variables $\kappa = 0$. The approximations (1.4) are 'asymptotic' (become correct as $m \to \infty$) and are the start of mathematical series in powers of $1/\sqrt{m}$, the next term being of size $1/m$; see Hall (1992). Such large sample results apply excellently to Monte Carlo experiments where m is large.

Example: Financial returns
Let us examine how this machinery works in a transparent situation where it is not needed. Sample mean and sample standard deviation calculated from m Gaussian simulations have in Figure 2.1 been plotted against m. The true values were $\xi = 0.5\%$ $\sigma = 5\%$ (which could be monthly returns from equity investments). All experiments were completely redone with new simulations for each m. That is why the curves jump so irregularly around the straight lines representing the true values.

The estimates tend to ξ and σ as $m \to \infty$. That we knew. But actually the experiment tells us something else. In relative terms *the sample mean is less accurately estimated than the volatility.*

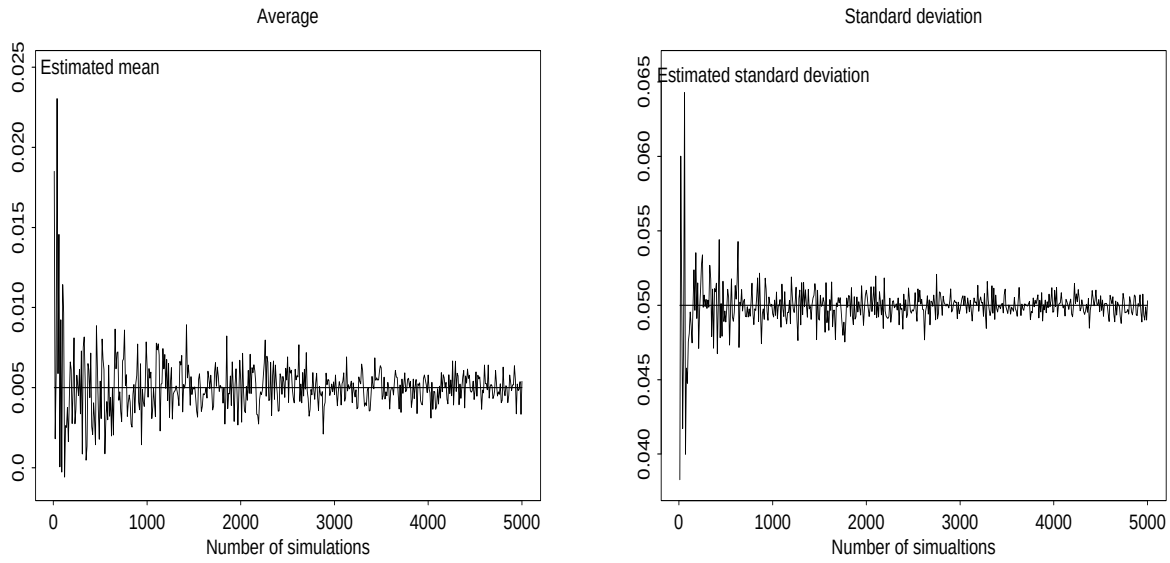

Figure 2.1 *Sample mean and standard deviation against the number of simulations for a Gaussian model. Straight lines are the true parameters.*

Suppose the simulations had been historical returns of equity instead with mean expectation and volatility estimated from them. Then, after 1000 months (about eighty years, a very long time) the relative error in the sample mean is still, perhaps, two thirds of the true value! Errors of that size would have a degrading effect on our ability to evaluate financial risk and makes the celebrated Markowitz theory of optimal investment in Chapter 5 much harder to use. When financial derivatives are discussed in Section 3.5 (and Chapter 14), it will emerge that the Black-Scholes-Merton theory removes these parameters from the pricing formulas, doubtless one of the reasons for their success.

This is an elementary case, and the main conclusion can be taken from the mathematics as well. Indeed, from the left hand side of (1.3) and (1.4)

$$\frac{\mathrm{sd}(\bar{X}^*)}{\xi} = \frac{\sigma}{\xi}\,\frac{1}{\sqrt{m}} \qquad \text{and} \qquad \frac{\mathrm{sd}(s^*)}{\sigma} = \sqrt{1/2 + \kappa/4}\,\frac{1}{\sqrt{m}}$$

and the coefficient σ/ξ is more than ten times larger than the one on the right (for which $\kappa = 0$). This explains why the estimate $\bar{X}^*$ is so inaccurate. In finance it is nearly always like that.

Percentiles

The percentile q_ϵ is the solution of either either of the equations

$$F(q_\epsilon) = 1 - \epsilon, \qquad\qquad F(q_\epsilon) = \epsilon,$$

depending on whether the upper or the lower version is sought. With insurance risk it is typically the former, in finance the latter. In either the case the Monte Carlo approximation is obtained by ordering the simulations. For the *upper percentile* the estimate is

$$q^*_\epsilon = X^*_{(\epsilon m)} \qquad \text{for} \qquad X^*_{(1)} \geq \ldots \geq X^*_{(m)}. \tag{1.5}$$

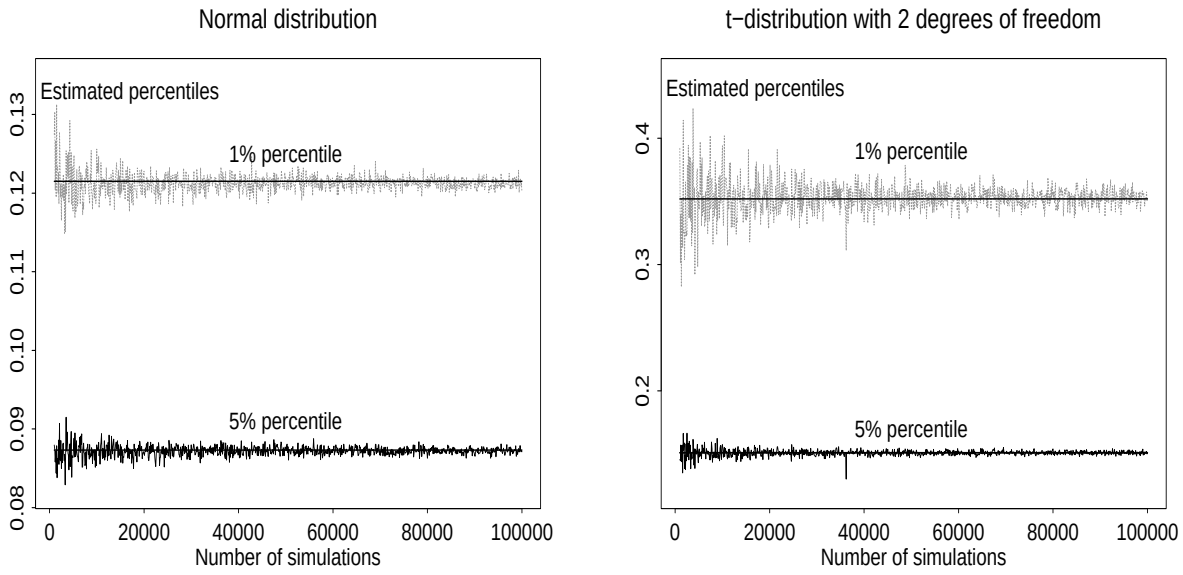

Figure 2.2 *Estimated percentiles of simulated series against the number of simulations.*

The lower one is exactly the same, except that the ranking on the right now is in ascending order. A useful, approximate expression for the error is available. Indeed

$$E(q_\epsilon^* - q_\epsilon) \doteq 0, \qquad \mathrm{sd}(q_\epsilon^*) \doteq \frac{a_\epsilon}{\sqrt{m}}, \quad a_\epsilon = \frac{\sqrt{\epsilon(1-\epsilon)}}{f(q_\epsilon)}, \tag{1.6}$$

which are again asymptotic results as $m \to \infty$. It is possible to evaluate $f(q_\epsilon)$ through density estimation (see below) and insert the estimate into (1.6) for a numerical estimate of $\mathrm{sd}(q_\epsilon^*)$.

The standard deviation depends on both the level ϵ and on the underlying density function. As ϵ is lowered, the value of a_ϵ increases drastically to the extent that

$$a_\epsilon \to \infty \quad \text{as} \quad \epsilon \to 0; \tag{1.7}$$

see Section 2.7 for the proof. *Very* many more simulations are needed for small ϵ far out into the tails of the distribution. That is no more than common sense. Perhaps it is less obvious that heavy-tailed distributions require more simulations for the same accuracy than light-tailed ones. A precise result is the following. Let $q_{1\epsilon}$ and $q_{2\epsilon}$ be percentiles under two different density functions $f_1(x)$ and $f_2(x)$ and suppose the second one has heavier tails. This means that

$$\frac{q_{2\epsilon}}{q_{1\epsilon}} \to \infty \quad \text{as} \quad \epsilon \to 0. \tag{1.8}$$

This yields (see Section 2.7 for the proof)

$$\frac{f_2(q_{2\epsilon})}{f_1(q_{1\epsilon})} \to 0 \quad \text{as} \quad \epsilon \to 0 \qquad \text{so that} \qquad \frac{a_{2\epsilon}}{a_{1\epsilon}} = \frac{f_2(q_{2\epsilon})}{f_1(q_{1\epsilon})} \to \infty \ \text{ as } \ \epsilon \to 0, \tag{1.9}$$

where $a_{i\epsilon} = \sqrt{\epsilon(1-\epsilon)}/f_i(q_{i\epsilon})$ are coefficients similar to a_ϵ in (1.6). Its right hand side establishes that the *second* and more heavy-tailed distribution demands more simulations.

Financial returns again
The experiments in Figure 2.1 have been repeated in Figure 2.2, but this time with the percentiles obtained from (1.5). Mean and volatility were (as before) 0.5% and 5%, and the exact values are the straight lines. The simulations are Gaussian on the left and the very heavy-tailed t-distribution with 2 degrees on the right; see Section 2.3 below for the precise definition of the latter.

