Chapter 7
Empirical Methods

In this chapter we consider a modeling approach that uses a set of historical data, such as bond prices, share prices, claim sizes, or exchange rates, to model the value at a future time $T > 0$ of portfolios whose values depend on a given set of assets and possibly also liabilities. Here we want the data to speak for themselves in the sense that the model for the future values should only be based on information available in the given historical data samples. The assumption we make is therefore that the information in the samples is representative of future values and that no additional probability beliefs of the modeler are relevant.

Historical share prices $S_{-n}, \ldots, S_0$ of a stock over the last $n+1$ time periods are not necessarily good representatives of possible values for the future share price S_1. But the sample of historical returns $R_{-k} = S_{-k+1}/S_{-k}$, for $k = n-1, \ldots, 1$, may be assumed to be a good representative of possible values for the future return $R_1 = S_1/S_0$ over the next time period. Similarly, the historical zero rates $r_{-n}, \ldots, r_0$, corresponding to a given time to maturity, may be transformed into zero rate changes $r_{-k+1} - r_{-k}$, for $k = n-1, \ldots, 1$, that can be viewed as good representatives of the possible zero rate change $r_1 - r_0$ over the next time period. If we believe in this approach, then appropriate transformations of the historical samples produce samples of the random values, e.g., returns, that determine the future portfolio values. If the generated sample of returns or value changes can be viewed as samples from independent and identically distributed random variables, then standard statistical techniques can be used to investigate the probability distribution of future portfolio values, expressed as known functions of future returns or value changes.

In this chapter, we will investigate this approach to modeling the future. This is a subjective approach just as any other approach (such as assigning a parametric probability distribution to the future portfolio value). However, it is fully nonparametric and is a reasonable approach if we believe that the mechanism that produced the returns in the past is the same as the mechanism that will produce returns in the future, even if the mechanism is unknown to us.

H. Hult et al., *Risk and Portfolio Analysis: Principles and Methods*, Springer Series in Operations Research and Financial Engineering, DOI 10.1007/978-1-4614-4103-8_7,

The first topic of this chapter is how to turn historical prices into a sample from the distribution of the future portfolio value under the assumption that returns over the next time period will be similar to those returns. This material is presented in Sect. 7.1. In Sect. 7.2 we consider the empirical distribution, which is the probability distribution derived from a data sample. The quantile function of the empirical distribution is the empirical quantile studied in Sect. 7.3. The main objective in these two sections is to investigate how the accuracy of empirical probabilities and quantiles, relative to the true unknown quantities they are measuring, varies with the sample size and characteristics of the unknown distribution from which the sample of observations is generated. Empirical distributions and quantiles provide natural estimators of value-at-risk (VaR) and expected shortfall (ES), which are presented in Sect. 7.4. Point estimates of risk measures are not particularly useful unless they are accompanied by estimates of their accuracy. Therefore, we analyze in detail methods for constructing confidence intervals for the quantities estimated by empirical estimators. In Sect. 7.5, we present a method for constructing exact confidence intervals for quantiles and a method for constructing approximative confidence intervals using the nonparametric bootstrap procedure. The latter method is further studied in Sect. 7.6, which deals with the uncertainty in estimates for solvency capital requirements for a nonlife insurer.

7.1 Sample Preparation

Denote the current time by 0, and consider a future time that we call time 1. Let V_1 be the random value of some portfolio at time 1 that we can express as a function of the vector $\mathbf{S}_1$ of asset prices at time 1. For the sake of clarity of presentation we take $\mathbf{S}_1$ to be the share prices of some stocks. It is assumed that we have access to a sample $\{\mathbf{S}_{-n}, \mathbf{S}_{-n+1}, \ldots, \mathbf{S}_0\}$ of vectors of historical prices from the n previous equally spaced points in time (e.g., days, weeks) and from the current time. It is clear that the sample points may be strongly dependent (the share price on any given day is strongly dependent on the previous day's price). Moreover, it is likely that the asset prices $\mathbf{S}_{-k}$, from k time periods ago, are quite different from what can be anticipated for $\mathbf{S}_1$, at least if k is large. The sample of historical asset prices may, however, be transformed into a sample of vectors of returns $\mathbf{R}_{-n+1}, \ldots, \mathbf{R}_0$, where

$$\mathbf{R}_{-k} = \left(R^1_{-k}, \ldots, R^d_{-k}\right)^{\mathrm{T}} \text{ with } R^l_{-k} = S^l_{-k}/S^l_{-k-1}$$

for $k = 0, \ldots, n-1$, and $l = 1, \ldots, d$. It is often reasonable to assume, supported by statistical analysis, that the points of the sample $\{\mathbf{R}_{-n+1}, \ldots, \mathbf{R}_0\}$ are weakly dependent and close to identically distributed and have distributional characteristics that are representative also for $\mathbf{R}_1$, the vector of percentage returns for the next time period. The portfolio value at time 1 is $V_1 = f(\mathbf{R}_1)$ for some function f that depends on information available at time 0 such as the current asset prices $\mathbf{S}_0$. Then

the sample $\{\mathbf{R}_{-n+1}, \ldots, \mathbf{R}_0\}$ of return vectors can be transformed into the sample $\{f(\mathbf{R}_{-n+1}), \ldots, f(\mathbf{R}_0)\}$ from the probability distribution of $V_1 = f(\mathbf{R}_1)$. If the vectors in the former sample are approximately independent copies of $\mathbf{R}_1$, then the vectors in the latter sample are approximately independent copies of V_1.

The transformation of historical prices into historical returns is not essential for the sample preparation scheme to work. Returns could, for instance, be replaced by something else, such as price differences. The essential point is that the original sample $\{\mathbf{S}_{-n}, \mathbf{S}_{-n+1}, \ldots, \mathbf{S}_0\}$ is transformed into a sample $\{\mathbf{Z}_{-n+1}, \ldots, \mathbf{Z}_0\}$, which in turn could be transformed into a sample $\{f(\mathbf{Z}_{-n+1}), \ldots, f(\mathbf{Z}_0)\}$ whose points may be viewed as independent copies of the future portfolio value V_1. This situation is the desired starting point for statistical analysis. From a sample of independent and identically distributed random variables drawn from the unknown probability distribution of V_1, statistical methods can be applied to investigate the probability distribution of the future portfolio value V_1.

The approach presented for generating a sample from the probability distribution of the future portfolio value is based on the assumption that changes in values in the past contain relevant information for assessing the probability distribution of changes in value from now until the future time we are considering. Determining the extent to which this assumption is reasonable requires some serious thinking. Big changes in the legal or political environment, monetary policies of governments or central banks, or other events may make it hard to justify this assumption.

Throughout the book we write $\{\mathbf{S}_{-n}, \mathbf{S}_{-n+1}, \ldots, \mathbf{S}_0\}$ for the random vectors of historical prices (and similarly for the sample of returns) and $\mathbf{s}_{-n}, \mathbf{s}_{-n+1}, \ldots, \mathbf{s}_0$ for the actual observations of the historical prices. The following example illustrates the sample preparation approach.

Example 7.1 (Sample preparation). Consider a portfolio consisting of long positions in two different assets, one unit of the first asset and two units of the second asset. The daily prices per unit of the two assets over the last 20 days are given by S_t^1 and S_t^2 for $t = -20, \ldots, 0$. Suppose the corresponding pairs of returns

$$\mathbf{R}_t = (R_t^1, R_t^2) = (S_t^1/S_{t-1}^1, S_t^2/S_{t-1}^2), \quad t = -19, \ldots, 0,$$

are independent and identically distributed. If V_1 is the value of the portfolio at time 1, then

$$V_1 = S_1^1 + 2S_1^2 = S_0^1 \frac{S_1^1}{S_0^1} + 2S_0^2 \frac{S_1^2}{S_0^2} = S_0^1 R_1^1 + 2S_0^2 R_1^2 = f(\mathbf{R}_1),$$

where $f(x, y) = S_0^1 x + 2S_0^2 y$. The random variables $\{f(\mathbf{R}_{-20+1}), \ldots, f(\mathbf{R}_0)\}$ can be viewed as a sample of independent copies of V_1.

It may happen that we have access to daily historical prices and want to use the data to investigate the probability distribution of the value of a portfolio a week (month or year) from now. Then there are different options available. Consider the

sample $\{\mathbf{S}_{-n}, \mathbf{S}_{-n+1}, \ldots, \mathbf{S}_0\}$ of vectors of historical prices and suppose we want to investigate the distribution of V_T, where $T > 1$. We assume that the original sample can be transformed into a sample $\{\mathbf{R}_{-n+1}, \ldots, \mathbf{R}_0\}$ of vectors of returns such that the vectors are approximately independent copies of $\mathbf{R}_1$ and that $V_T = f(\mathbf{R}_1 \cdots \mathbf{R}_T)$, where $\mathbf{R}_1 \cdots \mathbf{R}_T$ is interpreted as componentwise multiplication and $\mathbf{R}_1 \cdots \mathbf{R}_T$ is the vector of returns over the next period of length T.

