1 Monte Carlo thinking and technique

1.1 Introduction

To get started we need stochastic modelling and Monte Carlo, and this chapter is an introduction to both. The modelling part, though skinny is enough to see us through a lot of problems in the next chapter. There is a deliberate thought behind the manner in which models are presented. Emphasis is on how they are simulated in the computer, not on their probabilistic description. This is the constructive approach where mathematics is developed 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. Typicaly a risk variable X has many random sources, and it is usually hard, or even impossible, to find its density function $f(x)$ or its distribution function $F(x)$ through mathematical deductions. This is true even when the random mechanisms involved are simple to write down and fully known. It is here Monte Carlo comes in. Computer simulations $X_1^*, \ldots, X_m^*$ enables the distribution of X to be *approximated*. How that is done and the error it brings is best discussed at a general level. That is where we start (Section 2.2). Then come construction and design, an immense theme. Basis for stochastic simulation is the **uniform** random variable U for which every value between 0 and 1 is equally likely and the density function a horizontal straight line over the interval $(0, 1)$. 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. Computer software contains procedures for drawing 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 why be so basic? Gaussian and many other distributions can be sampled through software packages. Can't we ignore the theory and proceed directly to how they are used? A lot of work *can* be satsfactory carried out with no knowledge of underlying algorithms, yet they should be studied. Otherwise we would be at the mercy of what software vendors have chosen to implement. Consider large claims in property insurance. One of the popular models is the Pareto distribution (you see why in Chapter 9), but a Pareto generator is not always routinely available, and we should be able to set up one on our own. Then there is computational speed. Software packages have a tendency to run slowly. By writing a program in, say the C language, speed may be enhanced by a huge factor and even very much more if **quasi-randomness** (Section 4.7) is invoked. 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
Quantities sought are typically expectation, standard deviation, percentiles and probability density

function. This section demonstrates how they are worked out from simulations $X_1^*, \ldots, X_m^*$ of X, the error that brings 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 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

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

Monte Carlo estimates of ξ are unbiased, and their error may in theory 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, but if $\bar{X}^*$ is to be a price of something, high Monte Carlo accuracy may still be demanded.

For s^* the statistical properties are 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 $\kappa = \mathrm{kurt}(X)$ is the **kurtosis** of X (see Appendix A.1 for the definition). Are these results unknown, look them up (p. 355) in Kendall, Stuart and Ord (1994). They are needed in Section 5.7 too. For normal variables $\kappa = 0$. The approximations (1.4) are asymptotic and become exact as $m \to \infty$. Large sample results of this kind work excellently with Monte Carlo where m is large.

Example: Financial returns

Let us examine how this machinery performs 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 the experiment tells us something else. *The sample mean is in terms of relative error less accurately estimated than the standard deviation.* Suppose the simulations had been historical returns of equity. After 1000 months (about eighty years, a very long time) the relative error of 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 Section 5.3 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.
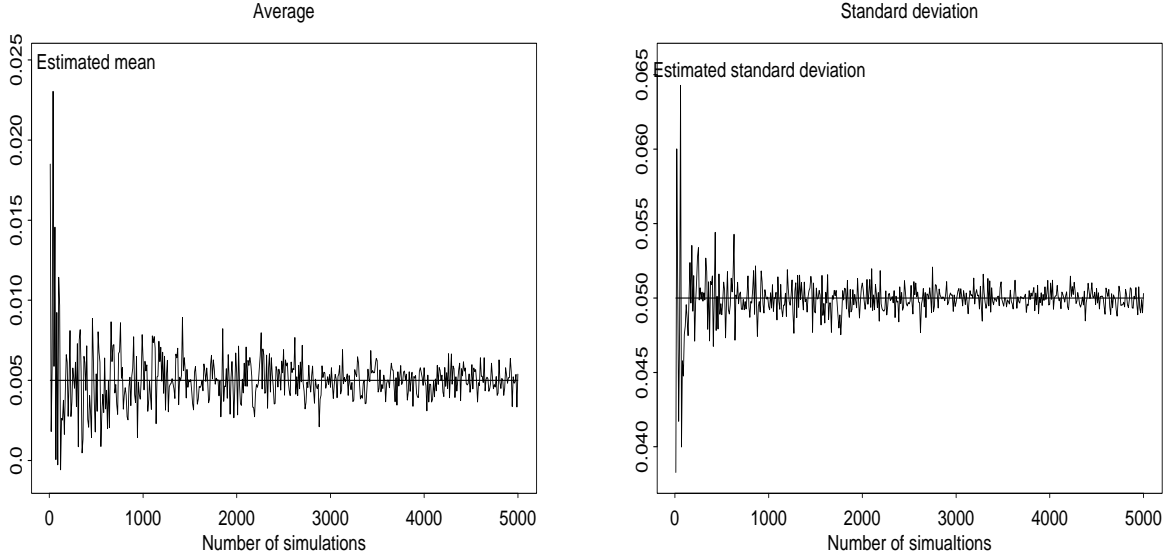
Figure 2.1 *Sample mean and standard deviation against the number of simulations for a Gaussian model. Straight lines are the true parameters.*

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

$$\frac{\mathrm{sd}(\bar{X}^*)}{\xi} = \frac{\sigma}{\xi}\frac{1}{\sqrt{m}} = \frac{10}{\sqrt{m}} \qquad \text{and} \qquad \frac{\mathrm{sd}(s^*)}{\sigma} \doteq \sqrt{1/2 + \kappa/4}\ \frac{1}{\sqrt{m}} \doteq \frac{0.71}{\sqrt{m}}$$

when the values of the parameters are inserted ($\kappa = 0$). The coefficients explain why $\bar{X}^*$ is so much more inaccurate. In Section 13.5 parameter errors of the celebrated Wilkie asset model follow a similar pattern.