Earlier assertions on Monte Carlo error are confirmed. The discrepancies are larger for $\epsilon = 1\%$ than for $\epsilon = 5\%$, and they are strongly inflated for the heavy-tailed distribution on the right. To appreciate these results, note the highly unequal scale of the vertical axes of the two figures which is the reason why the 5% curve on the right looks less spread out than the one on the left is. We have learned that *the smaller ε and the heavier tails, the higher number of simulations.*

Density estimation
Then there is the issue of estimating the density function $f(x)$ itself. The simplest way may be to read the simulations $X_1^*, \ldots, X_m^*$ into statistical software. But even then an idea of how such techniques operate is useful, all the more since there is a parameter to adjust. All plots of density functions in this book are obtained through the **Gaussian kernel** method where we choose a smoothing parameter $h > 0$ and use as estimate

$$f^*(x) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{\delta}\varphi\left(\frac{x - X_i^*}{\delta}\right) \qquad \text{where} \qquad \delta = hs^*. \tag{1.10}$$

Here $\varphi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ is the standard Gaussian density. As x is varied the estimate (1.10) traces out a curve which is the average of m Gaussian densities with standard deviation δ centered at the m simulations X_i^*. The statistical properties of the estimate, as derived in Chapter 2 in Wand and Jones (1995) is

$$E\{f^*(x) - f(x)\} \doteq \frac{1}{2}h^2 f''(x), \qquad \text{and} \qquad \mathrm{sd}\{f^*(x)\} \doteq 0.4466\sqrt{\frac{f(x)}{hm}}, \tag{1.11}$$

where $f''(x)$ is the second derivative. The estimate is biased! The choice of h is compromise between bias on the left (going *down* with h) and random variation on the right (going *up*). Usually commercial software is equipped with a sensible default value. In theory the choice depends on m, the 'best' value being proportional to the *fifth* root!

The curve $f^*(x)$ will contain random bumps if h is too low. That emerges clearly on the left in Figure 2.3 showing estimates based on $m = 1000$ simulations from the density function

$$f(x) = \frac{1}{2}x^2\exp(-x).$$

The estimates become smoother with the higher values of h on the right, but now the bias tend to drag the estimates away from the true function. It may for many purposes not matter too much if h is selected a little too low. Perhaps $h = 0.2$ is a suitable choice in Figure 2.3. A sensible general rule of the thumb for h could be the range $0.05 - 0.30$, but, as remarked above, it also depend on m. Other kernels than the Gaussian one can be used; see Wand and Jones (1995) or Scott (1992) for monographs on density estimation.

Monte Carlo error and selection of m

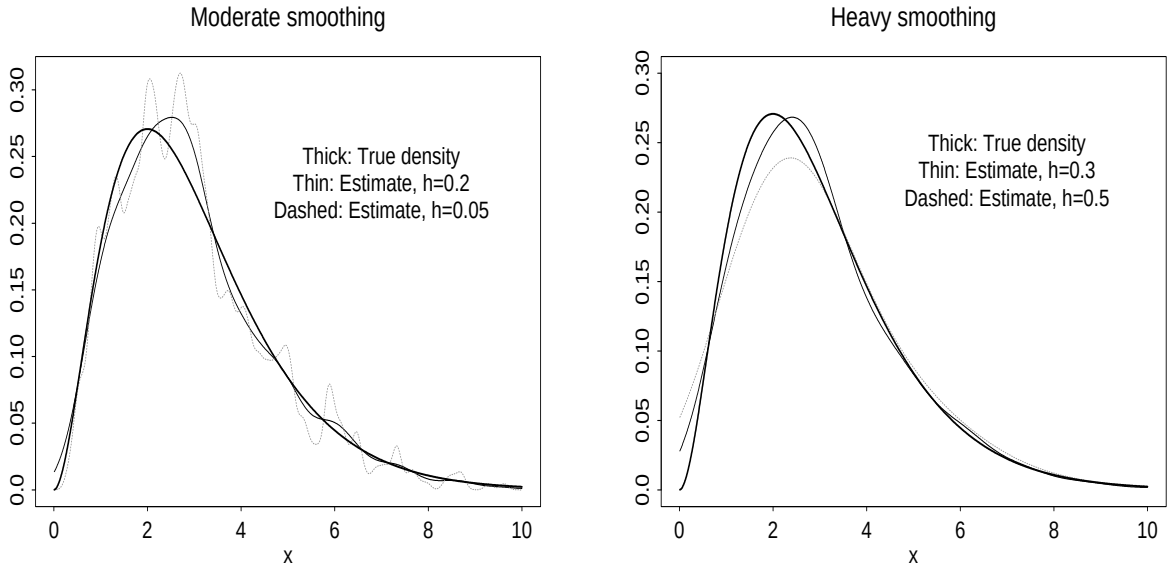

Figure 2.3 *Kernel density estimates based on* 1000 *simulations from model in the text, shown as the thick solid line in both plots.*

The discrepancy between a Monte Carlo approximation and its underlying, exact value nearly always becomes Gaussian as $m \to \infty$. For the sample mean this follows from the central limit theorem. Standard statistical large sample theory yields the result in most other cases; see the reading list at the end of the chapter. Thus, a Monte Carlo evaluation ψ^* of some quantity ψ is roughly Gaussian with mean ψ and standard deviation of the form $a/\sqrt{m}$, where a is a constant. That applied to all the examples above, except the density estimate (there is still a theory, but its details are different; see Scott, 1992). We also saw how an estimate a^* of a could be obtained from the simulations. The interval

$$\psi^* - 2\frac{a^*}{\sqrt{m}} < \psi < \psi^* + 2\frac{a^*}{\sqrt{m}} \tag{1.12}$$

contains ψ with approximately 95% confidence[1] that can be reported as a formal appraisal of Monte Carlo error. Here $a^* = s^*$ when ψ is he mean and $a^* = (s^*/\sqrt{2})\sqrt{1 + \kappa^*/2}$ for the standard deviation; see (1.3) and (1.4) right.

Such results can also be used for design. Suppose Monte Carlo standard deviation is required be below some level σ_0. If the equation $a^*/\sqrt{m} = \sigma_0$ is solved for m, we get as lower bound

$$m = \left(\frac{a^*}{\sigma_0}\right)^2; \tag{1.13}$$

i.e. the number of simulations must be at least that. For the idea to work you need the estimate a^*. Often the only way is to run a preliminary round of simulations, estimate ζ, determine m and complete the additional samples you need. That approach is a standard one with clinical trials in medicine! With some programming effort it is possible to automatize the process so that the computer takes care of it on its own.

[1]The precise 2.5% percentiles of the normal has been rounded off from 1.96 to 2.

1.3 Gaussian based sampling and modelling

Introduction

The **Gaussian** (or **normal**) model is the most famous of all probability distributions, arguably the most important one too. It is familiar from introductory courses in statistics, yet introduced from scratch below. The goal of this section is to review a main tool of probabilistic modelling, but it also an introduction to constructions in terms of stochastic representations. Defining models the way they are simulated in the computer has several advantages. One of them is that elementary default versions are so easy to extend. The present section is a case study in this line of thinking where stochastic volatilites, heavy tails, correlated variables are hung on an original defintion of Gaussian variables. Later (Chapter 5 and 13) further development introduces general normal variables and time-dependent (or **dynamic**) phenomena affecting expectations, volatilties and correlations alike.

The normal family

A normal random variable with mean ξ and standard deviation σ may be written

$$X = \xi + \sigma\varepsilon, \qquad \varepsilon \sim N(0,1), \tag{1.14}$$

where ε is the standard normal distribution, denoted $N(0,1)$. The Gaussian family of models is defined by varying $\xi = E(X)$ and $\sigma = \mathrm{sd}(X)$. Simulation is by means of $X^* = \xi + \sigma\varepsilon^*$, and the problem is how to generate ε^*. Let $\Phi(x)$ be the Gaussian integral[2] and $\Phi^{-1}(u)$ its inverse function (same as q_u for *lower* percentiles). It will be proved in Section 2.4 that ε can be represented as

$$\varepsilon = \Phi^{-1}(U), \qquad U \sim \text{uniform}, \tag{1.15}$$

and Gaussian variables can be sampled by combining (1.14) and (1.15). In summary:

Algorithm 2.1. Gaussian generator
0 Input: ξ and σ
1 Generate $U^* \sim$ uniform
2 Return $X^* \leftarrow \xi + \sigma\Phi^{-1}(U^*)$ *%Or $\Phi^{-1}(U^*)$ replaced by ε^* generated by software directly*

For this to be practical we must have a quick way to calculate $\Phi^{-1}(u)$. *Very* accurate and simple approximations are available; see Appendix C. For Jäckel (2002) this is actually the recommended method for Gaussian sampling.