Example 7.2 (Thinning of the sample). One way of obtaining a sample of vectors of returns over time periods of length T would be to start with the sample $\{\mathbf{S}_{-T[n/T]}, \ldots, \mathbf{S}_{-T}, \mathbf{S}_0\}$ and set

$$\mathbf{R}_{-k}^{(T)} = ((R^{(T)})_{-k}^1, \ldots, (R^{(T)})_{-k}^d)^{\mathrm{T}} \text{ with } (R^{(T)})_{-k}^l = S_{-Tk}^l / S_{-T(k+1)}^l$$

for $k = 0, \ldots, [n/T] - 1$, and $l = 1, \ldots, d$. Here $[y]$ denotes the largest integer smaller than or equal to y, i.e., $[y] = \max\{k \in \mathbb{N} : k \leq y\}$. The sample $\{\mathbf{R}_{-[n/T]+1}^{(T)}, \ldots, \mathbf{R}_0^{(T)}\}$ is a sample of vectors of returns over nonoverlapping time periods of length T. If these return vectors are independent copies of $\mathbf{R}_1 \cdots \mathbf{R}_T$, then $f(\mathbf{R}_{-[n/T]+1}^{(T)}), \ldots, f(\mathbf{R}_0^{(T)})$ are independent copies of V_T. The problem with this approach is that much of the possibly relevant information in the original sample $\{\mathbf{S}_{-n}, \mathbf{S}_{-n+1}, \ldots, \mathbf{S}_0\}$ is ignored and the sample size is reduced from n to $[n/T]$.

Example 7.3 (Historical simulation). An approach that, unlike the approach in Example 7.2, uses the entire original sample is to draw with replacement T vectors from the sample $\{\mathbf{R}_{-n+1}, \ldots, \mathbf{R}_0\}$ and form the componentwise product of these vectors, denoted by $\mathbf{R}_1^{*(T)}$. Repeat the procedure m times to obtain the sample $\{\mathbf{R}_1^{*(T)}, \ldots, \mathbf{R}_m^{*(T)}\}$ of fictive return vectors over time periods of length T. If the original return vectors $\mathbf{R}_{-n+1}, \ldots, \mathbf{R}_0$ are independent and identically distributed, then the vectors $\mathbf{R}_1^{*(T)}, \ldots, \mathbf{R}_m^{*(T)}$ are identically distributed but not independent since some of the random indices may take the same index value, but they are conditionally independent given $\mathbf{R}_{-n+1}, \ldots, \mathbf{R}_0$.

The sample $\{f(\mathbf{R}_1^{*(T)}), \ldots, f(\mathbf{R}_m^{*(T)})\}$ is a sample of size m, where the sample points are approximately distributed as V_T. This approach to generating a sample from the distribution of V_T is called a historical simulation. On the one hand, all the original sample points are used and the sample size m can be chosen arbitrarily large. On the other hand, the original return vectors appear as factors in more than one of the fictive return vectors $\mathbf{R}_k^{*(T)}$, so there may be substantial redundancy in the constructed sample of return vectors over periods of length T.

7.2 Empirical Distributions

Consider observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ of independent and identically distributed d-dimensional random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with a common unknown distribution function $F(\mathbf{x}) = \mathrm{P}(\mathbf{X} \leq \mathbf{x})$, where $\mathbf{X}$ is an independent copy of $\mathbf{X}_k$ and $\mathbf{X} \leq \mathbf{x}$ is

interpreted as an inequality for all the components; $\mathbf{X} \leq \mathbf{x}$ if and only if $X_j \leq x_j$ for $j = 1, \ldots, d$. Suppose that we want to compute some quantity $\theta = \theta(F)$ that depends on F, for instance, the mean, the variance, a quantile, or a risk measure. It is impossible to compute θ since F is unknown, but the observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ allow us to approximate the unknown distribution by that obtained from assigning a probability weight $1/n$ to each of the $\mathbf{x}_k$. That is, approximating the unknown $F(\mathbf{x})$ by the fraction $F_n(\mathbf{x})$ of the $\mathbf{x}_k$ that are smaller than or equal to $\mathbf{x}$,

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{k=1}^{n} I\{\mathbf{x}_k \leq \mathbf{x}\}.$$

The distribution function F_n is called the empirical distribution function of $\mathbf{x}_1, \ldots, \mathbf{x}_n$. The random counterpart, which is the empirical distribution associated with the random sample $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$, is given by

$$F_{n,\mathbf{X}}(\mathbf{x}) = \frac{1}{n} \sum_{k=1}^{n} I\{\mathbf{X}_k \leq \mathbf{x}\}.$$

Note that $F_{n,\mathbf{X}}$ is a random object whose outcome F_n is a distribution function.

The (strong) law of large numbers says that if $Z_1, Z_2, \ldots$ is a sequence of independent copies of a random variable Z for which the expected value $\mathrm{E}[Z]$ exists finitely, then

$$\frac{1}{n} \sum_{k=1}^{n} Z_k \to \mathrm{E}[Z] \quad \text{with probability 1 as } n \to \infty.$$

If we choose $Z_k = I\{\mathbf{X}_k \leq \mathbf{x}\}$, then $\mathrm{E}[Z_k] = \mathrm{P}(\mathbf{X}_k \leq \mathbf{x}) = F(\mathbf{x})$ and the law of large numbers implies that, with probability one, $\lim_{n\to\infty} F_{n,\mathbf{X}}(\mathbf{x}) = F(\mathbf{x})$. In particular, the empirical distribution function $F_{n,\mathbf{X}}$ is a good approximation of the unknown distribution function F as long as the sample size n is sufficiently large. Similarly, if we choose $Z_k = h(\mathbf{X}_k)$, then $\mathrm{E}[Z_k] = \mathrm{E}[h(\mathbf{X}_k)] = \mathrm{E}[h(\mathbf{X})]$ and the law of large numbers implies that, with probability one,

$$\int h(\mathbf{x}) dF_{n,\mathbf{X}}(\mathbf{x}) = \frac{1}{n} \sum_{k=1}^{n} h(\mathbf{X}_k) \to \mathrm{E}[h(\mathbf{X})] = \int h(\mathbf{x}) dF(\mathbf{x}) \quad \text{as } n \to \infty. \tag{7.1}$$

In particular, the expression on the left-hand side of (7.1) is a good approximation of the expression on the right-hand side as long as the sample size n is sufficiently large.

Example 7.4 (Sample mean and variance). Consider a sample $\{x_1, \ldots, x_n\}$ and the corresponding empirical distribution function F_n. The sample mean

$\overline{x} = (x_1 + \cdots + x_n)/n$ is simply the expected value of a random variable with the distribution function F_n:

$$\int x dF_n(x) = \frac{1}{n}\sum_{k=1}^{n} x_k = \overline{x}.$$

We know from (7.1) that $\overline{X} = (X_1 + \cdots + X_n)/n \to \mathrm{E}[X]$ with probability one as $n \to \infty$, and it is easy to see that $\mathrm{E}[\overline{X}] = \mathrm{E}[X]$. The variance of F_n is

$$\begin{aligned}\int x^2 dF_n(x) - \left(\int x dF_n(x)\right)^2 &= \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \frac{1}{n^2}\left(\sum_{k=1}^{n} x_k\right)^2 \\ &= \frac{1}{n}\left(\sum_{k=1}^{n} x_k^2 - n\overline{x}^2\right) \\ &= \frac{1}{n}\left(\sum_{k=1}^{n} x_k^2 - 2\sum_{k=1}^{n}\overline{x}x_k + \sum_{k=1}^{n}\overline{x}^2\right) \\ &= \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2.\end{aligned}$$

We know from (7.1) that, with probability one,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(X_k - \overline{X})^2 = \mathrm{Var}(X).$$

However, the expected value of the variance estimator is not equal to $\mathrm{Var}(X)$. Therefore, the variance is typically estimated by the sample variance

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2,$$

which satisfies $\mathrm{E}[S^2] = \mathrm{Var}(X)$.

Consider a subset B of $\mathbb{R}^d$ and suppose that we want to estimate the probability $P(B) = \mathrm{P}(\mathbf{X} \in B)$. Similarly to the empirical distribution function, we form the empirical estimator

$$P_{n,\mathbf{X}}(B) = \frac{1}{n}\sum_{k=1}^{n} I\{\mathbf{X}_k \in B\}. \tag{7.2}$$

Notice that $P_{n,\mathbf{X}}(B) = F_{n,\mathbf{X}}(\mathbf{x})$ if $B = \{\mathbf{y} : \mathbf{y} \leq \mathbf{x}\}$ and that the sum in (7.2) is $\text{Bin}(n, P(B))$-distributed. In particular, from the expected value $nP(B)$ and variance $nP(B)(1 - P(B))$ of the binomially distributed sum in (7.2) we find that

$$\text{E}[P_{n,\mathbf{X}}(B)] = P(B) \quad \text{and} \quad \text{Var}(P_{n,\mathbf{X}}(B)) = \frac{1}{n} P(B)(1 - P(B)).$$

Moreover, it follows from the law of large numbers that $\lim_{n\to\infty} P_{n,\mathbf{X}}(B) = P(B)$ with probability one.