Percentiles

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

$$\underset{upper}{F(q_\epsilon) = 1 - \epsilon} \qquad \text{or} \qquad \underset{lower}{F(q_\epsilon) = \epsilon,}$$

depending on whether the upper or the lower version is sought. With insurance risk it is typically the former, in finance the latter. The Monte Carlo approximation is obtained by sorting the simulations, for example in descending order as $X^*_{(1)} \geq \ldots \geq X^*_{(m)}$. Then
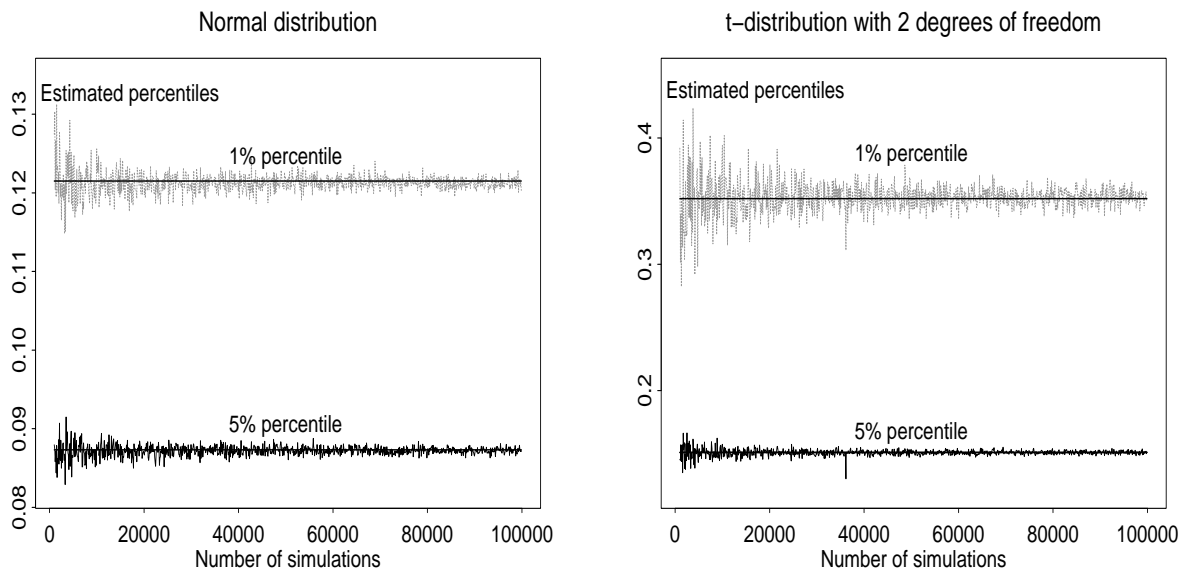
$$\underset{upper}{q^*_\epsilon = X^*_{(\epsilon m)}} \qquad \text{or} \qquad \underset{lower}{q^*_\epsilon = X^*_{((1-\epsilon)m)}} \tag{1.5}$$

with error approximately

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

which are again asymptotic results as $m \to \infty$; see Kendall, Stuart and Ord (1994), p. 382. It is possible to evaluate $f(q_\epsilon)$ through density estimation (see below) and insert the estimate into (1.6)

3

Figure 2.2 *Estimated percentiles of simulated series against the number of simulations.* **Note**: *Scale of the vertical axes unequal.*

for a numerical estimate of $\mathrm{sd}(q^*_\epsilon)$.

The experiment in Figure 2.1 has in Figure 2.2 left been repeated for 1% and 5%-percentiles. Simulation error is larger for the former which is no more than common sense, but it is substantiated by the fact that

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

which is proved in Section 2.7. *Very* many more simulations are required for small ϵ. What about the impact of the distribution itself? The second experiment on the right in Figure 2.2 has been run for the heavy-tailed t-distribution with two degrees of freedom (see Section 2.3 for the definition). Now the error has become much larger than they were for the normal on the left (scales of the vertical axes differ). A precise mathematical result is as follows. Let $q_{1\epsilon}$ and $q_{2\epsilon}$ be percentiles under two different density functions $f_1(x)$ and $f_2(x)$. Suppose the second one has heavier tail than the first. We may take this to mean that

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

and if $a_{i\epsilon} = \sqrt{\epsilon(1-\epsilon)}/f_i(q_{i\epsilon})$ are coefficients similar to a_ϵ in (1.6), then

$$\frac{a_{2\epsilon}}{a_{1\epsilon}} = \frac{f_2(q_{2\epsilon})}{f_1(q_{1\epsilon})} \to \infty \ \text{as} \ \epsilon \to 0; \tag{1.9}$$

see Section 2.7 for the proof. This tells us that simulation error is indeed larger with the second, more heavy-tailed distribution.

Density estimation
Another issue is how the density function $f(x)$ is visualized given simulations $X^*_1, \ldots, X^*_m$. Statistical software is available and works automatically, but it is still useful to have an idea of how
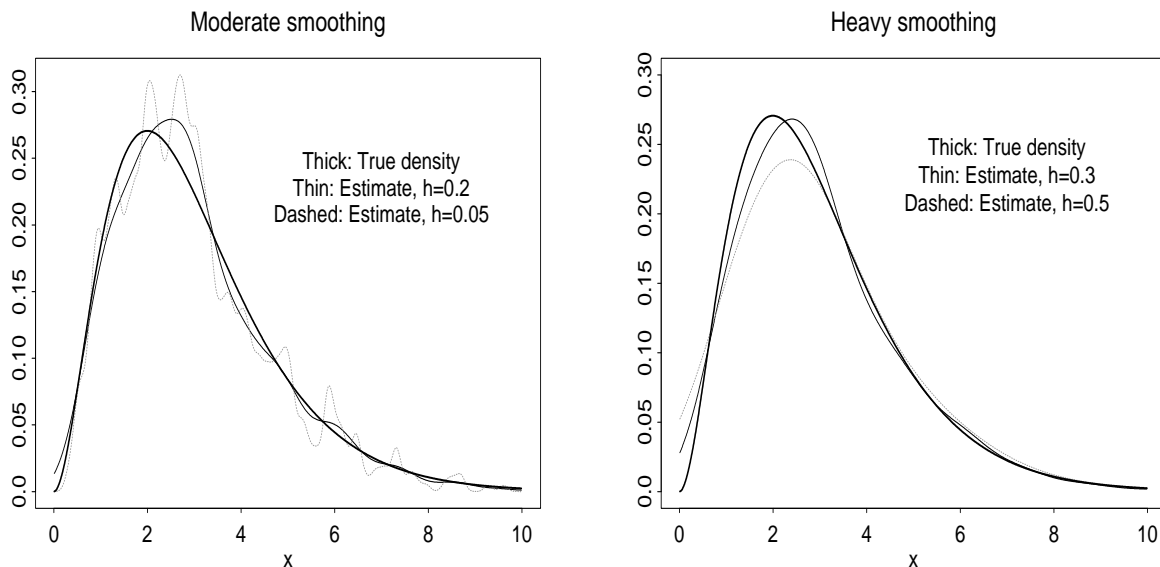
Figure 2.3 *Kernel density estimates based on* 1000 *simulations from model in the text, shown as the thick solid line in both plots.*

such techniques operate, all the more since there is a parameter to adjust. Density functions are in this book estimated from simulations by means of the **Gaussian kernel** method. A smoothing parameter $h > 0$ is then selected, and the estimate is the sum

$$f^*(x) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{hs^*} \varphi\left(\frac{x - X_i^*}{hs^*}\right) \qquad \text{where} \qquad \varphi(x) = \frac{1}{\sqrt{(2\pi)}} e^{-x^2/2}. \tag{1.10}$$