Modelling on logarithmic scale

Models constructed on logarithmic scale are common. Returns on equity (Section 1.3) is case in point, the standard model being

$$\log(1+R) = \xi + \sigma\varepsilon, \qquad \text{or} \qquad R = \exp(\xi + \sigma\varepsilon) - 1, \tag{1.16}$$

where $\varepsilon \sim N(0,1)$. Another example is the size of claims in property insurance, in this book denoted by Z. The model now reads

$$\log(Z) = \xi + \sigma\varepsilon, \qquad \text{or} \qquad Z = \exp(\xi + \sigma\varepsilon). \tag{1.17}$$

[2]Defined as

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2)dy.$$

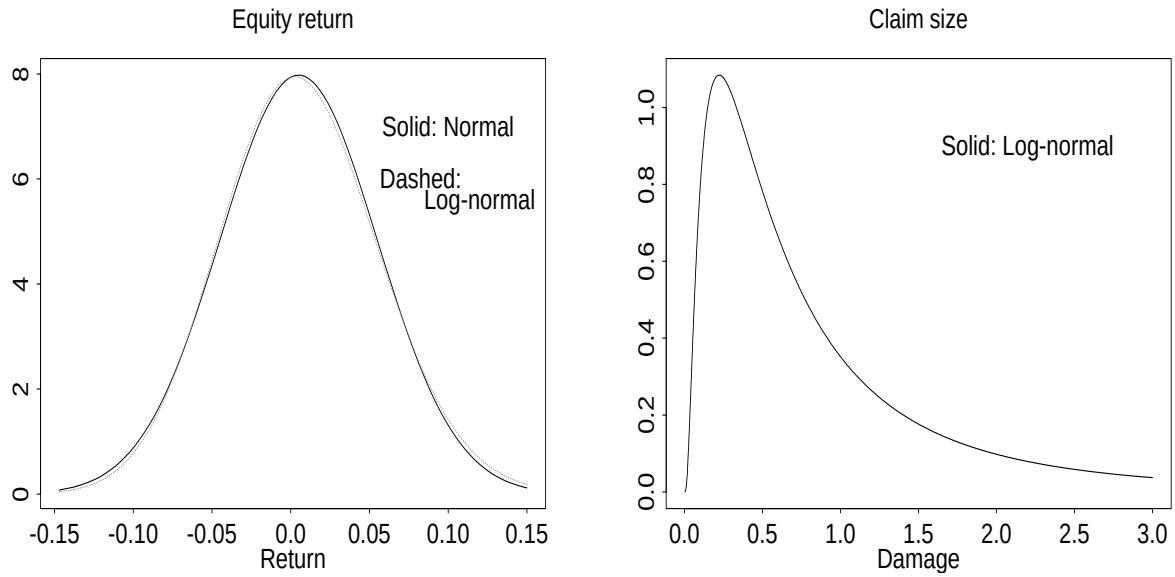

Figure 2.4 *Left: Normal and log-normal density functions for* $\xi = 0.005$*,* $\sigma = 0.05$*. Right: Log-normal for* $\xi = -0.5$*,* $\sigma = 1$*.*

Formulae for means and standard deviations are among the most important in the entire theory of risk. Indeed

$$E(R) = \exp(\xi + \frac{1}{2}\sigma^2) - 1, \qquad \text{and} \qquad E(Z) = \exp(\xi + \frac{1}{2}\sigma^2), \tag{1.18}$$

and

$$\text{sd}(R) \;=\; \text{sd}(Z) \;=\; E(Z)\{\exp(\sigma^2) - 1\}^{1/2}; \tag{1.19}$$

see Section 2.7 for the proof.

Sampling is easy:

Algorithm 2.2 Log-normal sampling
0 Input: ξ, σ
1 Draw $\varepsilon^* \sim$N(0,1) %*For example:* $U^* \sim$ *uniform,* $\varepsilon^* \leftarrow \Phi^{-1}(U^*)$
2 Return $R^* \leftarrow \exp(\xi + \sigma\varepsilon^*) - 1$, or $Z^* \leftarrow \exp(\xi + \sigma\varepsilon^*)$.

The models (1.16) and (1.17) are called **log-normal**. Mathematical expressions for their density function *can* be derived (you will find them in Appendix A). Two examples are plotted in Figure 2.4. Note the pronounced difference from left to right. Small σ (on the left) is appropriate for finance and yields a distribution close to the normal model, as predicted in Section 1.3. Higher values of σ (on the right) leads to pronounced skewness, as is typical for large claims in property insurance.

Stochastic volatility
Financial risk is in many situations better described by introducing a separate stochastic model for σ. This means that (1.14) is extended to

$$X = \xi_x + \sigma\varepsilon \qquad \text{where} \qquad \sigma = \xi_\sigma\sqrt{Z} \tag{1.20}$$

for a *positive* random variable Z. Typically Z is scaled so that $E(Z) = 1$ or $E(Z^2) = 1$. Then $\xi_\sigma = E(\sigma)$ or $\xi_\sigma = E(\sigma^2)$ making Z responsible for random fluctuations around a mean value. Since we are now dealing with two expectations, that of X has now been written ξ_x.

The effect of such a stochastic standard deviation or volatility is to make the tails of the distribution of X heavy. Why is that? Because the possibility of a very small/large ε *and* a very large Z *jointly* must lead to higher discrepancy from the mean ξ_x than the normal can portray on its own. Such models have drawn much interest in finance, and a *dynamic* version where σ is linked to earlier values will be introduced in Chapter 13. Sampling is an extension of Algorithm 2.1:

Algorithm 2.3 Gaussian with stochastic volatility
0 Input: ξ, σ_0, model for Z
1 Draw Z^* and $\sigma^* \leftarrow \xi_\sigma \sqrt{Z^*}$ *%Many possibilities for Z^*; see text*
2 Generate $U^* \sim$ uniform.
3 Return $X^* \leftarrow \xi_x + \sigma^* \Phi^{-1}(U^*)$ *%Or $\Phi^{-1}(U^*)$ replaced by ε^* generated by software directly*

The most common choice for Z is

$$Z = 1/Y,$$

where Y is Gamma variable with mean 1; see Section 2.5. Now X follows a **t-distribution** (Chapter 13). The example in Figure 2.2 right was run with $Y = -\log(U)$, which is an exponential distribution (see Section 2.5). This is a very strong form of stochastic volatility. Even *daily* equity returns typically have lighter tails than this.

Dependent normal pairs
Many situations demand normal variables that are correlated. Such models can for pairs X_1 and X_2 be built up by applying (1.14) twice. We then write

$$\begin{array}{ll} X_1 = \xi_{x1} + \sigma_1 \varepsilon_1 & \\ X_2 = \xi_{x2} + \sigma_2 \varepsilon_2. & \text{where} \end{array} \qquad \begin{array}{l} \varepsilon_1 = \eta_1 \\ \varepsilon_2 = \rho \eta_1 + \sqrt{1-\rho^2}\,\eta_2, \end{array} \tag{1.21}$$

and the sub-model on the right is the new feature. Here η_1 and η_2 are independent and $N(0,1)$ variables, and they produce a second pair ε_1 and ε_2 also $N(0,1)$, but now linked through η_1 influencing both. (for ε_2 a slight argument is needed; see Appendix A). The co-variation between ε_1 and ε_2 is controlled by the parameter ρ which concides with the ordinary correlation coefficient (Section 5.4). Simulation is straightforward. Generate η_1^* and η_2^* by Gaussian sampling and insert them for η_1 and η_2 in (1.21); see Algorithm 2.4 below for a formal organisation of commands.

The model provides one of the the most popular descriptions of pairs of equity returns R_1 and R_2. Using the log-normal we then take

$$R_1 = \exp(X_1) - 1, \quad R_2 = \exp(X_2) - 1,$$

where X_1 and X_2 are correlated Gaussians as above. Simulations of (R_1, R_2) based on $\xi_{x1} = \xi_{x2} = 0.5\%$ and $\sigma_1 = \sigma_2 = 5\%$ (monthly returns on equity) have been plotted in Figure 2.5. Variation of ρ captures different degrees of co-variation.

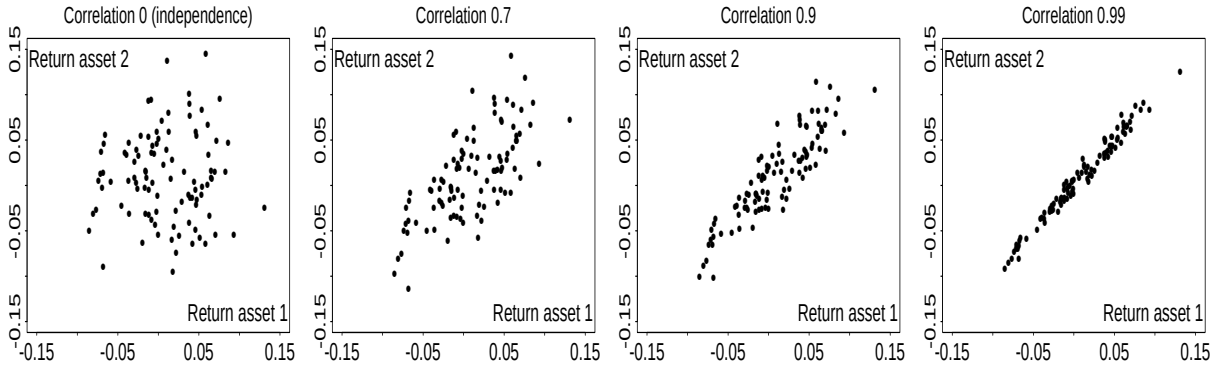

Figure 2.5 *Joint plot of* 100 *pairs of simulated equity returns; from the ordinary log-***normal** *model described in the text.*

Dependence *and* heavy tails

Returns of equity investments may be *both* dependent and heavy-tailed. Can that be handled? Easily! We simply combine (1.20) and (1.21), rewriting the latter as

$$\begin{aligned} X_1 &= \xi_{x1} + \sigma_1 \eta_1, & \sigma_1 &= \xi_{\sigma 1}\sqrt{Z_1} \\ X_2 &= \xi_{x2} + \sigma_2(\rho\eta_1 + \sqrt{1-\rho^2}\,\eta_2), & \sigma_2 &= \xi_{\sigma 2}\sqrt{Z_2}. \end{aligned} \tag{1.22}$$

Here $\xi_{\sigma 1}$ and $\xi_{\sigma 2}$ are fixed parameters and Z_1 and Z_2 are *positive* random variables playing the same role as Z in (1.20).

It is common to take $Z_1 = Z_2 = Z$ assuming fluctuations in σ_1 and σ_2 to be in perfect synchrony. The shape of the density functions of X_1 and X_2 must then be equal and non-normal to exactly the same degree. This has no special justification beyond mathematical convenience, but it does give joint density function of (X_1, X_2) a 'nice' mathematical form, see Appendix A. Exercise 2.4.5 play with an alternative. The effect on financial returns has been indicated in Figure 2.6 which has been set up from the same model as in Figure 2.5 except that now

$$Z_1 = Z_2 = 1/\{-\log(U)\}.$$

What is the change brought by stochastic volatility? When you take into account that *axes scales are almost tripled* compared to what they were in Figure 2.5, it becomes clear that strongly deviating returns has become much more frequent. By contrast the degree of dependence seem to have remained unchanged. It is in Exercise 5.2.7 proved that it must be so.