Example 7.5 (Estimation of small probabilities). In this example, we investigate the sample size needed for accurate empirical estimation of a small probability $P(B)$. A common measure of the accuracy of an estimator is the relative error—the standard deviation of the estimator divided by the estimated quantity. In this context, the relative error is given by

$$\frac{\text{Var}(P_{n,\mathbf{X}}(B))^{1/2}}{P(B)} = n^{-1/2} \Big(\frac{1}{P(B)} - 1 \Big)^{1/2}.$$

It is natural to require that the standard deviation of the estimator must be at least no greater than the probability to be estimated. Under this requirement, since $P(B)$ is assumed to be small, we find that $n \approx 1/P(B)$, which corresponds to a very large required sample size if $P(B)$ is small.

The accuracy of the estimator can be investigated by considering the probability

$$\text{P}\Big(\Big| \frac{P_{n,\mathbf{X}}(B) - P(B)}{P(B)} \Big| < \varepsilon \Big) = \text{P}\Big(1 - \varepsilon < \frac{P_{n,\mathbf{X}}(B)}{P(B)} < 1 + \varepsilon \Big).$$

Since the sum in (7.2) is $\text{Bin}(n, P(B))$-distributed, we find that

$$\text{P}\Big(\frac{P_{n,\mathbf{X}}(B)}{P(B)} < 1 + \varepsilon \Big) = \sum_{k=0}^{[n(1+\varepsilon)P(B)]} \binom{n}{k} P(B)^k (1 - P(B))^{n-k},$$

and similarly with $1 - \varepsilon$ instead of $1 + \varepsilon$.

Another approach to investigating the accuracy of the estimator when the sample size n is large is to apply the central limit theorem. If $Z_1, Z_2, \ldots$ is a sequence of independent copies of a random variable Z with finite expected value μ and standard deviation σ, then

$$\lim_{n\to\infty} \text{P}\Big(\frac{Z_1 + \cdots + Z_n - n\mu}{n^{1/2}\sigma} \leq x \Big) = \Phi(x) \quad \text{for all } x,$$

where Φ denotes the standard normal distribution function. Taking $Z_k = I\{\mathbf{X}_k \in B\}/n$ we find that

$$\lim_{n\to\infty} \mathrm{P}\left(\left(\frac{n}{P(B)(1-P(B))}\right)^{1/2}(P_{n,\mathbf{X}}(B) - P(B)) \le x\right) = \Phi(x) \quad \text{for all } x.$$

In particular, $P_{n,\mathbf{X}}(B)$ is approximately $\mathrm{N}(P(B), P(B)(1-P(B))/n))$-distributed if n is large.

7.3 Empirical Quantiles

Here we consider observations $x_1, \ldots, x_n$ from independent and identically distributed random variables $X_1, \ldots, X_n$ with a common unknown distribution function F defined on the real line $\mathbb{R}$. The empirical quantile function F_n^{-1} is the quantile function of the empirical distribution function F_n and therefore given by

$$F_n^{-1}(p) = \min\{x : F_n(x) \ge p\}.$$

Similarly, the empirical quantile function $F_{n,X}^{-1}$ is the quantile function of $F_{n,X}$. We will now show that the empirical quantile $F_{n,X}^{-1}(p)$ is the kth largest of the sample points $X_1, \ldots, X_n$ (and therefore the same holds for F_n^{-1} in terms of the sample points $x_1, \ldots, x_n$), where $k = k(n, p)$ depends on n and p. It turns out to be useful to order the sample $\{X_1, \ldots, X_n\}$ such that $X_{1,n} \ge \cdots \ge X_{n,n}$ (if F is continuous, then with probability one there are no $j \ne k$ such that $X_j = X_k$, i.e., no ties). Note that

$$\min\{x : F_{n,X}(x) \ge p\} = \min\left\{x : \sum_{k=1}^{n} I\{X_{k,n} \le x\} \ge np\right\}. \tag{7.3}$$

Since the sum $\sum_{k=1}^{n} I\{X_{k,n} \le x\}$ can only take integer values, we see that the right-hand side of (7.3) is equal to $X_{j,n}$ for some j. Which j? Note that for any j in the set $\{1, \ldots, n\}$,

$$\sum_{k=1}^{n} I\{X_{k,n} \le X_{j,n}\} = \sum_{k=j}^{n} I\{X_{k,n} \le X_{j,n}\} = n - j + 1,$$

and we must look for the largest j such that the last expression is greater than or equal to np. If we take $j = [n(1-p)] + 1$, then

$$n - j + 1 = n - [n(1-p)] \ge n - n(1-p) = np,$$

with equality if and only if np is an integer. In particular, every $j \ge [n(1-p)] + 2$ gives $n - j + 1 < np$. We conclude that the empirical quantile function is given by

$$F_{n,X}^{-1}(p) = X_{[n(1-p)]+1,n}, \quad p \in (0, 1),$$

a piecewise constant function on $(0, 1)$ with

$$X_{[n(1-p)]+1,n} = X_{k,n} \text{ if } p \in (1 - k/n, 1 - (k-1)/n]. \tag{7.4}$$

It can be shown that if F is strictly increasing, then $\mathrm{P}(\lim_{n\to\infty} F^{-1}_{n,X}(p) = F^{-1}(p)) = 1$ for all $p \in (0, 1)$. Therefore, the empirical quantile is an arbitrary good approximation of the true but unknown quantile if the sample size n is sufficiently large. We prove the following slightly weaker statement.

Proposition 7.1. *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with common distribution function F, and let $F_{n,X}$ be the empirical distribution function of the first n elements of the sequence. If F is strictly increasing in a neighborhood of $F^{-1}(p)$, then $\lim_{n\to\infty} \mathrm{P}(|F^{-1}_{n,X}(p) - F^{-1}(p)| > \varepsilon) = 0$ for every $\varepsilon > 0$.*

Proof. From the quantile transform, Proposition 6.1, we know that $F^{-1}(U)$ has distribution function F if U is uniformly distributed on $(0, 1)$. Therefore, we may consider a sequence of independent random variables $U_1, U_2, \ldots$ uniformly distributed on $(0, 1)$ and represent $X_1, \ldots, X_n$ as $F^{-1}(U_1), \ldots, F^{-1}(U_n)$. Write $U_{1,n} \geq \cdots \geq U_{n,n}$ for the ordered U_k. Note that $F^{-1}_{n,X}(p) = F^{-1}(U_{[n(1-p)]+1,n})$, and since F is strictly increasing in a neighborhood of $F^{-1}(p)$, it follows that F^{-1} is continuous at p. Note also that

$$\begin{aligned}
\{u : |F^{-1}(u) - F^{-1}(p)| > \varepsilon\} &= \{u : |F^{-1}(u) - F^{-1}(p)| > \varepsilon, |u - p| \geq \delta\} \\
&\quad \cup \{u : |F^{-1}(u) - F^{-1}(p)| > \varepsilon, |u - p| < \delta\} \\
&\subset \{u : |u - p| \geq \delta\} \\
&\quad \cup \{u : |F^{-1}(u) - F^{-1}(p)| > \varepsilon, |u - p| < \delta\},
\end{aligned}$$

and the continuity at p implies that

$$\lim_{\delta\to 0} \{u : |F^{-1}(u) - F^{-1}(p)| > \varepsilon, |u - p| < \delta\}$$

is the empty set. Therefore, for all $\delta > 0$,

$$\mathrm{P}(|F^{-1}_{n,X}(p) - F^{-1}(p)| > \varepsilon) \leq \mathrm{P}(|U_{[n(1-p)]+1,n} - p| \geq \delta) + C_\delta,$$

and since $\lim_{\delta\to 0} C_\delta = 0$, to complete the proof it only remains to show that $\lim_{n\to\infty} \mathrm{P}(|U_{[n(1-p)]+1,n} - p| \geq \delta) = 0$ for every $\delta > 0$.

We claim that $U_{k,n}$ is Beta$(n - k + 1, k)$-distributed. To verify this claim, we first recall that the Beta(a, b) distribution is a probability distribution on $(0, 1)$ with density function

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1},$$

where $\Gamma(n) = (n-1)!$. The Beta(a,b) distribution has mean $a/(a+b)$ and variance $ab(a+b)^{-2}(a+b+1)^{-1}$. The density function $f_{U_{k,n}}$ of $U_{k,n}$ can be expressed as

$$f_{U_{k,n}}(x) = \frac{d}{dx} \mathrm{P}(U_{k,n} \le x) = \lim_{\Delta \to 0} \frac{\mathrm{P}(U_{k,n} \in [x, x+\Delta])}{\Delta}.$$