As x is varied $f^*(x)$ traces out a curve which resembles the exact $f(x)$. The method averages m Gaussian density functions with standard deviation hs^* and centered at the m simulations X_i^*. Its statistical properties, derived in Chapter 2 in Wand and Jones (1995), are

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

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*). Commercial software is usually 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 small. This emerges clearly on the left in Figure 2.3 showing estimates based on $m = 1000$ simulations drawn from the density function

$$f(x) = \frac{1}{2}x^2 e^{-x}, \qquad x > 0.$$

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 rule of the thumb is to take h in the range $0.05 - 0.30$, but, as remarked above, it also depend on m.

Other kernels than the Gaussian one can also be used; see Wand and Jones (1995) or Scott (1992) for monographs on density estimation.

Monte Carlo error and selection of m

The discrepancy between a Monte Carlo approximation and its underlying, exact value is nearly always Gaussian as $m \to \infty$. For the sample mean this follows from the central limit theorem, and standard large sample theory from statistics yields the result in most other cases; see Appendix A.4. A Monte Carlo evaluation ψ^* of some quantity ψ is therefore 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 the details are different; see Scott, 1992). Let a^* be an estimate of a obtained from the simulations (how was explained above). The interval

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

then 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 the mean and $a^* = s^*\sqrt{1/2 + \kappa^*/4}$ when ψ is the standard deviation; see (1.3) and (1.4) right (for the kurtosis estimate κ^* consult Exercise 2.2.8).

Such results can also be used for design. Suppose Monte Carlo standard deviation exceeding σ_0 is unaccepable. The equation $a^*/\sqrt{m} = \sigma_0$ when solved for m yields

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

which is the number of simulations required. For the idea to work you need the estimate a^*. Often the only way is to run a preliminary round of simulations, estimate a, 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. The selection of m is further dicussed in Section 7.2.

1.3   Making the Gaussian work

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 built up from scratch below and is the first example of distributions being defined the way they are simulated in the computer. This allows more advanced topics like stochastic volatility, heavy tails and correlated variables to be introduced quickly, though their treatment here is only preliminary. *General*, dependent Gaussian variables require linear algebra and is dealt with in Chapter 5 which introduces time-dependent versions too.

The normal family

Normal (or Gaussian) variables are built up from standard-normal $N(0,1)$ variables (in this book denoted ε or η). Their distribution function is

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

---

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

$\Phi(x) = Q(z) \exp(-x^2/2)$ where
$z = 1/(1 - c_0 x)$ and $Q(z) = z(c_1 + z(c_2 + z(c_3 + z(c_4 + zc_5))))$
$c_0 = 0.2316419$ $\qquad c_1 = 0.127414796$ $\qquad c_2 = -0.142248368$
$c_3 = 0.710706870$ $\quad c_4 = -0.7265760135$ $\quad c_5 = 0.5307027145$

Table 2.1 *Approximation to the normal integral $\Phi(x)$ for $x \leq 0$,*
use $\Phi(x) = 1 - \Phi(-x)$ for $x > 0$

known as the Gaussian integral and needed on many occasions. Closed formluae are unavailable, but an accurate approximation with error less than $1.5 \, 10^{-7}$ (taken from Abramowitz and Stegun, 1965) is given in Table 2.1.

The normal **family** of random variables is defined as

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

where $\xi = E(X)$ and $\sigma = \text{sd}(X)$ are mean and standard deviation. Simulations are generated through $X^* = \xi + \sigma\varepsilon^*$, and the problem is how to draw ε^*. Let $\Phi^{-1}(u)$ be its inverse function (which was denoted q_u earlier). 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 $X^* \leftarrow \xi + \sigma\Phi^{-1}(U^*)$ $\qquad$ %*Or $\Phi^{-1}(U^*)$ replaced by ε^**
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ *generated by software directly*

For this to be practical we must have a quick way of calculating $\Phi^{-1}(u)$. *Very* accurate and simple approximations are available. The method in Table 2.2 was developed by Odeh and Evans (1974) and is for all u accurate to six decimal places. sufficient for most purposes. Even more accurate approximations can be found in Jäckel (2002) who recommends Algorithm 2.1 for Gaussian sampling.

Modelling on logarithmic scale
Models on logarithmic scale are common. Returns on equity (Section 1.3) is a case in point for which the standard model is

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

$\phi^{-1}(\epsilon) = z + Q_1(z)/Q_2(z)$ where $z = \{-2\log(1-\epsilon)\}^{1/2}$ and
$Q_1(z) = c_0 + z(-1 + z(c_2 + zc_3)),\ Q_2(z) = c_4 + z(c_5 + z(c_6 + z(c_7 + zc_8)))$
$c_0 = -0.322232431088$ $\qquad c_1 = -0.342242088547$ $\quad c_2 = -0.020423121024$
$c_3 = 0.0000453642210148$ $\quad c_4 = -0.099348462606$ $\quad c_5 = 0.58858157049$
$c_6 = -0.531103462366$ $\qquad c_7 = 0.0.10353775285$ $\qquad c_8 = -0.0038560700634$