Equi-correlation models

Many interacting Gaussian variables start out as above by taking

$$X_j = \xi_{xj} + \sigma_j \varepsilon_j \qquad j = 1, \ldots, J \tag{1.23}$$

where $\varepsilon_1, \ldots, \varepsilon_J$ are normal $N(0,1)$. Their general formulation is a somewhat complicated issue and is dealt with in in Section 5.4. A simple special case which will be used in the next chapter is the **equi-correlation** model for which

$$\varepsilon_j = \sqrt{\rho}\,\eta_0 + \sqrt{1-\rho}\,\eta_j \qquad j = 1, \ldots, J. \tag{1.24}$$

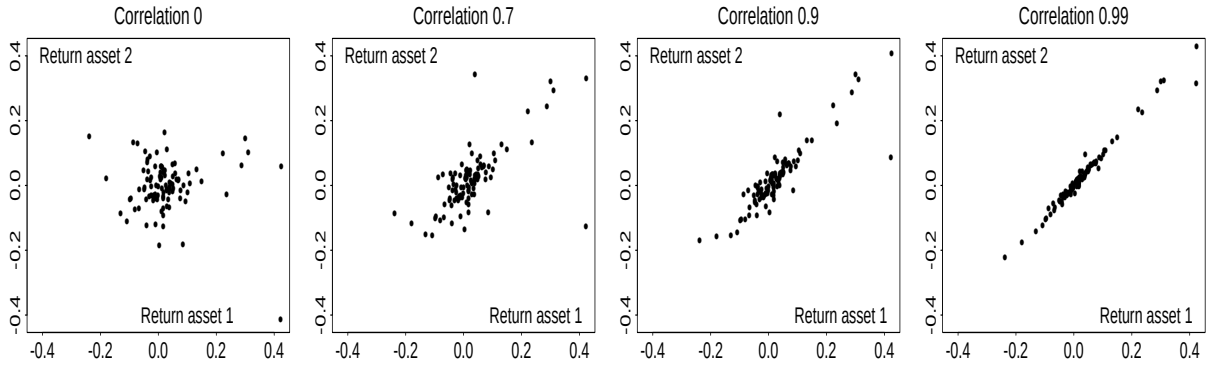

Figure 2.6 *Joint plot of* 100 *simulated financial returns; from* **stochastic volatility model** *(same as in Figure 2.5 otherwise) described in the text.*

Here $\eta_0, \eta_1, \ldots, \eta_J$ are independent and $N(0,1)$, and η_0 is responsible for relationships between all pairs of variables $(\varepsilon_i, \varepsilon_j)$. The parameter ρ (must be ≥ 0) is still a correlation coefficient, this time a common one for all pairs.

How correlated returns are generated under this model is summarized by the following algorithm:

Algorithm 2.4 Financial returns under equi-correlation

0 Input: $\xi_{x1}, \ldots, \xi_{xJ},\ \sigma_1, \ldots, \sigma_J,\ c_1 \leftarrow \sqrt{\rho},\ c_2 \leftarrow \sqrt{1-\rho}$

1 Generate $\eta_0^* \sim N(0,1)$ %*Common stochastic factor*

2 For $j = 1 \ldots, J$ do

3 Generate $\eta^* \sim N(0,1)$

4 $\epsilon^* \leftarrow c_1\eta_0^* + c_2\eta^*$ %*Randomness in j'th return*

5 $R_j^* \leftarrow \exp(\xi_{xj} + \sigma_j\varepsilon^*) - 1$ %*Stochastic volatility: Draw Z^* and let $\sigma_j^* \leftarrow \xi_{\sigma j}\sqrt{Z^*}$; use it for σ_j*

6 Return $R_1^*, \ldots, R_J^*$

How heavy-tailed models are included is indicated in the comment to Command 4. Some of the exercises at the end of the chapter play with this algorithm.

1.4 Creating sampling algorithms

Introduction

The simulation algorithms in the two preceding sections were (largely) model relationships copied in the computer. This is indeed the most common way stochastic simulation algorithms are developed and has influenced how models are presented. But we also need a toolbox of basic sampling techniques to work from. That is definitely an area for the clever, full of ingenious tricks. An example is the **Box-Muller** representation of Gaussian random variables. Suppose U_1 and U_2 are independent and uniform. Then

$$\eta_1 = \sqrt{-2\log(U_1)}\sin(2\pi U_2), \qquad \eta_2 = \sqrt{-2\log(U_1)}\cos(2\pi U_2) \tag{1.25}$$

are both $N(0,1)$ and also independent; see Devroye (1986) for a proof. This gives the Box-Muller generator:

Algorithm 2.5 Independent, normal pairs
1 Generate U_1^*, $U_2^* \sim$ uniform
2 $Y^* \leftarrow \sqrt{-2\log(U_1^*)}$
3 Return $\eta_1^* \leftarrow Y^* \sin(2\pi U_2^*)$, $\eta_2^* \leftarrow Y^* \cos(2\pi U_2^*)$

On output η_1^* and η_2^* are independent and $N(0,1)$. The algorithm is, despite its elegance, not particularly fast, but worth including for its simplicity (despite reservations in Jäckel, 2002). It is also an illustration of the inventiveness of sampling theory. Many useful procedures are ad-hoc and like the Box-Muller method adapted to concrete situations.

The intent here is not even remotely one of providing justice to the vast subject of generating random variables with given distributions. For that Devroye (1986) is still a good reference; see also Section 2.7. Our purpose is to select methods of practical usefulness in actuarial science. Actually the sampling procedures reviewed in the present chapter will take us far if we know how to apply and combine them them intelligently (more in Chapter 4). We are going to lean on two *general* techniques. The first is:

The inversion method
It was claimed above that a normal variable is generated through (1.15). This is actually a general sampling method known as **inversion**. Let $F(x)$ be a **strictly increasing** distribution function with inverse $F^{-1}(u)$. Define

$$X = F^{-1}(U) \quad \text{or} \quad X = F^{-1}(1-U), \qquad U \sim \text{Uniform}. \tag{1.26}$$

Consider the version on the left for which $U = F(X)$. Note that

$$\Pr(X \le x) = \Pr\{F(X) \le F(x)\} = \Pr\{U \le F(x)\} = F(x),$$

since $\Pr(U \le u) = u$. In other words, X defined by (1.26) left has the distribution function $F(x)$, and we have a general technique for the generating random variables. The other one based on $1-U$ is justified by U and $1-U$ having the same distribution. In summary:

Algorithm 2.6 Sampling by inversion
0 Input: The percentile function $F^{-1}(u)$
1 Draw $U^* \sim$ uniform
2 Return $X^* \leftarrow F^{-1}(U^*)$ or $X^* \leftarrow F^{-1}(1-U^*)$

In either case X^* has the desired distribution function $F(x)$. The two versions represent a so-called **antitetic** pair. It has a speed-enhancing potential that will be discussed in Chapter 4.

Whether Algorithn 2.6 is practical depends on the ease with which the percentile function $F^{-1}(u)$ can be computed. That condition is satisfied for Gaussian variables, and Algorithm 2.1 has now been justified. There are many additional examples in Section 2.5; first a *second* general technique.

The acceptance-rejection method
This is an example of a so-called **random stopping** rule and is more subtle than inversion. The idea is to *select* a density function $g(x)$ which is convenient to sample from. Simulations from $f(x)$ can still be obtained if we discard those that do not meet a certain acceptance criterion A. Magic?

It works like this. Let $g(x|A)$ be the density function of the simulations kept. By Bayes' formula (Appendix A)

$$g(x|A) = \frac{\Pr(A|x)g(x)}{\Pr(A)}, \tag{1.27}$$

and we must specify $\Pr(A|x)$, i.e. the probability that $X = x$ drawn from $g(x)$ is allowed to stand. Suppose we have been able to find a constant M such that

$$M \geq \frac{f(x)}{g(x)}, \qquad \text{all } x. \tag{1.28}$$

Let us examine what happens if x is accepted whenever a uniform random number U satisfies

$$U \leq \frac{f(x)}{Mg(x)},$$

where the right hand side is always less than one. Now

$$\Pr(A|x) = \Pr\left(U \leq \frac{f(x)}{Mg(x)}\right) = \frac{f(x)}{Mg(x)},$$

which in combination with (1.27) yields

$$g(x|A) = \frac{f(x)}{M\Pr(A)}.$$

The denominator must be one (otherwise $g(x|A)$ won't be a density function), and so

$$g(x|A) = f(x) \qquad \text{and} \qquad \Pr(A) = \frac{1}{M}. \tag{1.29}$$

We have indeed obtained the right distribution. In summary the algorithm runs as follows:

Algorithm 2.7 Rejection-acceptance sampling
0 Input $f(x)$, $g(x)$, M
1 Repeat
2 Draw $X^* \sim g(x)$
3 Draw $U^* \sim$ uniform
4 If $U^* \leq f(X^*)/Mg(X^*)$ then
stop and return X^*.

The expected number of repetitions equals $1/\Pr(A)$ and hence M by (1.29) right. Good designs are those with low M. Some of the smartest sampling algorithms in the business are of the acceptance-rejection type; notably Algorithms 2.11 and 2.12 below.

1.5 Some standard distributions

Introduction
The normal and log-normal models were reviewed above. With the four additional distributions introduced in this section they form a toolkit we shall rely on all through Part I. The presentation below is *very* sketchy, concentrating on mean and standard deviation and on how sampling is carried out. Poperties and genesis of these distributions are covered in Parts II and III where other models will be introduced too; see also some of the exercises to this section.

The Pareto distribution
Random variables X with density function

$$f(x) = \frac{\alpha/\beta}{(1+x/\beta)^{1+\alpha}}, \qquad x > 0 \tag{1.30}$$

are **Pareto** distributed. Here $\alpha > 0$ and $\beta > 0$ are positive parameters and negative values for X do not occur. The model is extremely heavy-tailed and often serves as model for large claims in property insurance; more on that in Chapter 9. Mean and standard deviation are

$$E(X) = \frac{\beta}{\alpha - 1}, \ \alpha > 1 \qquad \text{and} \qquad \mathrm{sd}(X) = E(X)\sqrt{\frac{\alpha}{\alpha-2}}, \ \alpha > 2. \tag{1.31}$$

They do not exist (i.e. is infinite) for other values of α than those shown. Real phenomena where α seems to be between 1 and 2 will be encountered later.