We want to compute the limit on the right-hand side above. To this end, we introduce the notation

$$\begin{aligned} A_x = \big\{ & n-k \text{ of the } U_j \text{ are in } (0,x) \text{ and} \\ & 1 \text{ of the } U_j \text{ is in } [x, x+\Delta] \text{ and} \\ & k-1 \text{ of the } U_j \text{ are in } (x+\Delta, 1) \big\} \end{aligned}$$

and notice that

$$\begin{aligned} \mathrm{P}(U_{k,n} \in [x, x+\Delta]) &= \mathrm{P}(A_x) + o(\Delta) \\ &= \frac{n!}{(n-k)!(k-1)!} x^{n-k} \Delta^1 (1-x-\Delta)^{k-1} + o(\Delta), \end{aligned}$$

where $o(\Delta)$ is the probability of the event $\{U_{k,n} \in [x, x+\Delta]\}$ when two or more of the U_j are in $[x, x+\Delta]$. Letting $\Delta \to 0$ gives

$$\begin{aligned} f_{U_{k,n}}(x) &= \frac{n!}{(n-k)!(k-1)!} x^{n-k}(1-x)^{k-1} \\ &= \frac{\Gamma(n+1)}{\Gamma(n-k+1)\Gamma(k)} x^{n-k+1-1}(1-x)^{k-1}, \end{aligned}$$

which confirms the claim that $U_{k,n}$ is Beta$(n-k+1, k)$-distributed. In particular,

$$\mathrm{E}[U_{k,n}] = \frac{n-k+1}{n+1} = 1 - \frac{k}{n+1}, \tag{7.5}$$

$$\mathrm{E}[U_{k,n}^2] = \frac{(n-k+1)(n-k+2)}{(n+1)(n+2)}. \tag{7.6}$$

Finally, take $p \in (0,1)$ and $k(n) = [n(1-p)] + 1$. Then from (7.5) and (7.6) we find that

$$\mathrm{E}[U_{k(n),n}] = 1 - \frac{[n(1-p)]+1}{n+1} \to 1 - (1-p) = p \quad \text{as } n \to \infty$$

and further that

$$n\,\mathrm{E}[(U_{k(n),n} - p)^2] \to p(1-p) \quad \text{as } n \to \infty. \tag{7.7}$$

The conclusion now follows from an application of Markov's inequality together with (7.7): for every $\delta > 0$

$$\mathrm{P}(|U_{k(n),n} - p\,| \geq \delta) \leq \delta^{-2}\,\mathrm{E}[(U_{k(n),n} - p)^2] \to 0 \quad \text{as } n \to \infty. \qquad \square$$

But how good is the empirical quantile as an approximation of the true quantile for finite sample sizes? It turns out that this question can be answered, at least in the sense that for a given distribution function F we can express the distribution of the empirical quantile in terms of F.

Let Y_x be the number of sample points exceeding x, i.e., the number of indices k for which $X_k > x$. It follows immediately that Y_x is $\mathrm{Bin}(n, q)$-distributed, where $q = \mathrm{P}(X_k > x) = 1 - F(x)$. We have

$$\begin{aligned}
\mathrm{P}(X_{1,n} \leq x) &= \mathrm{P}(Y_x = 0),\\
\mathrm{P}(X_{2,n} \leq x) &= \mathrm{P}(Y_x \leq 1),\\
&\vdots\\
\mathrm{P}(X_{j,n} \leq x) &= \mathrm{P}(Y_x \leq j-1).
\end{aligned}$$

Since $F^{-1}_{n,X}(p) = X_{[n(1-p)]+1,n}$, we have found that

$$\mathrm{P}(F^{-1}_{n,X}(p) \leq x) = \mathrm{P}(Y_x \leq [n(1-p)]) = \sum_{k=0}^{[n(1-p)]} \binom{n}{k} (1-F(x))^k\, F(x)^{n-k}.$$

For a given F these probabilities are easily evaluated on a computer. In particular, we can compute probabilities of the kind

$$\mathrm{P}\left(\left|\frac{F^{-1}_{n,X}(0.95) - F^{-1}(0.95)}{F^{-1}(0.95)}\right| < \varepsilon\right) = \mathrm{P}\left(1-\varepsilon < \frac{F^{-1}_{n,X}(0.95)}{F^{-1}(0.95)} < 1+\varepsilon\right),$$

i.e., the probability that the relative error is at most ε. Graphs showing these probabilities for different sample sizes and distributions can be found in Fig. 7.1.

The graph of the density function gives information about the concentration of the probability mass that is more easily interpreted than the graph of the distribution function. Differentiating the distribution function of the empirical quantile

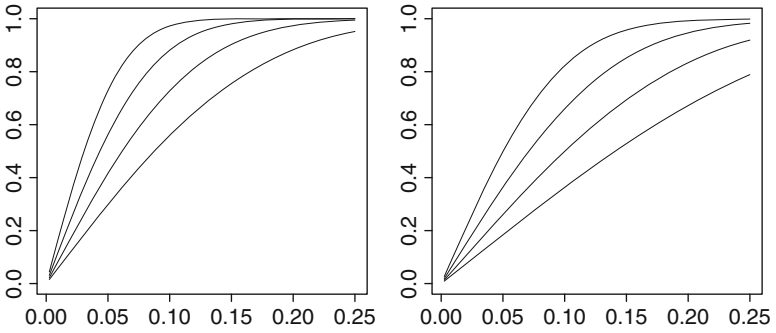

Fig. 7.1 Probabilities $\mathrm{P}(1-\varepsilon < F_{n,X}^{-1}(0.95)/F^{-1}(0.95) < 1+\varepsilon)$ for ε in $(0, 0.25)$ for sample sizes $n = 100, 200, 400, 800$ (*lower* to *upper curve*), where F is the standard normal distribution function in the *left plot* and standard lognormal in the *right plot*

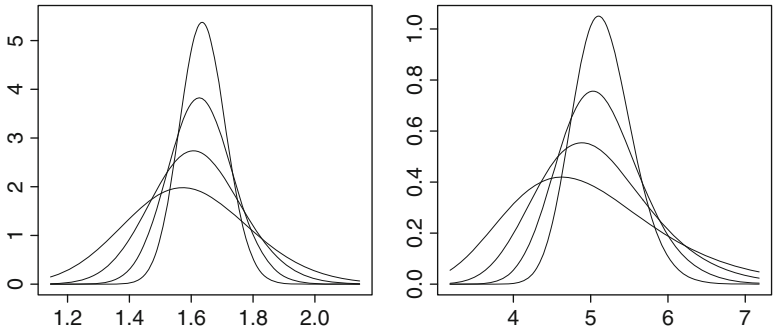

Fig. 7.2 Density functions of empirical quantile estimators $F_{n,X}^{-1}(0.95)$ for sample sizes $n = 100, 200, 400, 800$, where F is the standard normal distribution function in the *left plot* and standard lognormal in the *right plot*

$$\mathrm{P}(F_{n,X}^{-1}(p) \le x) = \sum_{k=0}^{[n(1-p)]} \binom{n}{k} (1-F(x))^k F(x)^{n-k},$$

i.e., the right-hand expression above, gives the density function of the empirical quantile (assuming that F has a density f). It is given by

$$\frac{n!}{[n(1-p)]!(n-[n(1-p)]-1)!}(1-F(x))^{[n(1-p)]}F(x)^{n-[n(1-p)]-1}f(x). \quad (7.8)$$

The graphs of density function (7.8) of the empirical quantile for different sample sizes and distributions are shown in Fig. 7.2.

Now we consider another approach to investigating the accuracy of the empirical quantile estimator $F_{n,X}^{-1}(p) = X_{[n(1-p)]+1,n}$ based on the sample $\{X_1, \ldots, X_n\}$ of independent copies of X with distribution function F. The approach considered

here is appropriate for rather large sample sizes n. Here we want to determine the sample size n required for bounding the root mean square error (RMSE) of the empirical quantile by $10^{-d} F^{-1}(p)$ for some $d \geq 1$. More precisely, we want to determine the smallest integer n such that

$$\mathrm{E}[(X_{[n(1-p)]+1,n} - F^{-1}(p))^2]^{1/2} \leq 10^{-d} F^{-1}(p). \tag{7.9}$$

From Proposition 6.1 we know that we can represent $X_1, \ldots, X_n$ in terms of F and independent random variables $U_1, \ldots, U_n$ uniformly distributed on $(0, 1)$ as $F^{-1}(U_1), \ldots, F^{-1}(U_n)$. In particular, $X_{[n(1-p)]+1,n} = F^{-1}(U_{[n(1-p)]+1,n})$.

Under the assumption that F is differentiable with a density function f we may use a Taylor expansion of F^{-1} around point p to approximate

$$\begin{aligned} X_{[n(1-p]+1,n} &= F^{-1}(U_{[n(1-p]+1,n)} \\ &= F^{-1}(p) + \frac{d}{dp} F^{-1}(p)(U_{[n(1-p]+1,n} - p) + \text{remainder term} \\ &= F^{-1}(p) + \frac{1}{f(F^{-1}(p))}(U_{[n(1-p]+1,n} - p) + \text{remainder term}. \end{aligned}$$

If we ignore the remainder term and use (7.6), then we may approximate the mean square error of the empirical quantile estimator by

$$\begin{aligned} \mathrm{E}\left[\left(X_{[n(1-p)]+1,n} - F^{-1}(p)\right)^2\right] &\approx \frac{1}{f(F^{-1}(p))^2} \mathrm{E}\left[\left(U_{[n(1-p)]+1,n} - p\right)^2\right] \\ &= \frac{1}{f(F^{-1}(p))^2} \frac{p(1-p)}{n}. \end{aligned} \tag{7.10}$$

Can we ignore the error term? By Taylor's formula we find that if f does not vary that much in a neighborhood of $F^{-1}(p)$, then, as $n \to \infty$,

$$\mathrm{E}\left[(X_{[n(1-p)]+1,n} - F^{-1}(p))^2\right] \frac{1}{f(F^{-1}(p))^2} \frac{p(1-p)}{n} + O(n^{-2}).$$

The notation $O(n^{-2})$ as $n \to \infty$ means that the remainder term divided by n^{-2} is bounded as $n \to \infty$.