Table 2.2 *Approximation to Gaussian percentiles $\Phi^{-1}(\epsilon)$ for $\epsilon \geq 1/2$. Use*
$\Phi^{-1}(\epsilon) = -\Phi^{-1}(1 - \epsilon)$ for $\epsilon < 1/2$.
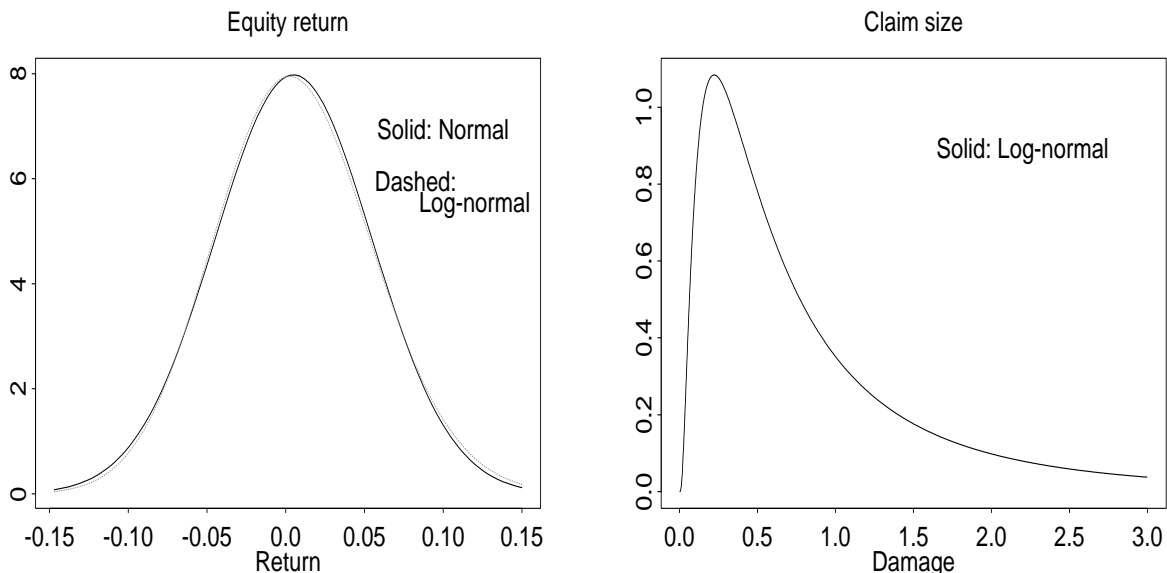
Figure 2.4 *Left: Normal and log-normal density functions for $\xi = 0.005$, $\sigma = 0.05$. Right: Log-normal for $\xi = -0.5$, $\sigma = 1$.*

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

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

Mean and standard deviation are

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

and

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

These formulae are among the most important ones in the entire theory of risk. Sampling is easy:

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

The models (1.16) and (1.17) are known as the **log-normal**. Mathematical expressions for their density functions are available, but are in this book not needed at all. Examples of their shape (obtained from simulations) are shown 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 postulated 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 adding a stochastic model for σ so that (1.14) is extended to

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

There are now two ξ-parameters ξ_x and ξ_σ distinguished through their subscripts. The random variable Z is positive and might be scaled so that $E(Z) = 1$ or $E(Z^2) = 1$ which means that $\xi_\sigma = E(\sigma)$ or $\xi_\sigma^2 = E(\sigma^2)$. What is the effect on X? Principally that a very small or large ε may occur jointly with a very large Z which opens for larger deviations from ξ_x than the normal is able to capture alone. The distribution has become **heavier-tailed**. Models where the standard deviation (in finance called **volatility**) are stochastic have drawn much interest in finance, and dynamic versions 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

$$Z = 1/G,$$

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

Dependent normal pairs
Many situations demand **correlated** normal variables. Such models are constructed by applying the linear reresentation (1.14) several times. A normal pair (X_1, X_2) is defined as

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

and a new feature is the sub-model on the right based on independent $N(0,1)$ variables η_1 and η_2. Both of ε_1 and ε_2 are $N(0,1)$ too, but they have now become dependent (or **co-variating**) in a way controlled by ρ. It will emerge in Section 5.4 that ρ is the ordinary correlation coefficient.. 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.

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

$$R_1 = e^{X_1} - 1, \qquad \text{and} \qquad R_2 = e^{X_2} - 1,$$

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 $\sigma_1 = \sigma_2 = 5\%$ (could be monthly returns on equity) have been plotted in Figure 2.5. The
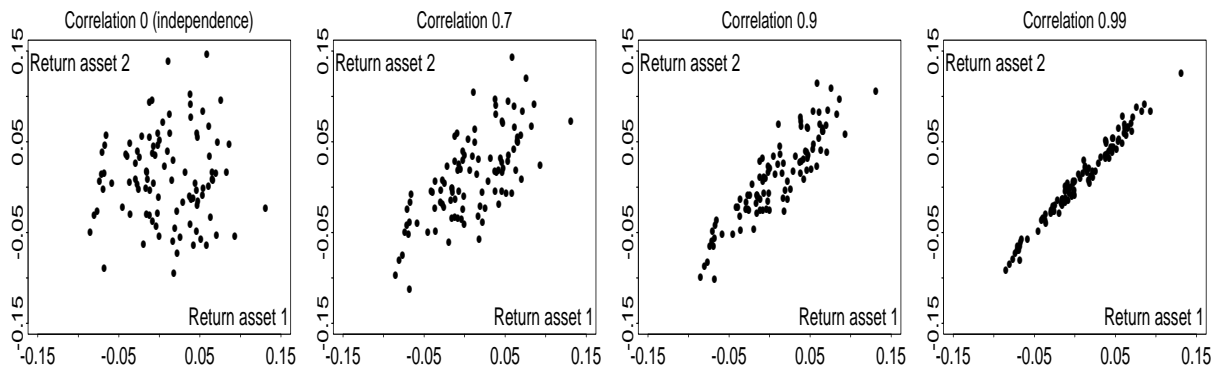
Figure 2.5 *Joint plot of* 100 *pairs of simulated equity returns; from the ordinary log-***normal** *model described in the text.*

effect of varying ρ is pronounced, yet the variation for each variable alone isn't affected at all.

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

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

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, but it does lead to a joint density function on a 'nice' mathematical form. Not much is made of this in the present book, and Exercise 2.4.5 plays 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 what it was; see Exercise 5.2.7.