Distribution function and its inverse of (1.30) are

$$F(x) = 1 - (1+x/\beta)^{-\alpha}, \quad x > 0 \qquad \text{and} \qquad F^{-1}(u) = \beta\{(1-u)^{-(1/\alpha)} - 1\}, \tag{1.32}$$

where the latter is found by solving the equation $F(x) = u$. This yields the following Pareto generator from the *second* version of the inversion algorithm:

Algorithm 2.8 Pareto generator
0 Input α and β
1 Generate $U^* \sim$ uniform
2 Return $X^* \leftarrow \beta\{(U^*)^{-(1/\alpha)} - 1\}$ *%X^* Pareto distributed*

The exponential distribution
Suppose $\beta = \alpha\xi$ is inserted into the Pareto density (1.30) while ξ is kept fixed and α is allowed to become infinite. Then

$$f(x) = \frac{\xi^{-1}}{(1+(x/\xi)\alpha^{-1})^{1+\alpha}} \to \frac{\xi^{-1}}{\exp(x/\xi)}, \quad \text{as} \quad \alpha \to \infty,$$

and we have obtained the **exponential** density function

$$f(x) = \frac{1}{\xi}\exp(-x/\xi), \qquad x > 0. \tag{1.33}$$

Mean and standard deviation are

$$E(X) = \xi \qquad \text{and} \qquad \mathrm{sd}(X) = \xi, \tag{1.34}$$

and for the distribution and percentile functions we have the expressions

$$F(x) = 1 - \exp(x/\xi) \qquad \text{and} \qquad F^{-1}(u) = -\xi \log(1-u).$$

Inversion (Algorithm 2.6) yields the following sampling method:

Algorithm 2.9 Exponential generator
0 Input ξ
1 Draw $U^* \sim$ uniform
2 Return $X^* \leftarrow -\xi \log(U^*)$ $\qquad\qquad$ $\%X^*$ *exponential*

There is a connection to Algorithm 2.8. Insert $\beta = \alpha\xi$ on the last line there and let $\alpha \to \infty$. The fact that the exponential distribution is a limiting member of the Pareto family is of some importance with extremes; see Section 9.4.

The Poisson distribution
Suppose X_1, X_2,... are independent and exponentially distributed with $\xi = 1$. It can then be proved (see Section 2.7 and also Exercise 8.2.4) that

$$\Pr(X_1 + \ldots + X_n < \lambda \le X_1 + \ldots + X_{n+1}) = \frac{\lambda^n}{n!} \exp(-\lambda) \tag{1.35}$$

for all $n \ge 0$ and all $\lambda > 0$. The right hand side are **Poisson** probabilities; i.e defining the density function

$$\Pr(N = n) = \frac{\lambda^n}{n!} \exp(-\lambda), \qquad n = 0, 1, \ldots \tag{1.36}$$

This model is the central one for claim frequency in property insurance, and a lot will be said about it in Chapter 8. Its mean and variance are equal; i.e.

$$E(N) = \lambda \qquad \text{and} \qquad \text{sd}(N) = \sqrt{\lambda}. \tag{1.37}$$

The main point at the moment is that (1.35) tells us how Poisson variables are sampled. Utilize that $X_j = -\log(U_j)$ is exponential if U_j is uniform and follow the sum $X_1 + X_2 + \ldots$ until it exceeds λ, in other words:

Algorithm 2.10 Poisson generator
0 Input λ,$Y^* \leftarrow 0$
1 For $n = 1, 2, \ldots$ do
2 Draw $U^* \sim$ uniform and $Y^* \leftarrow Y^* - \log(U^*)$
3 If $Y^* \ge \lambda$ then
stop and return $N^* \leftarrow n - 1$.

This is a random stopping rule of a kind different from acceptance-rejection. We count how long it takes for (1.35) to be satisfed and return the number of trials minus one. This simple procedure is often good enough

More on Poisson sampling
Poisson counts are central in most evaluations of risk in property insurance (all in this book) and

their sampling important. Yet very often the simple Algorithm 2.10 is good enough (to see why consult Section 10.3). However, it does slow down for large λ and if speed *is* critical, we may turn to the method of Atkinson (1979) which was constructed to deal with that issue precisely:

Algorithm 2.11 Atkinson's Poisson generator
0 Input: $c \leftarrow 0.767 - 3.36/\lambda, \quad a \leftarrow \pi/\sqrt{3\lambda}, \quad b \leftarrow \lambda a, \quad d \leftarrow \log(c/a) - \lambda$
1 Repeat

2 Repeat
3 Draw $U^* \sim$ uniform and $X^* \leftarrow \{b - \log(1/U^* - 1)\}/a$ until $X^* > -0.5$

4 $N^* \leftarrow [X^* + 0.5]$ and draw $U^* \sim$ uniform
5 If $b - aX^* - \log\{\{1 + \exp(b - aX^*)\}^2/U^*\} < d + N^* \log(\lambda) - \log(N^*!)$ stop and return N^*

Before running the algorithm it is necessary compute (recursively!) and store the sequence $\log(n!)$ up to some number the Poisson variable has microsopic chances to exceed (5λ could be a sensible choice). The method is derived through rejection sampling; see Cassela and Robert (1998). Atkinson recommends that $\lambda > 30$ for his procedure to be used. Devroye (1986) contains other possibilities; see also the discrete sampling procedures in Section 4.2.

The Gamma distribution
One of the most important models in actuarial science is without doubt the **Gamma** family of distributions which plays several differerent roles. The probability density function is

$$f(x) = \frac{(\alpha/\xi)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\alpha x/\xi), \quad x > 0. \tag{1.38}$$

Here $\Gamma(\alpha)$ is the so-called **Gamma** function[3]. Mean and standard deviation are

$$E(X) = \xi, \qquad \text{and} \qquad \text{sd}(X) = \xi/\sqrt{\alpha}. \tag{1.39}$$

Following Mccullagh and Nelder (1992) expectation is one of the two parameters (often Gamma models are presented slightly different.) The case $\xi = 1$ will be called the **standard** Gamma and denoted Gamma(α).

The Gamma distribution isn't that easy to sample. Its percentile function is complicated computationally, and unlike the normal case there isn't simple approximations to drawn on. Inversion sampling is not promising, and neither are there convenient stochastic representations. Cases like those are candidates for acceptance-rejection. Algorithm 2.11 is a simple outcome of that idea. Let $\xi = 1$ and generate proposals through

$$g(x) = \exp(-x), \quad x > 0.$$

[3]It is defined through

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x)\, dx$$

and coincides with the factorials when α is an integer; i.e. $\Gamma(n) = (n-1)!$.

When $\xi = 1$, the ratio $f(x)/g(x)$ attains its maximum at $x = 1$ (differentiate and see) and so

$$M = \frac{\alpha^\alpha}{\Gamma(\alpha)} \exp(-\alpha + 1) \qquad \text{which yields} \qquad \frac{f(x)}{Mg(x)} = \exp\{(\alpha - 1)(\log(x) - x)\},$$

and we are lead to the following algorithm:

Algorithm 2.12 Simple Gamma generator for $\alpha \geq 1$
0 Input: α and ξ
1 Repeat
2 Draw $U^* \sim$ uniform and $X^* \leftarrow -\log(U^*)$ *%Proposal (the exponential)*
3 Draw $U^* \sim$ uniform
4 If $\log(U^*) \leq (\alpha - 1)(\log(X^*) - X^*)$ then *%The test is carried out here*
$X^* \leftarrow \xi X^*$, stop and return X^*

The method works for moderate $\alpha \geq 1$. When $\alpha > 100$ say, the acceptance rates fall below 10%, and the method becomes slow.

More on Gamma sampling
Vast improvements over Algorithm 2.12 are possible through more subtle proposal distributions. The following procedure, justified in Devroye (1986), is effective for all $\alpha \geq 1$:

Algorithm 2.13 Fast Gamma generator for $\alpha \geq 1$
0 Input: ξ, α and $b = \alpha - 1$, $c = 3\alpha - 0.75$.
1 Repeat
2 Sample $U^* \sim$ uniform
3 $W^* \leftarrow U^*(1 - U^*)$, $Y^* \leftarrow \sqrt{c/W^*}(U^* - 0.5)$, $X^* \leftarrow b + Y^*$
6 If $X^* > 0$ then
7 Sample $V^* \sim$ uniform$(0, 1)$
8 $Z^* \leftarrow 64(W^*)^3(V^*)^2$
9 If $Z^* \leq 1 - 2(Y^*)^2/X^*$ **or** if $\log(Z^*) \leq 2\{b(\log(X^*/b) - Y^*)\}$
then **stop** and return $X^* \leftarrow \xi X^*/\alpha$.

The loop is repeated until the stop criterion is satisfied.

The case $\alpha < 1$ is referred back to $1 + \alpha$ through **Stuart's** theorem; i.e.

$$X = YU^{1/\alpha} \sim \text{Gamma}(\alpha) \qquad \text{if} \qquad Y \sim \text{Gamma}(1 + \alpha),\ U \sim \text{Uniform};$$

see Devroye (1986). Commands in the computer are summarized as follows:

Algorithm 2.14 Gamma generator for $\alpha < 1$
0 Input: ξ, α
1 Sample $Z^* \sim$ Gamma$(1 + \alpha)$ *%From Algorithm 2.11 or 2.12*
2 Sample $U^* \sim$ uniform
3 Return $Z^* \leftarrow \xi Z^*(U^*)^{1/\alpha}$

Together Algorithms 2.12 and 2.13 offer rapid sampling of general Gamma variables.