If we want the bounded RMSE in (7.9) and approximate the mean square error by (7.10), then we find that the sample size n should be

$$n \approx \frac{p(1-p)}{f(F^{-1}(p))^2 F^{-1}(p)^2} 10^{2d}.$$

We illustrate the size of n for the exponential and the Pareto distribution. The following example shows that large sample sizes are required to obtain small relative errors for the empirical estimator of high quantiles.

Example 7.6 (Bound on the RMSE). Consider an exponential distribution with $F(x) = 1 - e^{-x}$ for $x > 0$. Then $F^{-1}(p) = -\log(1-p)$ and $f(F^{-1}(p)) = \exp\{\log(1-p)\} = 1-p$. If $p = 0.99$ and $d = 2$ (RMSE of 1% of $F^{-1}(p)$), then we need a sample size n such that

$$n \approx \frac{p}{(1-p)[\log(1-p)]^2} 10^4 \approx 4.7 \cdot 10^4.$$

Consider a Pareto distribution with $F(x) = 1 - x^{-3}$ for $x > 1$. Then $F^{-1}(p) = (1-p)^{-1/3}$ and $f(F^{-1}(p)) = 3(1-p)^{4/3}$. If $p = 0.99$ and $d = 2$ (a RMSE of 1% of $F^{-1}(p)$), then we need a sample size n such that

$$n \approx \frac{p}{9(1-p)} 10^4 = 1.1 \cdot 10^5.$$

7.4 Empirical VaR and ES

If X is the value at time 1 of a financial portfolio, then $\mathrm{VaR}_p(X) = F_L^{-1}(1-p)$, where $L = -X/R_0$, where R_0 is the return on the reference instrument (a risk-free zero-coupon bond, say). Given a sample $L_1, \dots, L_n$ of independent copies of L, the empirical estimate of $\mathrm{VaR}_p(X)$ is therefore given by

$$\widehat{\mathrm{VaR}}_p(X) = L_{[np]+1,n},$$

where $L_{1,n} \geq \cdots \geq L_{n,n}$ is the ordered sample. Note that $\widehat{\mathrm{VaR}}_p(X)$ is simply the empirical $(1-p)$-quantile of L_k.

To compute the empirical VaR estimate from a sample of historical prices, we first transform the prices into a sample $\{L_1, \dots, L_n\}$ and then compute the VaR estimate as an empirical quantile. The following example illustrates this procedure.

Example 7.7 (Empirical VaR*).* Consider a portfolio consisting of long positions in two different assets, one unit of the first asset and two units of the second asset. Historical daily prices per unit of the two assets over the last 20 days are given by S_t^1 and S_t^2 for $t = -20, \dots, 0$. Assume the corresponding pairs of returns

$$\mathbf{R}_t = (R_t^1, R_t^2) = (S_t^1/S_{t-1}^1, S_t^2/S_{t-1}^2), \quad t = -19, \dots, 0,$$

are independent and identically distributed. The value of the portfolio at time 0 is given by $V_0 = S_0^1 + 2S_0^2$, and the value at time 1 is $V_1 = S_1^1 + 2S_1^2$. Suppose we want to compute the empirical $\mathrm{VaR}_{0.05}$ estimate of $X = V_1 - V_0 R_0$, where for

simplicity we set $R_0 = 1$. We may express X as the value of a function evaluated at the point $\mathbf{R}_1$:

$$\begin{aligned} X = V_1 - V_0 &= (S_1^1 - S_0^1) + 2(S_1^2 - S_0^2) \\ &= S_0^1(R_1^1 - 1) + 2S_0^2(R_1^2 - 1) = g(\mathbf{R}_1). \end{aligned}$$

Under the assumption that the $\mathbf{R}_t$, for $t = -19, \ldots, 0$, are independent copies of $\mathbf{R}_1$, we can easily construct independent copies of X by setting $X_k = g(\mathbf{R}_{-20+k})$ for $k = 1, \ldots, 20$. Setting $L_k = -X_k$ and ordering the sample of L_k as $L_{1,20} \geq \cdots \geq L_{20,20}$, we compute the empirical estimate of $\mathrm{VaR}_{0.05}(X)$ as $\widehat{\mathrm{VaR}}_{0.05}(X) = L_{[20\cdot 0.05]+1,20} = L_{2,20}$.

Example 7.8 (Thinning versus historical simulation). Suppose today is November 3, 2010, and we have just invested an amount of 100 in the Dow Jones Industrial Average (DJIA) stock market index. We want to analyze the risk we face from holding the position over a period of 20 trading days. The value of the position 20 trading days from today is $V_{20} = 100R_1 \cdots R_{20}$, where $R_1, \ldots, R_{20}$ are the daily returns over the time period under consideration. We want to estimate $\mathrm{VaR}_p(V_{20} - 100)$ (the effect of interest rates are ignored) based on the 801 historical index values of DJIA from August 30, 2007 to November 2, 2010.

If the thinning approach in Example 7.2 is used, then we use every 20th value of the sample of historical DJIA values to obtain the sample $\{R_{-39}^{(20)}, \ldots, R_0^{(20)}\}$ of historical 20-day returns. We set

$$X_k = 100(R_{-40+k}^{(20)} - 1) \quad \text{and} \quad L_k = -X_k \quad \text{for } k = 1, \ldots, 40$$

and estimate $\mathrm{VaR}_p(V_{20} - 100)$ as $L_{[40p]+1,40}$.

If the historical simulation approach in Example 7.3 is used, then we may choose $m = 5{,}000$ in Example 7.3 and use the sample of historical DJIA values to obtain the sample $\{R_1^{*(20)}, \ldots, R_{5000}^{*(20)}\}$ of fictive 20-day returns. We set

$$X_k = 100(R_k^{*(20)} - 1) \quad \text{and} \quad L_k = -X_k \quad \text{for } k = 1, \ldots, 5000$$

and estimate $\mathrm{VaR}_p(V_{20} - 100)$ as $L_{[5000p]+1,5000}$.

The left plot in Fig. 7.3 shows the left tail of the empirical distribution function of X_k for the two approaches. Because the thinning approach gives a sample of small size 40, the staircase shape is pronounced and the tail estimate is unreliable. The historical simulation approach gives a much smoother, and likely more reliable, estimate of the left tail of the distribution function of $V_{20} - 100$. The right plot in Fig. 7.3 shows the estimates of $\mathrm{VaR}_p(V_{20} - 100)$ as a function of p for the two approaches.

We now present the empirical ES estimator. Recall that the ES at level p of a portfolio with value X at time 1 is given by

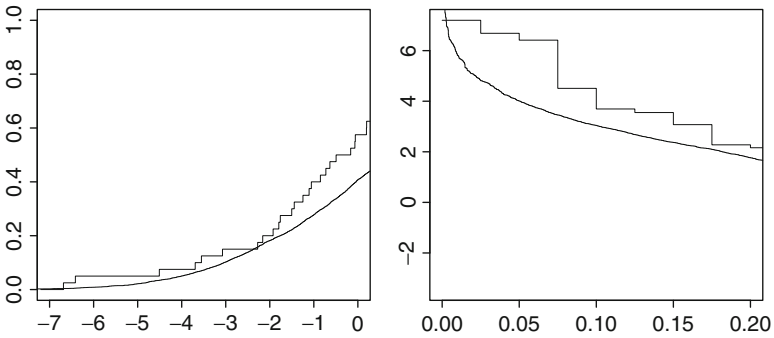

Fig. 7.3 *Above*: observed upper triangle of paid claims; *Below*: unobserved triangle of outstanding claims

$$\mathrm{ES}_p(X) = \frac{1}{p}\int_0^p \mathrm{VaR}_u(X)du.$$

The empirical ES estimator is obtained by simply replacing $\mathrm{VaR}_p(X)$ by its empirical estimator $\widehat{\mathrm{VaR}}_p(X) = L_{[np]+1,n}$, where $L_k = -X_k/R_0$ and $L_{1,n} \geq \cdots \geq L_{n,n}$ is the ordered loss sample. This implies

$$\begin{aligned}\widehat{\mathrm{ES}}_p(X) &= \frac{1}{p}\int_0^p L_{[nu]+1,n}du \\ &= \frac{1}{p}\left(\sum_{k=1}^{[np]}\frac{L_{k,n}}{n} + \left(p - \frac{[np]}{n}\right)L_{[np]+1,n}\right). \end{aligned} \tag{7.11}$$

If $[np]$ is an integer, then the expression in the last display reduces to the sample mean of the $[np]$ largest losses. To clarify how one arrives at the expression in (7.11), suppose that $[np] \geq 2$, and notice that in this case

$$\begin{aligned}&\int_0^p L_{[nu]+1,n}du \\ &= \int_0^{1/n} L_{1,n}du + \cdots + \int_{([np]-1)/n}^{[np]/n} L_{[np],n}du + \int_{[np]/n}^p L_{[np]+1,n}du \\ &= \frac{1}{n}L_{1,n} + \cdots + \frac{1}{n}L_{[np],n} + \left(p - \frac{[np]}{n}\right)L_{[np]+1,n}.\end{aligned}$$

Example 7.9 (Empirical ES*).* Consider the historical daily prices of two assets A and B, listed in Table 7.1. You are considering taking a position corresponding to a long position of two units of asset A and three units of asset B.