Equi-correlation models

Suppose there are many interacting Gaussian variables. We start out as above by taking

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

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

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

10
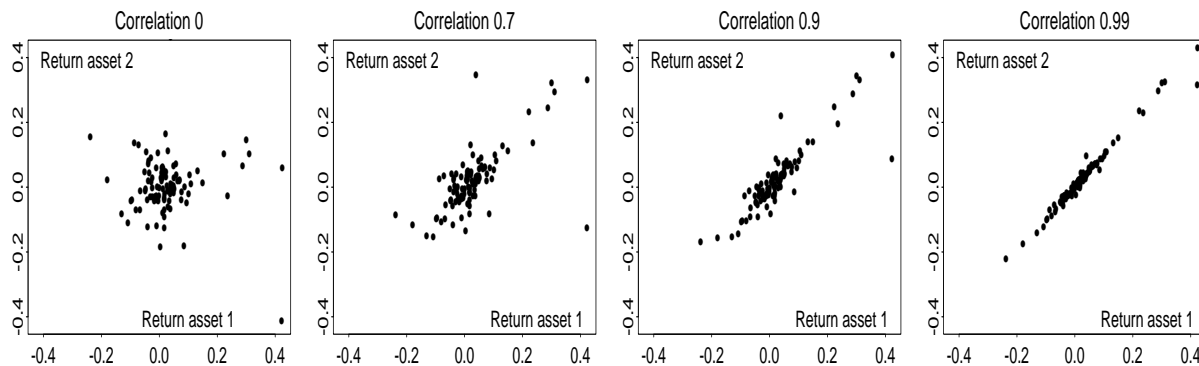
Figure 2.6 *Joint plot of* 100 *simulated financial returns; from* **stochastic volatility model** *(same as in Figure 2.5 otherwise) described in the text.*

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 common for all pairs.

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

> **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 introduced through the comment on Line 5. Some of the exercises at the end of the chapter play with this algorithm.

1.4   Generating non-uniform random variables

Introduction

The simulation algorithms in the two preceding sections were model relationships copied in the computer. This is indeed the most common way stochastic simulation algorithms are developed and has in this book influenced the way probabilistic models are being presented. But there are other ways too. Sampling 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

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

are both $N(0,1)$ and also independent; consult p.38 in Hörmann, Leydold and Derflinger (2004) 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 $\eta_1^* \leftarrow Y^* \sin(2\pi U_2^*), \quad \eta_2^* \leftarrow Y^* \cos(2\pi U_2^*)$

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. 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; see Section 2.7 for references. Our target is methods of practical usefulness in actuarial science. Actually the handful of sampling procedures in Section 2.5 take us far if we know how to apply and combine them intelligently. The present section presents three *general* techniques.

Inversion
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

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

Consider the specification 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 sampling technique. The second version 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 $X^* \leftarrow F^{-1}(U^*) \quad \text{or} \quad X^* \leftarrow F^{-1}(1-U^*)$

In either case X^* has the desired distribution function $F(x)$. The two variants 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 the next section, but first a *second* general technique:

Acceptance-rejection
Acceptance-rejection is a **random stopping** rule and much more subtle than inversion. The idea is to sample from a density function $g(x)$ of our choice. Simulations that do *not* meet a certain acceptance criterion A are discarded, and the rest will then come from the original density function

$f(x)$. Magic? It works like this. Let $g(x|A)$ be the density function of the simulations keept. By Bayes' formula (consult Section 6.2 if necessary)

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

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

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

and suppose X is accepted whenever a uniform random number U satisfies

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

Note that 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.

Example: A Gamma sampler
Some of the smartest algorithms in the business are of the acceptance/rejection type. Here is an example illustrating how it works. Consider the Gamma density

$$f(x) = Cx^{\alpha-1}x^{-\alpha x}, \qquad x > 0$$

where $\alpha > 0$ is a parameter and C a constant. Sampling isn't straightforward, and acceptance/rejection is often used. A simple scheme when $\alpha \geq 1$ is to take $g(x) = e^{-x}$, $x > 0$ with

distribution function $1 - e^{-x}$ and inversion sampler $X^* = -\log(U^*)$. It is easy to verify that $f(x)/g(x)$ attains its maximum at $x = 1$ (differentiate and see). Hence

$$M = \frac{f(1)}{g(1)} = Ce^{-\alpha+1} \qquad \text{so that} \qquad \frac{f(x)}{Mg(x)} = e^{(\alpha-1)(\log(x)-x)}.$$

and the Gamma sampler for $\alpha > 1$ becomes

$$X^* \leftarrow -\log(U^*), \qquad \text{accepted if} \quad X^* > (\alpha - 1)(\log(X^*) - X^*).$$

This is a reasonably efficient for moderate α (but not for larger ones, don't use it when $\alpha > 50$). A better (but more complex) scheme is presented in the next section.

Ratio of uniforms

This is another random stopping rule and applies to positive variables only. It is due to Kinderman and Monahan (1977) and requires $f(x)$ and $x^2 f(x)$ to be bounded functions. Let a and b be finite constants such that

$$a \geq \max_{x \geq 0} \sqrt{f(x)} \qquad \text{and} \qquad b \geq \max_{x \geq 0} x\sqrt{f(x)}. \tag{1.30}$$

They should for maximum efficiency be as small as possible (*equalities* in (1.30) are best). Let U_1 and U_2 be uniform random variables and introduce

$$Y = aU_1 \qquad \text{and} \qquad X = bU_2/Y.$$

Suppose $Y = y$ is fixed. Then X is uniform over the interval $(0, b/y)$ so that its conditional density function is $f(x|y) = y/b$ for $0 < x < b/y$ (if conditional and joint distributions is unfamiliar ground consult Chapter 6.) Multiply with $f(y) = 1/a$ (the density function of Y), and the joint density function of (X, Y) appears as

$$f(x, y) = \frac{y}{ab}, \qquad 0 < y < a, \ 0 < x < b/y.$$

Let A be the event $Y < \sqrt{X}$ and note that if $y < \sqrt{f(x)}$, then

$$y < \sqrt{f(x)} \leq a \qquad \text{and} \qquad y < \sqrt{f(x)} \leq b/x \quad \text{so that} \quad x < b/y$$

which means that A is inside the region where $f(x, y)$ is positive. But then the density function of X given that A has occurred must be

$$f(x|A) = \int_0^{\sqrt{f(x)}} C\frac{y}{ab}\, dy = \frac{C}{2ab}f(x),$$

and this only makes sense if $C = 2ab$. It follows that $f(x|A) = f(x)$, and we have:

Algorithm 2.8 Ratio of uniforms
0 Input: $f(x)$, a, b and $c = b/a$
1 Repeat
2   Draw uniforms U_1^* and U_2^*
3      $X^* \leftarrow cU_2^*/U_2^*$
4   If $aU_1^* < \sqrt{f(X^*)}$ then   stop and return X^*.

14

Good designs are those that get the search done quickly. Implementation may be carried out in terms of any function proportional to $f(x)$.