1.6 Mathematical arguments

Section 2.2

The limit relationship (1.9) Only the *upper* percentiles will be considered; the lower ones are similar. Suppose

$$\frac{q_{1\epsilon}}{q_{2\epsilon}} \to 0, \quad \text{as} \quad \epsilon \to 0,$$

which is the condition (1.8) in Section 2.2. Note that both numerator and denominator tend to zero. Hence, l'Hôpital's rule yields

$$\frac{\frac{\partial q_{1\epsilon}}{\partial \epsilon}}{\frac{\partial q_{2\epsilon}}{\partial \epsilon}} \to 0, \quad \text{as} \quad \epsilon \to 0.$$

Differentiate both sides of $F_i(q_{i\epsilon}) = 1 - \epsilon$ with respect to ϵ, $i = 1, 2$. By the chain rule

$$f_1(q_{1\epsilon})\frac{\partial q_{1\epsilon}}{\partial \epsilon} = -1, \quad \text{and} \quad f_2(q_{2\epsilon})\frac{\partial q_{2\epsilon}}{\partial \epsilon} = -1, \tag{1.40}$$

so that

$$\frac{f_2(q_{2\epsilon})}{f_1(q_{1\epsilon})} = \frac{\frac{\partial q_{1\epsilon}}{\partial \epsilon}}{\frac{\partial q_{2\epsilon}}{\partial \epsilon}} \to 0, \quad \text{as} \quad \epsilon \to 0 \qquad \text{and} \qquad \frac{a_{1\varepsilon}}{a_{2\varepsilon}} = \frac{f(q_{2\epsilon})}{f(q_{1\epsilon})},$$

as claimed in (1.9).

The limit relationships (1.7) Again only the *upper* percentile is treated. Note that a_ϵ in (1.6) right can be rewritten

$$a_\epsilon = \sqrt{\frac{1-\epsilon}{b_\epsilon}} \qquad \text{where} \qquad \delta_\epsilon = \frac{f(q_\epsilon)^2}{\epsilon}.$$

and we must examine b_ϵ. If the density function $f(x)$ has a derivative $f'(x)$, l'Hôpital's rule may be used. The limit of b_ϵ is then that of

$$2f(q_\epsilon)f'(q_\epsilon)\frac{\partial q_\epsilon}{\partial \epsilon} = -2f'(q_\epsilon)$$

similar to (1.40). Since $q_\epsilon \to \infty$ as $\epsilon \to 0$ it follows that $b_\epsilon \to 0$ and hence $a_\epsilon \to \infty$ if

$$f'(x) \to 0 \quad \text{as} \quad x \to \infty.$$

It is possible to construct pathological cases when this does not hold, but in practice the condition is valid.

Section 2.5

Algorithm 2.10 Let $X_1, \ldots, X_n$ be stochastically independent with common density function

$$f(x) = \exp(-x), \quad x > 0.$$

If we define

$$p_n(\lambda) = \Pr(X_1 + \ldots + X_n < \lambda \leq X_1 + \ldots + X_{n+1}),$$

the assertion (1.35) behind the Poisson generator is that

$$p_n(\lambda) = \frac{\lambda^n}{n!} \exp(-\lambda).$$

The proof is exercise in conditional probabiltities. For $n > 1$ we may write the probability as the integral

$$p_n(\lambda) = \int_0^\infty \Pr(x + X_2 \ldots + X_n < \lambda \leq x + X_2 + \ldots + X_{n+1} | X_1 = x) f(x)\, dx,$$

or

$$p_n(\lambda) = \int_0^\infty \Pr(X_2 \ldots + X_n < \lambda - x \leq X_2 + \ldots + X_{n+1}) f(x)\, dx.$$

This can be written

$$p_n(\lambda) = \int_0^\lambda p_{n-1}(\lambda - x) f(x)\, dx, \qquad n = 1, 2, \ldots.$$

which starts at

$$p_0(\lambda) = \Pr(X_1 > \lambda) = \exp(-\lambda).$$

The result is certainly true for $n = 0$, and if it is true for $n - 1$, then

$$p_n(\lambda) = \int_0^\lambda \frac{(\lambda - x)^{n-1}}{(n-1)!} e^{-(\lambda - x)} e^{-x}\, dx = \int_0^\lambda \frac{(\lambda - x)^{n-1}}{(n-1)!}\, dx\, e^{-\lambda} = \frac{\lambda^n}{n!} \exp^{-\lambda},$$

and it holds for n as well.

1.7 Further reading

1.8 Exercises

Introduction
These exercises are meant to promote Monte Carlo technique and are preliminary to problem solving in the next chapter. Some topics of more general importance are also introduced here. Q-Q plotting (Exercises 2.2.2-2.2.5) is a convenient way of comparing distributions and are used on many occasions later. For some of the exercises the underlying answer is known permitting us to examine how well Monte Carlo works. If you find problems overly simplistic, remember that they are only an aid to tackle realistic situations later where the answer is *not* known. Quite a lot about Monte Carlo performance can be learned from simple examples.

Section 2.2
Exercise 2.2.1 Consider *Gaussian* financial returns R for which $\xi = 0.5\%$ and $\sigma = 5\%$. They might well be monthly ones. **a)** Run Monte Carlo experiments with $m = 100$, $m = 1000$ and $m = 10000$ simulations and in each case compute means $\bar{X}^*$ and and standard deviation s^*. **b)** Judge the *relative* accuracy in per cent; i.e

$$e_r^* = (\frac{\bar{X}^*}{\xi} - 1) \times 100 \qquad \text{or} \qquad e_r^* = (\frac{s^*}{\sigma} - 1) \times 100.$$

c) How good are the chances of determining ξ and σ if we are dealing with *historical* data instead of simulated ones?

Exercise 2.2.2 a) Generate $m = 1000$ Monte Carlo returns $R_1^*, \ldots, R_m^*$ assuming them to be normal with $\xi = 0.5\%$ and $\sigma = 5\%$. **b)** Order them in *ascending* order as

$$R_{(1)}^* \leq \ldots \leq R_{(m)}^*$$

and for $i = 1, 2 \ldots, m$

$$\text{plot } R_{(i)}^* \text{ against } \Phi^{-1}(u_i) \qquad \text{where} \qquad u_i = \frac{i - 1/2}{m}.$$

Here $\Phi^{-1}(u)$ is the inverse normal integral. **c)** Repeat when $R_1^*, \ldots, R_m^*$ are generated under $\xi = 0$ and $\sigma = 1$ (which could come from property insurance). **d)** You understand why the plot in c) is a straight line at angle 45°. Why is it another straight line in b)?

Exercise 2.2.3 The procedure in Exercise 2.2.2 where ordered simulations (or historical data!) were plotted against percentiles are known as a **Q-Q** plots. Arguably it is the most efficient way of checking graphically whether a given distribution fits. If it doesn't, the shape deviates from a straight line. **a)** Draw a Monte Carlo sample $Z_1^*, \ldots, Z_m^*$ from the Pareto distribution with $\alpha = 5$ and $\beta = 1$ using Algorithm 2.8. Take $m = 1000$. **b)** Order as

$$Z_{(1)}^* \leq \ldots \leq Z_{(m)}^*$$

and plot $Z_{(i)}^*$ against $\Phi^{-1}(u_i)$ as in Exercise 2.2.2. **c)** Comment on how the tails of the Pareto distribution show up in the discrepancies from the straight line. There is a *general* story here.

Exercise 2.2.4 Q-Q plotting may be carried out against any distribution. The Gaussian percentiles $\Phi^{-i}(u_i)$ are then replaced by general ones

$$F^{-1}(u_i) \qquad \text{where} \qquad u_i = \frac{i - 1/2}{m}$$

and ordered simulations like $R_{(i)}^*$ or $Z_{(i)}^*$ plotted against $F^{-1}(u_i)$. **a)** Compute the percentiles of the Pareto distribution when $\alpha = 5$ and $\beta = 1$ using (1.32). Take $m = 1000$ and store them. **b)** Draw $m = 1000$ simulations from the *same* Pareto distribution and Q-Q plot against the percentiles in **a)**. **c)** Repeat b) with Pareto simulations from $\alpha = 5$ and $\beta = 0.5$. Comment? **d)** Repeat **b)** one more time, but now with $\alpha = 3$ and $\beta = 1$. What has happened to the plot? **e)** Simulate $m = 1000$ normal variables with $\xi = 0.5\%$ and $\sigma = 5\%$ and Q-Q plot against the Pareto percentiles in a) as before. Anything different compared to Exercise 2.2.3b)?

Exercise 2.2.5 Q-Q plots with *fake* shapes emerge when the number of simulations is small. With the Monte Carlo experiments themselves that is not important (since m is large), but it is a highly relevant point with historical data. **a)** Generate normal Monte Carlo samples ($\xi = 0.5\%$ and $\sigma = 5\%$) for $m = 20$ and Q-Q plot against the mother distribution. Do this five times. Comments? **b)** Repeat the exercise for the Pareto distribution when $\alpha = 5$ and $\beta = 1$, but now use $m = 100$. **c)** Try to formulate some general lessons of the exercise.

Exercise 2.2.6 The accuracy of Monte Carlo evaluations of standard deviations hinges on the kurtosis of X; see (1.4). Kurtosis is defined as

$$\kappa = \frac{E(X - \xi)^4}{\sigma^4} - 3$$

where $\xi = E(X)$ and $\sigma = \text{sd}(X)$. Its meaning will be illustrated by the stochastic volatility model (1.20); i.e. $X = \xi + \sigma_0 \sqrt{Z}\varepsilon$ where ε is $N(0, 1)$. **a)** Show that

$$(X - \xi)^2 = \sigma_0^2 Z \varepsilon^2 \qquad \text{so that} \qquad \sigma^2 = E(X - \xi)^2 = \sigma_0^2 E(Z).$$

b) By utilising (see Appendix A) that $E(\varepsilon^4) = 3$ also show that

$$(X-\xi)^4 = \sigma_0^4 Z^2 \varepsilon^4 \qquad \text{which yields} \qquad E(X-\xi)^4 = 3\sigma_0^4 E(Z^2).$$

c) Now deduce that

$$\kappa = 3\left(\frac{\mathrm{sd}(Z)}{E(Z)}\right)^2 \qquad \text{so that} \qquad \kappa = 0 \quad \text{when } X \text{ is normal.}$$

d) Explain why $\kappa \doteq 3\mathrm{var}(Z)$ if $E(Z) \doteq 1$. For most stochastic volatility models used in practice this is *approximately* true.