Table 7.1 Historical daily prices of two assets A and B

Day	−20	−19	−18	−17	−16	−15	−14
Asset A	81.75	81.35	80.4	81.05	83.35	83.00	83.30
Asset B	81.25	81.00	81.5	81.50	81.85	81.25	81.45
Day	−13	−12	−11	−10	−9	−8	−7
Asset A	86.0	85.5	84.50	84.00	84.05	82.35	83.45
Asset B	83.5	83.5	83.75	86.00	85.75	84.60	83.85
Day	−6	−5	−4	−3	−2	−1	0
Asset A	83.50	84.4	86.9	85.90	82.55	83.75	84.75
Asset B	84.55	84.0	84.3	84.75	85.35	87.00	85.75

Table 7.2 Sample of X_k values and corresponding ordered L_k values

Sample of X_k-values: transformation of historical prices									
−1.62	−0.39	1.37	5.91	−2.60	1.25	11.97	−0.99	−1.21	5.91
−0.65	−6.88	−0.02	2.25	0.15	5.94	−0.58	−4.79	7.44	−1.67
Ordered sample of corresponding L_k values									
6.88	4.79	2.60	1.67	1.62	1.21	0.99	0.65	0.58	0.39
0.02	−0.15	−1.25	−1.37	−2.25	−5.91	−5.91	−5.94	−7.44	−11.97

To evaluate the riskiness of this investment, you want to compute the empirical ES estimate $\mathrm{ES}_p(X)$, where $p = 0.06$ and X is the difference between the value of the position tomorrow and its current value. We may express X as a function of the vector $\mathbf{R}_1 = (R_1^{\mathrm{A}}, R_1^{\mathrm{B}})$ of returns over the next day as

$$X = V_1 - V_0 = 2S_0^{\mathrm{A}}(R_1^{\mathrm{A}} - 1) + 3S_0^{\mathrm{B}}(R_1^{\mathrm{B}} - 1) = f(\mathbf{R}_1).$$

From Table 7.1 we can compute the corresponding vectors of historical returns. The function f then transforms these vectors into the sample of X_k values shown in Table 7.2 (rounded off to two decimal points). Setting $L_k = -X_k$ and ordering the L_k gives the ordered sample of L_k values in Table 7.2. From (7.11) we find that the ES estimate based on the values $l_{1,n} \geq \cdots \geq l_{n,n}$ is

$$\frac{1}{p}\left(\sum_{k=1}^{[np]} \frac{l_{k,n}}{n} + \left(p - \frac{[np]}{n}\right) l_{[np]+1,n}\right).$$

Here, with $n = 20$ and $p = 0.06$ and the values $l_{k,n}$ in Table 7.2 we get

$$\frac{1}{0.06}\left(\frac{6.88}{20} + (0.06 - 0.05)4.79\right) \approx 6.53.$$

7.5 Confidence Intervals

Suppose we have observations $x_1, \ldots, x_n$ of independent and identically distributed random variables $X_1, \ldots, X_n$ from an unknown distribution function F and we want to know the value $\theta = \theta(F)$ of some quantity that is determined by the unknown F. Examples include the mean, the variance, some quantile, or some risk measure that depends on F. We may estimate θ by the empirical estimator $\widehat{\theta} = \theta(F_{n,X})$ obtained by computing θ from $F_{n,X}$ instead of F. However, a point estimate is not meaningful unless we have some way of assessing its accuracy. Since we can never know whether the observations $x_1, \ldots, x_n$ are representative outcomes from the unknown distribution F, we can never know whether the empirical estimate $\widehat{\theta}_{\text{obs}} = \theta(F_n)$ based on these observations is close to the true value θ. What we can do is compute a confidence interval for θ.

Let us first recall what a confidence interval is. Given $q \in (0, 1)$, we want to form a stochastic interval (A, B), where $A = f_A(X_1, \ldots, X_n)$ and $B = f_B(X_1, \ldots, X_n)$ for some functions f_A and f_B such that

$$\mathrm{P}(A < \theta < B) = q,$$

i.e., the stochastic interval (A, B) covers the value θ with probability q. Clearly, we want q to be close to 1, e.g., $q = 0.95$, and at the same time we want that the length of the interval is likely to be small. The interval (a, b), where $a = f_A(x_1, \ldots, x_n)$ and $b = f_B(x_1, \ldots, x_n)$, is a confidence interval for θ with confidence level q. We may say that we feel confident at level q that the interval (a, b) covers the value θ. Note that q is not the probability that the specific interval (a, b) covers θ (either it does or it does not), but q is the probability that the procedure generating the interval will produce an interval covering θ if fed with a new random sample of the same size from the same probability distribution. Often we want to find a double-sided interval so that

$$\mathrm{P}(A < \theta < B) = q, \quad \mathrm{P}(A \geq \theta) = \mathrm{P}(B \leq \theta) = (1 - q)/2.$$

Since F is unknown, we do not know the functions f_A, f_B, but we can construct approximate confidence intervals. If θ is a quantile of F, i.e., $\theta = F^{-1}(p)$, then we can actually find exact confidence intervals for θ, but not for arbitrary confidence levels q.

7.5.1 Exact Confidence Intervals for Quantiles

Suppose we have observations $x_1, \ldots, x_n$ of outcomes of independent and identically distributed random variables $X_1, \ldots, X_n$ with common unknown continuous distribution function F. Suppose further that we want to construct a confidence

interval (a, b) for the quantile $F^{-1}(p)$, where $a = f_A(x_1, \ldots, x_n)$ and $b = f_B(x_1, \ldots, x_n)$ such that

$$\mathrm{P}(A < F^{-1}(p) < B) = q, \quad \mathrm{P}(A \geq F^{-1}(p)) = \mathrm{P}(B \leq F^{-1}(p)) = (1-q)/2,$$

where q is a confidence level, $A = f_A(X_1, \ldots, X_n)$, and $B = f_B(X_1, \ldots, X_n)$. Since F is unknown, we cannot find a and b. However, we can look for $i > j$ and the smallest $q' \geq q$ such that

$$\begin{aligned} &\mathrm{P}(X_{i,n} < F^{-1}(p) < X_{j,n}) = q', \\ &\mathrm{P}(X_{i,n} \geq F^{-1}(p)) \leq (1-q)/2, \quad \mathrm{P}(X_{j,n} \leq F^{-1}(p)) \leq (1-q)/2. \end{aligned} \tag{7.12}$$

It remains to compute the probabilities in (7.12). Let $Y_{F^{-1}(p)}$ be the number of sample points exceeding $F^{-1}(p)$, i.e., the number of indices k for which $X_k > F^{-1}(p)$. It follows immediately that $Y_{F^{-1}(p)}$ is $\mathrm{Bin}(n, r)$-distributed, where $r = \mathrm{P}(X_k > F^{-1}(p)) = 1 - F(F^{-1}(p))$. From Proposition 6.1 we know that the continuity of F implies that $F(F^{-1}(u)) = u$ for all $u \in (0, 1)$. In particular, $Y_{F^{-1}(p)}$ is $\mathrm{Bin}(n, 1-p)$-distributed. The probabilities in (7.12) are easily expressed in terms of the probabilities of $Y_{F^{-1}(p)}$, which are very easily computed (with the assistance of some appropriate software). We have

$$\begin{aligned} \mathrm{P}(X_{1,n} \leq F^{-1}(p)) &= \mathrm{P}(Y_{F^{-1}(p)} = 0), \\ \mathrm{P}(X_{2,n} \leq F^{-1}(p)) &= \mathrm{P}(Y_{F^{-1}(p)} \leq 1), \\ &\vdots \\ \mathrm{P}(X_{j,n} \leq F^{-1}(p)) &= \mathrm{P}(Y_{F^{-1}(p)} \leq j-1). \end{aligned}$$

Similarly, $\mathrm{P}(X_{i,n} \geq F^{-1}(p)) = 1 - \mathrm{P}(Y_{F^{-1}(p)} \leq i-1)$. We may now can compute the probabilities $\mathrm{P}(X_{j,n} \leq F^{-1}(p))$ and $\mathrm{P}(X_{i,n} \geq F^{-1}(p))$ for different i and j until we find indices that satisfy (7.12).