Gamma sampling again

For illustration consider again the Gamma density $f(x) = Cx^{\alpha-1}e^{-\alpha x}$ for $\alpha > 1$. The constant C is immaterial (cancels on Line 4 in Algorithm 2.8), and we may take $f(x) = x^{\alpha-1}e^{-\alpha x}$. Then

$$\sqrt{f(x)} = e^{\{(\alpha-1)\log(x)-\alpha x\}/2} \qquad \text{and} \qquad x\sqrt{f(x)} = e^{\{(\alpha+1)\log(x)-\alpha x\}/2}.$$

with maxima at $x = 1 - 1/\alpha$ and $x = 1 + 1/\alpha$ respectively. It follows that a and b in Algorithm 2.8 become

$$a = e^{(\alpha-1)(\log(1-1/\alpha)-1)/2} \qquad \text{and} \qquad b = e^{(\alpha+1)(\log(1+1/\alpha)-1)/2}$$

so that

$$c = \frac{b}{a} = e^{\{(\alpha+1)\log(1+1/\alpha)-(\alpha-1)(\log(1-1/\alpha)\}2-1}.$$

Gamma-distributed variables are for $\alpha > 1$ returned by the scheme

$$X^* \leftarrow c\frac{U_2^*}{U_1^*} \qquad \text{accepted if} \quad aU_1^* < e^{(\alpha\log(X^*)-\alpha X^*)/2}.$$

The method works reasonably for all α (and excellently when α is small), but Algorithm 2.13 below (though more complex) is still superior.

1.5 Some standard distributions

Introduction

Normals and log-normals were reviewed above, and four new distributions are now added. These six families of distributions 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 later where still other models will be introduced; see also some of the exercises to this section.

The Pareto distribution

Random variables X with density function

$$f(x) = \frac{\alpha/\beta}{(1 + x/\beta)^{1+\alpha}}, \qquad x > 0 \tag{1.31}$$

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 serve as model for large claims in property insurance; more on that in Chapter 9. Mean and standard deviation are

$$E(X) = \frac{\beta}{\alpha - 1}, \ \ \alpha > 1 \qquad \text{and} \qquad \text{sd}(X) = E(X)\sqrt{\frac{\alpha}{\alpha - 2}}, \ \ \alpha > 2. \tag{1.32}$$

They do not exist (i.e. is infinite) for other values of α than those shown. Real phenomena with α between 1 and 2 (so that the variance is infinite) will be encountered in Chapter 7.

The distribution function and its inverse of (1.31) are

$$F(x) = 1 - (1 + x/\beta)^{-\alpha}, \quad x > 0 \qquad \text{and} \qquad F^{-1}(u) = \beta\{(1-u)^{-(1/\alpha)} - 1\}, \qquad (1.33)$$

where the latter is found by solving the equation $F(x) = u$. The second version of the inversion algorithm now yields the following Pareto sampler:

Algorithm 2.9 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 (1.31) while ξ is kept fixed and α is allowed to become infinite. Then

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

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

$$f(x) = \frac{1}{\xi}e^{-x/\xi}, \qquad x > 0. \tag{1.34}$$

The fact that the exponential distribution is a limiting member of the Pareto family is of importance for extreme value methods; see Section 9.5

Mean and standard deviation of exponential variables are

$$E(X) = \xi \qquad \text{and} \qquad sd(X) = \xi, \tag{1.35}$$

and distribution and percentile functions become

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

Inversion (Algorithm 2.6) yields the following sampling method:

Algorithm 2.10 Exponential generator
0 Input ξ
1 Draw $U^* \sim$ uniform
2 Return $X^* \leftarrow -\xi \log(U^*)$ *%X* exponential*

There is a connection to Algorithm 2.9 which refects the way the exponential model was constructed from Pareto. If you insert $\beta = \alpha\xi$ on the last line of Algorithm 2.9 and let $\alpha \to \infty$, the preceding algorithm emerges.

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

$$\Pr(X_1 + \ldots + X_n < \lambda \le X_1 + \ldots + X_{n+1}) = \frac{\lambda^n}{n!}e^{-\lambda} \tag{1.36}$$

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

$$\Pr(N = n) = \frac{\lambda^n}{n!}e^{-\lambda}, \qquad n = 0, 1, \ldots. \qquad (1.37)$$

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.

$$E(N) = \lambda \qquad \text{and} \qquad \text{sd}(N) = \sqrt{\lambda}. \qquad (1.38)$$

The main point for the moment is that (1.36) 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 2.11 Poisson generator
0 Input $\lambda, Y^* \leftarrow 0$
1 For $n = 1, 2, \ldots$ do
2   Draw $U^* \sim$ uniform   and $Y^* \leftarrow Y^* - \log(U^*)$
3   If $Y^* \geq \lambda$ 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 (1.36) to be satisfied and return the number of trials minus one.

More on Poisson sampling
Poisson counts are so central in property insurance that it is worthwile elaborating a bit on its sampling. Actually the simple Algorithm 2.11 is often good enough (you see why Section 10.3), though it does slow down for large λ. If speed is critical, we may turn to the method of Atkinson (1979) which was constructed to deal with precisely that issue:

Algorithm 2.12 Atkinson's Poisson generator
0 Input: $c \leftarrow 0.767 - 3.36/\lambda, \qquad a \leftarrow \pi/\sqrt{3\lambda}, \qquad b \leftarrow \lambda a, \qquad d \leftarrow \log(c/a) - \lambda$
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 which 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 is without doubt the **Gamma** family of distributions which will be encountered repeatedly in differerent roles. The density function is

$$f(x) = \frac{(\alpha/\xi)^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\alpha x/\xi}, \ \ x > 0 \quad \text{where} \quad \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx. \tag{1.39}$$

Here $\Gamma(\alpha)$ is the **Gamma** function which satisifes $\Gamma(n) = (n-1)!$, coinciding with the factorials when α is an integer. Mean and standard deviation are

$$E(X) = \xi, \qquad \text{and} \qquad \text{sd}(X) = \xi/\sqrt{\alpha}. \tag{1.40}$$

and following McCullagh and Nelder (1992) the expectation has been made one of the two parameters (Gamma models are often presented differently.) The case $\xi = 1$ will be called the **standard** Gamma and denoted Gamma(α).