Exercise 2.2.7 Use (1.4) to explain how the accuracy of a standard deviation estimate depends on kurtosis. Explicitly, compare the cases $\kappa = 6$ and $\kappa = 0$ ($\kappa = 6$ could well be a reasonable value for *daily* equity returns).

Exercise 2.2.8 The standard kurtosis estimate is

$$\kappa^* = \frac{\lambda_4^*}{s^{*4}} - 3 \qquad \text{where} \qquad \lambda_4^* = \frac{1}{m}\sum_{i=1}^{m}(X_i^* - \bar{X}^*)^4$$

Here λ_4^* is the *fourth order moment*. **a)** Motivate this estimate. We shall test it on log-normal data $X = \exp(\xi + \sigma\varepsilon)$ where ε is $N(0,1)$. **b)** The parameter ξ does not matter. Do you see why? **c**) Simulate log-normal data when $\sigma = 0.05$. Use $m = 100$, $m = 1000$ and $m = 10000$ and estimate each time the kurtosis. **d)** Repeat c) when $\sigma = 1$. **e)** Compare the results with the the theoretical expression which for the kurtosis of the log-normal which is

$$\kappa = \frac{e^{6\sigma^2} - 4e^{3\sigma^2} + 6e^{\sigma^2} - 3}{(e^{\sigma^2}-1)^2}.$$

The small σ may correspond to monthly assets returns in finance and the large ones to the size of claims in property insurance. When is the kurtosis easiest to estimate?

Exercise 2.2.9 For this exercise use a procedure for density estimation in a software package or implement (1.10) on your own. There is smoothing parameter h to adjust and we shall examine how it affects the performance of the estimate. **a)** Draw a log-normal sample based on $\xi = 0.5\%$ and $\sigma = 5\%$ using $m = 100$. **b)** Apply the estimate with $h = 0.1, 0.2$ and 0.3. Comment! **c)** Repeat the exercise with $m = 1000$. **d)** Repeat b) and c) when $\xi = 0$ and $\sigma = 1$. What seems to be the conclusions from this exercise?

Exercise 2.2.10 Use the results in Section 2.2.2 to detail the confidence interval (1.12) when ψ is the mean, the standard deviation and the percentile.

Exercise 2.2.11 Usually the Monte Carlo standard deviation is approximately of the form $\zeta/\sqrt{m}$ which equals σ_0 if $m = (\zeta/\sigma_0)^2$; see (1.13). Of course, ζ is not known, but we can get around that through a preliminary, smaller experiment. That makes the entire scheme

$$\underbrace{X_1^*, \ldots, X_{m_1}^*}_{\textit{First round}} \longrightarrow \zeta^*, \qquad m = (\zeta^*/\sigma_0)^2 \qquad \text{and then} \qquad \underbrace{X_{m_1+1}^*, \ldots, X_m^*}_{\textit{Second round}}.$$

After ζ has been estimated from the first round, the main, *second* experiment is run with the number of simulations determined. **a)** If we are dealing with the mean, then $m = (s^*/\sigma_0)^2$ where s^* is the sample standard deviation of the first m_1 simulations. Explain why. **b)** If X is $N(0,1)$ and $m_1 = 100$, run the preliminary experiment five times, estimate each time s^* and report how much the estimated m varies. **c)** Repeat b) when is X is Pareto distributed with parameters $\alpha = 2$ and $\beta = 1$. **c)** What you simulate in practice is quite likely to follow a distribution between these two extremes. Did $m_1 = 100$ seem enough with the Pareto model?

Exercise 2.2.12 Suppose the Monte Carlo experiment is run to estimate the ϵ-percentile. Show that we in the set-up of the preceding exercise should use

$$m = \frac{\epsilon(1-\epsilon)}{\{f^*(q_\epsilon^*)\}^2 \sigma_0^2}$$

for the second part of the experiment. Here q_ϵ^* is the preliminary estimate of the percentile and $f^*(q_\epsilon^*)\}^2$ the density estimate.

Section 2.3

Exercise 2.3.1 We shall in this exercise compare normal and log-normal models for financial returns through simulations. The alternatives are

$$\underset{\textit{normal model}}{R = \xi + \sigma\varepsilon} \qquad \text{and} \qquad \underset{\textit{log-normal model}}{\tilde{R} = (1+\xi)\exp(-\tfrac{1}{2}\sigma^2 + \sigma\varepsilon) - 1}$$

where $\varepsilon \sim N(0,1)$. **a)** Explain why $E(R) = E(\tilde{R})$. **b)** Suppose $\xi = 0.02\%$ and $\sigma = 1.5\%$ (which could be true for *daily* equity returns) Draw $m = 10000$ simulations from each distribution, sort *each* sequence separately in *ascending* order as

$$\underset{\textit{normal model}}{R^*_{(1)} \le \ldots \le R^*_{(m)}} \qquad \text{and} \qquad \underset{\textit{log-normal model}}{\tilde{R}^*_{(1)} \le \ldots \le \tilde{R}^*_{(m)}}$$

and plot corresponding pairs $(R^*_{(i)}, \tilde{R}^*_{(i)})$ from the two sequences against each other. **c)** Repeat b) for $\xi = 5\%$ and $\sigma = 23.7\%$ (perhaps *annual* equity return). **d)** Draw conclusions from these two rounds of experiments.

Exercise 2.3.2 The issue resembles the one in Exercise 2.3.1, although now

$$R = \xi + \sigma\varepsilon \qquad \text{and} \qquad \tilde{R} = \exp(\tilde{\xi} + \tilde{\sigma}\varepsilon) - 1$$

where the parameters (ξ, σ) and $(\tilde{\xi}, \tilde{\sigma})$ differ. As usual $\varepsilon \sim N(0,1)$. **a)** Show that if

$$\tilde{\sigma} = \sqrt{1 + (\sigma/\xi)^2} \qquad \text{and} \qquad \tilde{\xi} = \log(\xi) - \frac{1}{2}\tilde{\sigma}^2$$

then $E(R) = E(\tilde{R})$ and $\mathrm{sd}(R) = \mathrm{sd}(\tilde{R})$. **b**) Determine $\tilde{\xi}$ and $\tilde{\sigma}$ if $\xi = 5\%$ and $\sigma = 23.7\%$. **c)** Repeat the experiment in Exercise 2.3.1c with these parameters; i.e. generate ordered, simulated returns $R^*_{(i)}$ and $\tilde{R}^*_{(i)}$ under the two models and plot the pairs $(R^*_{(i)}, \tilde{R}^*_{(i)})$ for $i = 1, \ldots, m$ when $m = 10000$. **d)** Comment on the difference between the two models.

Exercise 2.3.3 a) Draw a sample of 1000 log-normals $Z = \exp(\sigma\varepsilon)$ when $\sigma = 0.05$, $\sigma = 0.4$, $\sigma = 1.0$ and $\sigma = 2$. **b)** Estimate in each of the four cases the density function and plot it. **c)** Comment on the distribution as a model for financial returns and for size of claims in property insurance.

Exercise 2.3.4 Consider the stochastic volatility model (1.20) for log-returns; i.e. assume that

$$R = \exp(X) - 1, \qquad \text{where} \qquad X = \xi + \sigma_0\sqrt{Z}\,\varepsilon, \quad \varepsilon \sim N(0,1).$$

A possible model for Z is to make it log-normal, for example $Z = \exp(-\tau^2 + 2\tau\eta)$ where $\eta \sim N(0,1)$, $\tau \ge 0$ and where η is independent of ε. **a)** Explain why $\sqrt{Z}$ is also a log-normal variable. **b)** Use the formulae for mean and standard deviation of such variables in Section 2.3 to deduce that

$$E(\sqrt{Z}) = 1 \qquad \text{and} \qquad \mathrm{sd}(\sqrt{Z}) = \sqrt{e^{\tau^2} - 1},$$

and the degree of stochastic volatility goes up with τ.

Exercise 2.3.5 a) Implement a program for sampling R under the model of the preceding exercise. Suppose $\xi = 0.5\%$ and $\sigma_0 = 5\%$ (R could then be monthly return of equity). **b)** Draw $m = 1000$ simulations of R when $\tau = 0.5$, estimate the density function and plot it (it is inaccessible through ordinary mathematics now!). **c)** Redo **b)** when $\tau = 0.001$ and comment on the different shapes of the plots.

Exercise 2.3.6 Consider again the model for R introduced in Exercise 2.3.4 and the simulation program in Exercise 2.3.5. Suppose $\xi = 0.5\%$ and $\sigma_0 = 5\%$. **a)** Run the program $m = 10000$ times when $\tau = 0.5$ and compute the ε-percentiles of R for $\varepsilon = 0.01$, 0.05, 0.50, 0.95 and 0.99. **b)** Redo when $\tau = 0.001$. **c)** Compare the results in a) and b) and comment.

Section 2.4
Exercise 2.4.1 Consider the bivariate normal model (1.21). **a)** Simulate it ($m = 100$) when

$$\xi_1 = \xi_2 = 5\%, \qquad \sigma_1 = \sigma_2 = 25\% \qquad \text{and} \qquad \rho = 0.2,\ \rho = 0.7\ \rho = 0.95,$$

and make scatter-plots in each of these three cases. **b)** Redo a) for log-returns; i.e convert X_1 and X_2 to R_1 and R_2 through $R_1 = \exp(X_1) - 1$ and $R_2 = \exp(X_2) - 1$. This example could be annual returns for equity.

Exercise 2.4.2 Suppose a financial portfolio has placed equal weights on the two assets of the preceding exercise. This means that portfolio return is $\mathcal{R} = (R_1 + R_2)/2$; see (**??**) in Section 1.3. **a)** Simulate $\mathcal{R}$ $m = 10000$ times when $\rho = 0.2$ and compute the percentiles for $\varepsilon = 1$, 5%, 50% and 95%. **b)** Redo a) for $\rho = 0.5$ and $\rho = 0.95$ and compare the sets of percentiles computed.