Example 7.10 (Exact intervals for quantiles). Suppose we have a sample $\{X_1, \ldots, X_{200}\}$ of independent and identically distributed random variables with common unknown continuous distribution function F and we want a confidence interval for $F^{-1}(0.95)$ with confidence level $q' \approx q = 0.95$. Since $Y_{F^{-1}(0.95)}$ is $\mathrm{Bin}(200, 0.05)$-distributed, we find that

$$\begin{aligned} \mathrm{P}(X_{5,200} \leq F^{-1}(0.95)) &= \mathrm{P}(Y_{F^{-1}(0.95)} \leq 4) \approx 0.0264, \\ \mathrm{P}(X_{17,200} \geq F^{-1}(0.95)) &= 1 - \mathrm{P}(Y_{F^{-1}(0.95)} \leq 16) \approx 0.0238. \end{aligned}$$

Therefore, $\mathrm{P}(X_{17,200} < F^{-1}(0.95) < X_{5,200}) \approx 0.95$, so $(x_{17,200}, x_{5,200})$ is a confidence interval for $F^{-1}(0.95)$ with a confidence level of approximately 95%. The length of the confidence interval depends on the sample points, which in turn depends on the unknown distribution function F. Figure 7.4 shows 100 outcomes

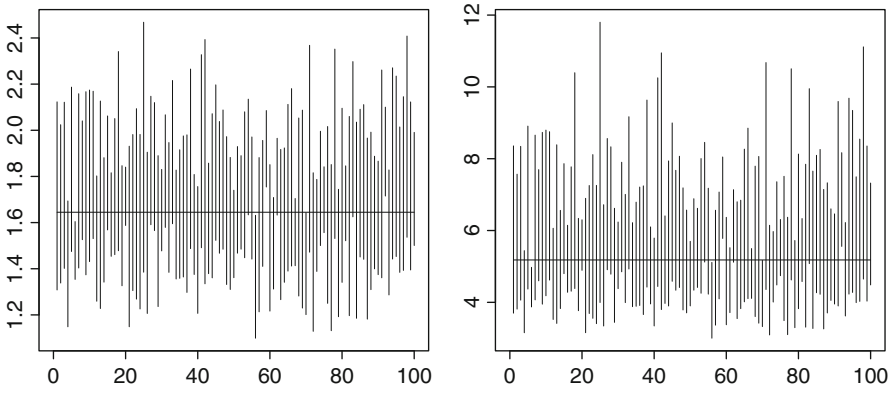

Fig. 7.4 *Each plot* shows empirical confidence intervals $(x_{17,200}, x_{5,200})$ for $F^{-1}(0.95)$ with confidence level 95% for 100 samples of size 200. *Left plot*: empirical confidence intervals for F the standard normal distribution; *right plot*: for F the standard lognormal distribution function

$(x_{17,200}, x_{5,200})$ of the empirical confidence interval for $F^{-1}(0.95)$ for F standard normal (left plot) and for F standard lognormal (right plot). Notice that the 25th confidence interval for the lognormal F says that if we had the 25th lognormal sample, then we could only say that we are rather sure that the 95% quantile of the unknown distribution lies somewhere between 4 and 11.8. This illustrates the difficulty of accurately estimating quantile values.

7.5.2 Confidence Intervals Using the Nonparametric Bootstrap

For quantiles we have seen how to construct exact confidence intervals. However, for risk measures, which unlike VaR are not simply quantile values, and for other quantities such as moments and loss probabilities this approach does not work. We will here investigate a useful method for constructing approximate confidence intervals called the nonparametric bootstrap method.

Suppose we have observations $x_1, \ldots, x_n$ of independent and identically distributed random variables $X_1, \ldots, X_n$ and we want to estimate some quantity $\theta = \theta(F)$ that depends on the unknown distribution F of X_k. For instance, θ could be the p-quantile $\theta = F^{-1}(p)$ giving $\widehat{\theta}_{\text{obs}} = x_{[n(1-p)]+1,n}$ or the mean $\theta = \int x dF(x)$ giving $\widehat{\theta}_{\text{obs}} = (x_1 + \cdots + x_n)/n$. We want to construct a confidence interval for θ with confidence level q.

If F were known, we could compute the value θ analytically or approximate it numerically. Alternatively, we could simulate a large sample from F to approximately compute θ as the empirical estimate. The problem here is that we do not know F and we only have one sample $\{x_1, \ldots, x_n\}$ of size n from F.

One way to produce more samples is to randomly draw with replacement n times from the set of observations $x_1, \ldots, x_n$ to produce a sample $\{X_1^*, \ldots, X_n^*\}$. The sample points X_k^* are independent and F_n-distributed (uniformly distributed on the set of the original observations $x_1, \ldots, x_n$). Some of the X_k^* are likely to be equal, even if the x_k are all different. The probability that $X_j^* \neq X_k^*$ for all $j \neq k$ is very small; the probability that none of the x_ks is drawn twice among the n tries is $n!/n^n$. Write F_n^* for the empirical distribution of $X_1^*, \ldots, X_n^*$ and $\widehat{\theta}^* = \theta(F_n^*)$ for the estimate of θ based on the sample $\{X_1^*, \ldots, X_n^*\}$. Even though $\{X_1^*, \ldots, X_n^*\}$ is not a sample from F, it has most of the characteristics of a sample from F as long as n is sufficiently large. In particular, the probability distribution of $\widehat{\theta}^*$ is likely to be close to the probability distribution of $\widehat{\theta}$. Whereas the probability distribution of $\widehat{\theta}$ is unknown (since F is not known), the probability distribution of $\widehat{\theta}^*$ can be approximated arbitrarily well by repeated resampling N times for N large enough.

An approximative confidence interval $I_{\theta,q}$ for θ of confidence level q using the nonparametric bootstrap method is constructed as follows.

- For each j in the set $\{1, \ldots, N\}$ draw with replacement n times from the sample $\{x_1, \ldots, x_n\}$ to obtain the sample $\{X_1^{*(j)}, \ldots, X_n^{*(j)}\}$ and the corresponding empirical distribution function $F_n^{*(j)}$.
- Compute the estimates $\widehat{\theta}_j^* = \theta(F_n^{*(j)})$ of θ and the residuals $R_j^* = \widehat{\theta}_{\text{obs}} - \widehat{\theta}_j^*$ for $j = 1, \ldots, N$.
- Form the interval

$$I_{\theta,q} = (\widehat{\theta}_{\text{obs}} + R^*_{[N(1+q)/2]+1,N}, \widehat{\theta}_{\text{obs}} + R^*_{[N(1-q)/2]+1,N}),$$

where $R^*_{1,N} \geq \cdots \geq R^*_{N,N}$ is the ordering of the sample $\{R_1^*, \ldots, R_N^*\}$.

Why is the interval $I_{\theta,q}$ a reasonable approximative confidence interval for θ? Here is one way of motivating the procedure.

Let G denote the distribution function of $\theta - \widehat{\theta}$. Then

$$\begin{aligned} q &= \mathrm{P}(G^{-1}((1-q)/2) < \theta - \widehat{\theta} < G^{-1}((1+q)/2)) \\ &= \mathrm{P}(\widehat{\theta} + G^{-1}((1-q)/2) < \theta < \widehat{\theta} + G^{-1}((1+q)/2)). \end{aligned}$$

Therefore, $(\widehat{\theta}_{\text{obs}} + G^{-1}((1-q)/2), \widehat{\theta}_{\text{obs}} + G^{-1}((1+q)/2))$ is a confidence interval for θ of level q. The problem is that we do not know the distribution function G.

The success of the bootstrap relies on the validity of the bootstrap principle, which says that G can be well approximated by the empirical distribution G_N^* of $R_1^*, \ldots, R_N^*$. Then the quantiles $G^{-1}((1-q)/2)$ and $G^{-1}((1+q)/2)$ can be well approximated by the empirical counterparts $R^*_{[N((1+q)/2)]+1,N}$ and $R^*_{[N(1-q)/2]+1,N}$, which leads to the interval $I_{\theta,q}$. We need n to be sufficiently large to make it plausible that the bootstrap principle holds so $\theta - \widehat{\theta}$ and $\widehat{\theta}_{\text{obs}} - \widehat{\theta}^*$ are approximately equally distributed. This requirement is investigated in the following example.

Example 7.11 (Bootstrap intervals for quantiles). Suppose we want to construct confidence intervals for $\theta = \mathrm{VaR}_{0.05}(V_1 - V_0)$, where V_0 and V_1 are respectively the current value and the value tomorrow of a long position in some stock index. Since the time period here is only 1 day, we ignore the impact of interest rates. We may express V_1 in terms of the return R_1 of the stock index as $V_1 = V_0 R_1$, and we assume that $\log R_1$ is normally distributed with zero mean and standard deviation 0.01. For simplicity we also assume that $V_0 = 1$. This implies that

$$\mathrm{VaR}_{0.05}(V_1 - V_0) = F^{-1}_{V_0 - V_1}(0.95) = V_0\big(1 - F^{-1}_{R_1}(0.05)\big) \approx 0.016314.$$

In reality we would not know with certainty the return distribution or, therefore, the value of $\mathrm{VaR}_{0.05}(V_1 - V_0)$. However, we may—under the right circumstances—believe that the past n index returns can be seen as sample points from the distribution of the future return R_1 and in that case transform the historical returns into outcomes $l_1, \ldots, l_n$ of $L_1, \ldots, L_n$ that are independent copies of $L = V_0 - V_1$. The problem we investigate here is how to construct and evaluate confidence intervals for $F_L^{-1}(0.95)$ given the sample $\{l_1, \ldots, l_n\}$. We have already seen how we can construct confidence intervals for quantiles, and this approach is applicable here since $\mathrm{VaR}_{0.05}(V_1 - V_0) = F_L^{-1}(0.95)$. However, the aim here is to investigate the nonparametric bootstrap approach to construct approximative confidence intervals and evaluate it by comparing it to the approach for quantiles.