Sampling is a bit problematic. There are no convenient stochastic representations to lean on and the percentile function is complicated computationally (no fast and accurate approximations available) which makes inversion sampling unattractive. There are several good acceptance-rejection procedures available among which the following method due to Best (1978) is one of the best ones when $\alpha > 1$:

> **Algorithm 2.13 Gamma generator for $\alpha \geq 1$**
> 0 Input: ξ, α and $b = \alpha - 1$, $c = 3\alpha - 0.75$.
> 1 Repeat
> 2 $\quad$ Sample $U^* \sim$ uniform
> 3 $\quad$ $W^* \leftarrow U^*(1 - U^*), \quad Y^* \leftarrow \sqrt{c/W^*}(U^* - 0.5), \quad X^* \leftarrow b + Y^*$
> 6 $\quad$ If $X^* > 0$ then
> 7 $\quad\quad$ Sample $V^* \sim$ uniform$(0, 1)$
> 8 $\quad\quad$ $Z^* \leftarrow 64(W^*)^3(V^*)^2$
> 9 $\quad\quad$ If $Z^* \leq 1 - 2(Y^*)^2/X^* \quad$ **or** $\quad$ if $\log(Z^*) \leq 2\{b(\log(X^*/b) - Y^*)\}$
> $\quad\quad\quad\quad$ 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 a result due to Stuart (1962); i.e.

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

Here Y and U are independent. Computer commands are summarized as follows:

> **Algorithm 2.14 Gamma generator for $\alpha < 1$**
> 0 Input: ξ, α
> 1 Sample $Z^* \sim$ Gamma$(1 + \alpha)$ $\quad\quad\quad$ *%From Algorithm 2.13*
> 2 Sample $U^* \sim$ uniform
> 3 Return $Z^* \leftarrow \xi Z^*(U^*)^{1/\alpha}$

Together Algorithms 2.13 and 2.14 yield quick sampling, though slower than many of the earlier algorithms.

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 $q_{1\epsilon}/q_{2\epsilon} \to 0$ as $\epsilon \to 0$ which is the condition (1.8) in Section 2.2. Since both numerator and denominator tend to zero as $\epsilon \to 0$, we may apply l'Hôpital's rule which yields

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

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

$$f_1(q_{1\epsilon})\frac{\partial q_{1\epsilon}}{\partial \epsilon} = -1, \quad \text{and} \quad f_2(q_{2\epsilon})\frac{\partial q_{2\epsilon}}{\partial \epsilon} = -1, \tag{1.41}$$

so that

$$\frac{f_2(q_{2\epsilon})}{f_1(q_{1\epsilon})} = \frac{\dfrac{\partial q_{1\epsilon}}{\partial \epsilon}}{\dfrac{\partial q_{2\epsilon}}{\partial \epsilon}} \to 0 \quad \text{as} \quad \epsilon \to 0$$

and the ratio $f(q_{1\epsilon}/f(q_{2\epsilon}) \to \infty$ 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

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

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

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

using (1.41). 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 $f(x) = \exp(-x)$ for $x > 0$. To verify the Poisson generator in Algorithm 2.10 we have to evaluate the probability

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

which is exercise in conditional probabiltities. Let $n > 1$ and note that

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

19

or

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

which can also be written

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

This is a recursion starting at

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

The solution is

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

as claimed in (1.36). This 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   Bibliographical notes

Statistiscs Parts of this chapter have drawn on fairly elementary results from statistics. Kendall, Stuart and Ord (1994) is a thorough, practical review of this topic containing many of the central distributions. The non-parametric aspect is treated (for example) in Wasserman (2006) whereas Scott (1992) and Wand and Jones (1995) are specialist monographs on density estimation. For univariate distributions see Johnson, Kotz and Balakrishnan (1994) (the continuous case) and Johnson, Kemp and Kotz (2005) (the discrete one), Balakrishnan and Nevzorov (2003) (both continuous and discrete) and in an actuarial and financial context Klugman, Panjer and Willmot (1998) and Kleiber and Kotz (2003). Many of the most common distributions in general insurance are also reviewed in Panjer and Wilmot (1992), Beirlant, Teugels and Vynckier (1996) and Klugman (2004). Gaussian models and stochastic volatility are treated much more thoroughly in Chapters 5 and 13 with references given in Sections 5.9 and 13.8.

Sampling The best handbook ever written on the sampling of non-uniform random variables may be Devroye (1986) (most of these algorithms had been discovered by 1986). An alternative is Hörmann, Leydold, and Derflinger (2004). Many of the continuous distributions used in this book can be sampled by inversion, but the Gamma model as an exception. Algorithms 2.13 and 2.14 are due to Best (1978) and Stuart (1962). Other possibilities for Gamma sampling are presented in Gentle (2003), for example the method due to Cheng and Feast (1979). Smart algorithms for some of the central distributions in actuarial science are presented in Ahrens and Dieter (1974).

Programming What platforms should you go for? High-level software packages are Splus or R (which is the same), MATLAB, Maple and Mathematica. All of them allow easy implementation with sampling generators for the most common distributions available as built-in routines. Much

information is provided by the web-sites[2]; for textbooks consult Venables and Ripley (2002), Zivot and Wang (2003) or Krause and Olson (2005) (for Splus), Hunt, Lipsman and Rosenberg (2001) or Otto and Denier (2005) (MATLAB), Cornil and Testud (2002) or Dagpunar (2007) (Maple) and Wolfram (1999), Rose and Smith (2001) or Landau and Wangberg (2005) (Mathematica). Many problems can be successfully handled by these platforms, and you may even try Excell. The advantage is quick implementation, but for large problems such programs may run uncomfortably slowly (vectorization helps; avoid for-loops if possible). If you are an Excell user, you might be familiar with Visual Basic which is still another possibility (see Schneider (2006) for a reference), but if speed is needed, choose C, Fortran or Pascal. All experiments in this book have been coded in Fortran90, and in most cases the computer time was seconds or less. Introductions to these programming languages are Stoustrup (1997) and Harbison and Steele (2002) (for C), Ellis, Philips and Lahey (1994) and Chivers and Sleightholme (2006) (Fortran) and Savitch (1995) (Pascal). Parallel processing may allow even higher speed, but this hasn't been used much in insurance and finance. Grama, Gupta, Korypis and Kumar (2003) is a general introduction (with examples from engineering and natural sciences); see also Nakano (2004) in the context of statistics.

Abramowitz, M. and Stegun, I. (1965). *Handbook of Mathematical Functions.* Dover, New York.

Ahrens, J. and Dieter, U. (1974). Computer Methods for Sampling from Gamma, Beta, Poisson and Binomial Distributions. *Computing*, 12, 223-246.

Atkinson, A.C. (1979). The Computer Generation of Poisson Random Variables. *Applied Statistiscs*, 28, 29-35.

Balakrishnan, N. and Nevzorov, V.B. (2003). *A Primer on Statistical Distributions.* John Wiley & Sons, Hoboken, New Jersey.