Exercise 2.4.3 Suppose the financial portfolio of the preceding exercise is based on $J = 5$ assets instead still with equal weights on all. The portfolio return is now $\mathcal{R} = (R_1 + \ldots + R_5)/5$. **a)** Implement Algorithm 2.4 for financial returns that are log-normal with common correlation coefficient ρ. **b)** Determine the percentiles of $\mathcal{R}$ when $\xi = 5\%$ and $\sigma = 25\%$ for all five assets and $\rho = 0.2$. **c)** Redo b) when $\rho = 0.5$ and 0.95. **d)** Compare the evaluations in b) and c) with the analogous ones in Exercise 2.4.2. Any patterns?

Exercise 2.4.4 Consider a heavy-tailed bivariate model of the form

$$\begin{array}{l} R_1 = \exp(X_1) - 1 \\ R_2 = \exp(X_2) - 1 \end{array} \quad \text{where} \quad \begin{array}{l} X_1 = \xi + \sigma_0\sqrt{Z_1}\,\varepsilon_1 \\ X_2 = \xi + \sigma_0\sqrt{Z_2}\,\varepsilon_2. \end{array} \quad \text{and} \quad Z_1 = Z_2 = Z.$$

Here ε_1 and ε_2 are $N(0,1)$ with correlation ρ. As in Exercise 2.3.4 $Z = \exp(-\tau^2 + 2\tau\eta)$ for $\eta \sim N(0,1)$. **a)** Implement a program that samples (R_1, R_2). **b)** Calculate the 1%, 5%, 50% and 95% percentiles of the portfolio return $\mathcal{R} = (R_1 + R_2)/2$ under conditions similar to those in Exercise 2.4.2; i.e take $\xi = 5\%$, $\sigma_0 = 25\%$, $\rho = 0.5$ and let $\tau = 0.5$. **c)** What's the effect of the heavy tails when you compare with the $\rho = 0.5$ evaluations in Exercise 2.4.2?

Exercise 2.4.5 Consider the model of the preceding exercise, but now allow Z_1 and Z_2 to be different. A simple construction is

$$Z_1 = \exp(-\tau_1^2 + 2\tau_1\eta_1) \qquad \text{and} \qquad Z_2 = \exp(-\tau_2^2 + 2\tau_2\eta_2)$$

where η_1 and η_2 are $N(0,1)$ with correlation $\rho_\eta = \text{cor}(\eta_1, \eta_2)$. **a)** Explain why the model is the same as in the preceding exercise if $\tau_1 = \tau_2$ and $\rho_\eta = 1$. **b)** Revise the program in Exercise 2.4.4a) so that it covers the present situation. **c)** Calculate the 1%, 5%, 50% and 95% percentiles of the portfolio return $\mathcal{R} = (R_1 + R_2)/2$ when $\xi = 5\%$, $\sigma_0 = 25\%$, $\rho = 0.5$, $\tau_1 = \tau_2 = 0.5$ and $\rho_\eta = 0.0$. Compare with the results from Exercise 2.4.4.

Exercise 2.4.6 An avant-garde model would be to allow stochastic *correlations*. If it appears far-fetched, the idea has nevertheless been proposed (and substantiated) in academic literature, for example in Ball and

Torus (2000). With the machinery in Section 2.4 it is not hard to build such models for financial returns. For example, starting from the same angle as before let $R_j = \exp(\xi + \sigma_0\varepsilon_j) - 1$ for $j = 1, 2$ where ε_1 and ε_2 are N(0,1) with correlation coefficient ρ for which

$$\rho = \frac{(1+\rho_0)e^{\tau\eta} - (1-\rho_0)}{(1+\rho_0)e^{\tau\eta} + (1-\rho_0)} \qquad \text{where} \qquad \eta \sim N(0,1).$$

a) Verify that $-1 < \rho < 1$ and that ρ_0 is the median in the distribution for ρ [Hint: The median appears when $\eta = 0$.]. **b)** How do you make ρ a fixed parameter and what's its value then? **c)** Implement a program that samples (R_1, R_2) under this model. **d)** Compute the 1%, 5%, 50% and 95% percentiles of the portfolio return $\mathcal{R} = (R_1 + R_2)/2$ now using $\xi = 5\%$, $\sigma_0 = 25\%$, $\rho_0 = 0.5$ and $\tau_1 = 0.5$. You may again compare with results in Exercise 2.4.2

Section 2.5
Exercises 2.5.1-4 introduce probability distributions that have been proposed (and used) in property insurance. None of them admits simple matematical expressions for mean and variance. An alternative way of interpreting their parameters is to use **median** and **quantile difference** i.e.

$$\text{med}(X) = q_{0.5} \qquad \text{and} \qquad \text{qd}(X) = q_{0.75} - q_{0.25} \tag{1.41}$$

where q_ε is the lower ε-percentile of the distribution function $F(x)$; i.e the solution of the equation $F(q_\varepsilon) = \varepsilon$. The quantile difference is a measure of spread.

Exercise 2.5.1 The **Weibull** model comes from engineering orginally. Its distribution function is

$$F(x) = 1 - \exp\{-(x/\beta)^\alpha\}, \quad x > 0.$$

Here $\alpha, \beta > 0$ are parameters. **a)** Show that

$$X^* = \beta(-\log U^*)^{1/\alpha}$$

is the inversion sampler. **b)** Use this to derive mathematical expressions for med(X) and qd(X); see (1.41). **c)** Generate $m = 1000$ simulations for $\beta = 1$ and $\alpha = 1.0$, 3.15 and 5.0. Plot in each case density estimates and comment. **d)** Run $m = 10000$ simulations for $\alpha = 3.15$ and $\beta = 1$ and run a Q-Q plot against the normal distribution. Any comments?

Exercise 2.5.2 The **Fréchet** distribution

$$F(x) = \exp\{-(x/\beta)^{-\alpha}\}, \quad x > 0,$$

is of a so-called extreme value type. Again $\alpha, \beta > 0$ are parameters. **a)** Derive its inversion sampler and **b)** Determine med(X) and qd(X); see (1.41).

Exercise 2.5.3 Still another distribution sometimes used in property insurance is the **logistic** one for which

$$F(x) = 1 - \frac{1+\alpha}{1+\alpha\exp(x/\beta)}, \quad x > 0.$$

Once again the parameters $\alpha, \beta > 0$. **a)** Derive the inversion sampler. **b)** Determine mathematical expressions for med(X) and qd(X); see (1.41).

Exercise 2.5.4 The **Burr** model has three positive parameters α_1, α_2 and β and its distribution function is

$$F(x) = 1 - \{1 + (x/\beta)^{\alpha_1}\}^{-\alpha_2}, \quad x > 0.$$

a) Derive its inversion sampler. **b)** Find mathematical expressions for med(X) and qd(X), see (1.41).

Section 2.6
Exercise 2.6.1 Let Y be exponentially distributed with density function $\exp(-y)$, $y > 0$ and let $X = \beta Y^{1/\alpha}$ with $\alpha, \beta > 0$. **a)** Show that

$$\Pr(X \le x) = \Pr(Y \le (x/\beta)^\alpha) = 1 - \exp\{-(x/\beta)^\alpha\}, \qquad x > 0.$$

b) Use Exercise 2.5.1 to identify the model for X as the Weibull distribution.

Exercise 2.6.2 a) Draw $m = 1000$ Poisson variables when $\lambda = 5$, 20 and 100. **b)** In each of the three cases use a Q-Q plot to compare against the normal distribution. Comments?

Exercise 2.6.3 Let $N_1 = M_4 + M_7$ where M_4 and M_7 are Poisson distributed with parameters $\lambda = 4$ and $\lambda = 7$ respectively and let N_2 be Poisson with parameter $\lambda = 11$. **a)** Generate $m = 1000$ Monte Carlo samples of N_1 and then **b)** the same number of simulations from N_2. **c)** Compare the the distributions of N_1 and N_2 by Q-Q plotting their *ordered* simulations against each other. Any comments? For the general story see Chapter 8.

Exercise 2.6.4 We shall in this exercise consider sums of exponentially distributed variables, as in Algorithm 2.10, but now with a *fixed* number of terms. Let $Y = X_1 + \ldots + X_5$, where $X_1, \ldots, X_5$ are exponentially distributed. **a)** Sample Y one thousand times. **b)** Sample the same number of times from a Gamma distribution with shape parameter $\alpha = 5$. **c)** Compare the two distributions by plotting the ordered simulations against each other as in the preceding exercise. Again there is a more general story. It is presented in Chapter 9.

Exercise 2.6.5 One way to inestigate the efficiency of the Gamma simulator in Algorithm 2.11 is to check how often the acceptance criterion holds. With a slight rephrasal let U^* and V^* be uniform random variables. What we seek is the probability of the event

$$\log(U^*) \le (\alpha - 1)(\log(X^*) - X^*) \qquad \text{where} \qquad X^* = -\log(V^*).$$

Run 100000 simulations for $\alpha = 2$, 20, 100 and 1000 and estimate the acceptance probability. A smarter way is given in the next exercise!

Exercise 2.6.6 a) Implement the Gamma generator Algorithm 2.11. **b)** Generate $m = 1000$ simulations when $\alpha = 2$ and $\xi = 1$. **c)** Check the program by plotting a density function estimated from the simulations. **d)** Redo (possibly with smaller m) for $\alpha = 100$ and establish that the procedure now is more time-consuming. To understand why we shall try to find out how many repetitions are needed for accept to occur. The simplest way is to compute the constant M prior to Algorithm 2.11 in Section 2.6; i.e.

$$M = \frac{\alpha^\alpha}{\Gamma(\alpha)} \exp(-\alpha + 1) \qquad \text{where for integers } n \qquad \Gamma(n) = (n-1)!$$

e) Explain from the theory in Section 2.5 why M equals the average number of trials for each simulation. **f)** Compute it for $\alpha = 2$, 20, 100 and 1000 and compare with the assessments in Exercise 2.6.5. Any comments? Such sensitive performance is typical for the rejection/acceptance method. Cleverness is needed!