Beirlant, J., Teugels, J.L and Vynckier, P. (1996). *Practical Analysis of Extreme Values.* Leuven University Press, Leuven.

Best, P.J. (1978). Letter to the Editor. *Applied Statistics*, 28, 181.

Cheng, R.C.H and Feast, G.M. (1979). Some Simple Gamma Variable Generators. *Applied Statistics*, 28, 290-295.

Chivers, I.D. and Sleightholme, J. (2006). *Introduction to Programming with Fortran.* Springer-Verlag, London.

Cornil, J-M. and Testud, P. (2002). *Introduction to Maple V.* Springer, Berlin.

Devroye, L. (1986). *Non-uniform Random Variate Generation.* Springer-Verlag, Berlin.

Dagpunar, J.S. (2007). *Simulation and Monte Carlo with Application in Finance* John Wiley & Sons, Chichester.

Ellis, T.M.R., Philips, I.R. and Lahey, T.M. (1994). *Fortran 90 Programming.* Addison-Wesley, Harlow, Essex

Grama, A., Gupta, A., Korypis, G. and Kumar, V, second ed. (2003). *Introduction to Parallel Computing.* Pearson/Addison-Wesley, Harlow, Essex.

Gentle, J.E. (2003). *Random Number Generation and Monte Carlo Methods.* Springer-Verlag, New York.

Harbison, S.P. and Steele, G.L. (2002), fifth ed. *C: a Referee Manual.* Prentice Hall, Englewood Cliffs, New Jersey.

Hörmann, W., Leydold, J. and Derflinger, G. (2004). *Automatic Non-Uniform Random Variate Generation.* Springer-Verlag, Berlin.

---

[2]http://www.r-project.org for R (and Splus), http://www.mathworks.com for MATLAB, http://www.maplesoft.com for Maple and http://www.wolfram.com for Mathematica.

Hunt, B.R., Lipsman, R.L. and Rosenberg, J.M. (2001). *A Guide to Matlab.* Cambridge University Press, Cambridge.

Jäckel. P. *Monte Carlo Methods in Finance.* John Wiley & Sons, Chichester.

Johnson, N.L. Kotz, S. and Balakrishnan, N (1994). *Continuous Univariate, Distributions.* John Wiley & Sons, New York.

Johnson, N.L., Kemp, A.W. and Kotz, S. (2005). *Univariate Discrete Distributions.* John Wiley & Sons, Hoboken, New Jersey

Kendall, M.G., Stuart, A. and Ord, K. (1994), sixth ed. *Kendall's Advanced Theory of Statistiscs. Volume 1. Distribution Theory.* Edward Arnold, London.

Kinderman, A.J. and Monahan, J.F. (1977). Computer Generation of Random Variables Using Ratio of Uniform Deviates. *ACM Transactions of Mathematical Software.*, 3, 257-260

Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economic and Actuarial Sciences.* John Wiley & Sons, Hoboken, New Jersey.

Klugman, S.A. (2004). Continuous Parametric Distributions. In *Encyclopedia of Actuarial Science*, Teugels, J, and Sundt, B. (eds) John Wiley & Sons, Chichester, 357-362.

Klugman, S.A., Panjer, H. and Willmot, G.E. (1998). *Loss Models: from Data to Decisions.* John Wiley & Sons, New York.

Krause, A. and Olson, M. (2005). *The Basics of Splus.* Springer Science + Business Media, New York

Landau, R.H. and Wangberg, R. (2005). *A first Course in Scientific Computing: Symbolic, graphic, and Numerical Modeling Using Maple, Java, Mathematica and Fortran90.* Princeton University Press, Princeton, New Jersey.

McCullagh, P. and Nelder, J.A. (1989), second ed. *Generalized Linear Models.* Chapman & Hall, London.

Nakano, J. (2004). Parallel Computing Techniques. In *Handbook of Computational Statistics. Concepts and Methods.* Gentle, J.E., Härdle, W. and Mori, Y (eds). Springer Verlag, New York, 237-266.

Panjer, H. and Wilmot, G. E. (1992). *Insurance Risk Models* The Actuarial Foundation, Schaumburg, Illinois.

Odeh, R.E. and Evans, J.O. (1974). The Percentage Points of the Normal distribution. *Applied Statistics*, 23, 96-97.

Otto, S.R. and Denier, J.P. (2005). *An Introduction to Programming and Numerical methods in MATLAB.* Springer Verlag, London.

Rose, C. and Smith, M.D. (2001). *Mathematical Statistics with Matematica.* Springer Verlag, New York.

Savitch, W. (1995). *Pascal: an Introduction to the Art and Science of Programming.* Benjamin/Cummings. Redwood City, California.

Schneider, D.I. (2006). *An Introduction to Programming Using Visual Basic 2005.* Pearson/Prentice Hall. Upper Saddle River, New Jersey.

Scott, P.W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization.* John Wiley & Sons, New York.

Stoustrup, P. (1997), third ed. *The C++ Programming Language.* Addison-Wesley, Reading, Massachusetts.

Venables, W.N. and Ripley, B. (2002), fourth ed. *Modern Applied Statistics with S.* Springer Verlag, New York.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* Chapman & Hall/CRC, Boa Raton,

Florida.

Wasserman, L. (2006). *All of Nonparametric Statistics.* Springer, New York.

Wolfram, S. (1999). *The Matematica Book..* Wolfram Media, Champaign, Illinois.

Zivot, E. and Wang, J. (2003). *Modelling Financial Time Series with S-plus.* Springer-Verlag Berlin.

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 (1.33). 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 $\sigma = \mathrm{sd}(X)$. 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

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

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

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

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

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

25

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

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

A possible model for Z is to make it log-normal, for example $Z = \exp(-\tau^2 + 2\tau\eta)$ where $\eta \sim N(0, 1)$, $\tau \geq 0$ 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

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

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 $\rho_\eta = \mathrm{cor}(\eta_1, \eta_2)$. **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.

$$\mathrm{med}(X) = q_{0.5} \qquad \text{and} \qquad \mathrm{qd}(X) = q_{0.75} - q_{0.25} \tag{1.42}$$

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

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

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 (1.42). **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

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

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 (1.42).

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

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

Once again the parameters $\alpha, \beta > 0$. **a)** Derive the inversion sampler. **b)** Determine mathematical expressions for med(X) and qd(X); see (1.42).

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

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

a) Derive its inversion sampler. **b)** Find mathematical expressions for med(X) and qd(X), see (1.42).

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

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

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!