Recall that the accuracy of the nonparametric bootstrap approach for constructing confidence intervals is likely to be good if $\theta - \widehat{\theta}$, where $\widehat{\theta} = L_{[0.05n]+1,n}$, and $\widehat{\theta}_{\mathrm{obs}} - \widehat{\theta}^*$ have approximately the same probability distribution. The upper left plot in Fig. 7.5 shows a histogram from 2,000 simulations of $\theta - \widehat{\theta}$. The upper right and middle plots in Fig. 7.5 show histograms of 2,000 bootstrap simulations of $\widehat{\theta}_{\mathrm{obs}} - \widehat{\theta}^*$ based on resampling from three different outcomes of $L_1, \ldots, L_{500}$. Based on these plots, we definitely see a resemblance between the distribution of $\theta - \widehat{\theta}$ and $\widehat{\theta}_{\mathrm{obs}} - \widehat{\theta}^*$, but it is clear that much information has been lost in the bootstrap world. One might suspect that increasing the number N of resampling runs could improve things. The middle left and lower left plots show bootstrap simulations of $\widehat{\theta}_{\mathrm{obs}} - \widehat{\theta}^*$ based on the same sample $\{l_1, \ldots, l_{500}\}$, where the number N is 2,000 for the middle left plot and 10,000 for the lower left plot. The same comparison is shown in the middle right and the lower right plots but based on another sample $\{l_1, \ldots, l_{500}\}$. We observe that increasing the number N of bootstrap simulations to a very large number does not improve things much.

Finally, we compute 50 confidence intervals for θ of confidence level 0.95 with the exact method (left plot in Fig. 7.6) and with the nonparametric bootstrap method (right plot in Fig. 7.6). We observe that the results are very similar. The differences among the confidence intervals are to a much greater extent due to the differences among the 50 outcomes of the random sample $\{L_1, \ldots, L_{500}\}$ than to the particular method used. We conclude that here the bootstrap method performs rather well.

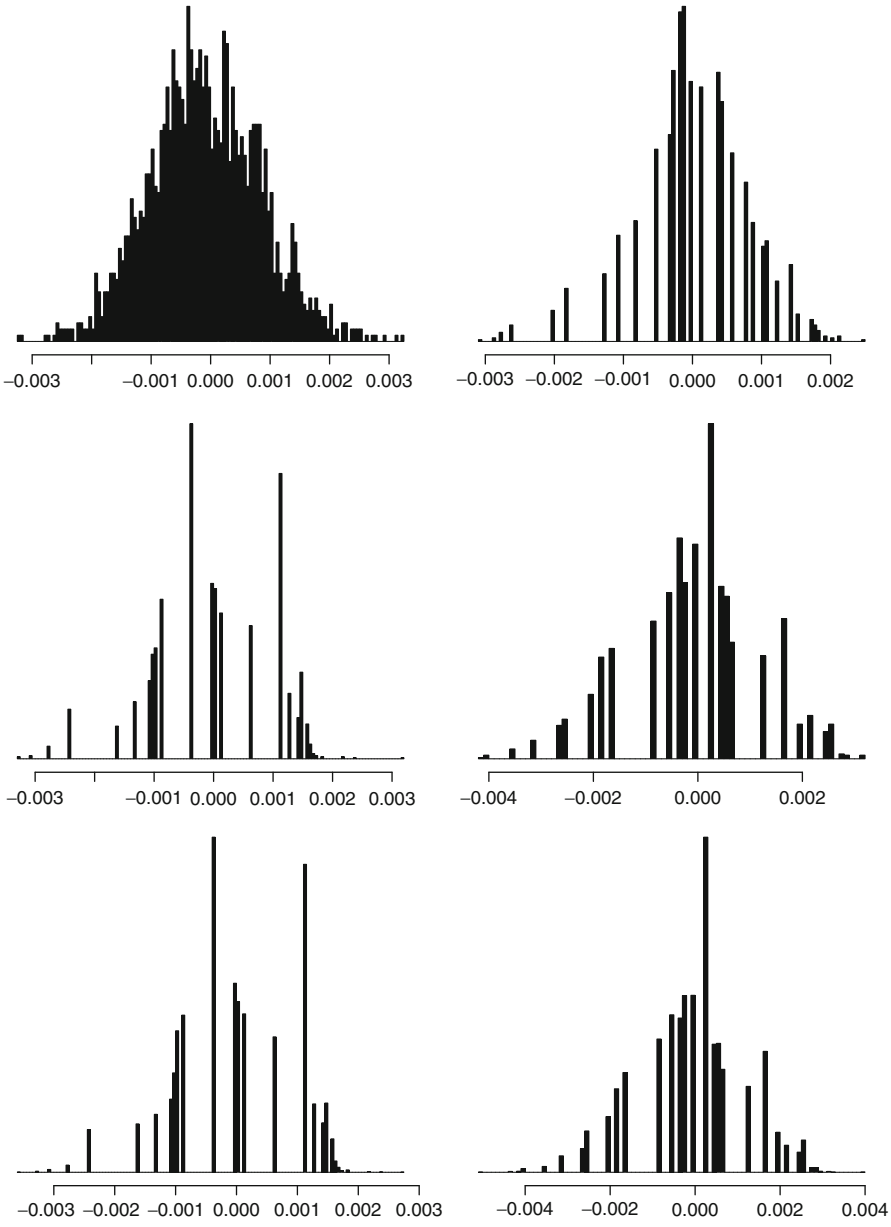

Fig. 7.5 *Upper left plot*: histogram of 2,000 outcomes of $\theta - \hat{\theta}$ based on 2,000 outcomes of $\{L_1, \ldots, L_{500}\}$. Each of the remaining plots shows centered bootstrap estimates $\hat{\theta}_{\text{obs}} - \hat{\theta}^*$. *Upper right* and *middle plots*: histograms based on $N = 2{,}000$ resampling runs for three different outcomes of the sample $\{L_1, \ldots, L_{500}\}$. *Middle* and *lower left plots*: based on the same original sample, the number of resampling runs is 10,000 for the *lower left plot* instead of 2,000. Similarly for *middle* and *lower right plots*

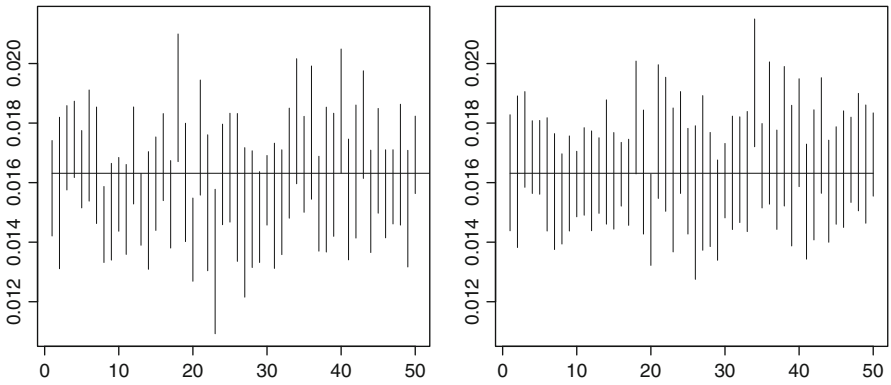

Fig. 7.6 *Each plot* shows 50 confidence intervals for $\mathrm{VaR}_{0.05}(V_1 - V_0)$ based on simulated samples of size 500. *Left plot*: result for the exact method for quantiles; *right plot*: result for nonparametric bootstrap method

7.6 Bootstrapping in Nonlife Insurance

This section is devoted to an application of the nonparametric bootstrap in an insurance context. Consider a nonlife insurer who is about to quantify risk with a 1-year horizon. The risk is quantified in terms of a solvency capital requirement (SCR). The SCR is, as in Example 6.1, given by

$$\mathrm{SCR} = \rho(A_1 - A_0 R_0 - L_1 + L_0 R_0),$$

where time is measured in years, A_0 and A_1 are the values of the assets at times 0 and 1, L_0 and L_1 are the values of the liabilities at times 0 and 1, and ρ is a risk measure, e.g., VaR or ES. To compute the SCR, the insurer must determine the distribution of $A_1 - L_1$—the value of the assets minus the value of the liabilities 1 year from now.

To compute L_0, the current value of the liabilities, the insurer adopts a claim reserving technique called the chain ladder, which will be explained below. L_0 consists of two parts: the value of the outstanding payments of incurred but not yet settled claims and the total value of the payments due to claims that will occur over the next year. The current value of the liabilities is computed as the sum of the discounted predicted future payment amounts.

7.6.1 Claims Reserve Prediction Via the Chain Ladder

The prediction of the future payment amounts is based on a historical record of paid claims. It is assumed that all claims that occurred at least $n + 1$ years ago

Table 7.3 *Above*: observed upper triangle of paid claims; *below*: unobserved triangle of outstanding claims

	Development year					
Origin	0	1	2	$\cdots$	$n-1$	n
$-n-1$	$C_{-n-1,0}$	$C_{-n-1,1}$	$C_{-n-1,2}$	$\cdots$	$C_{-n-1,n-1}$	$C_{-n-1,n}$
$-n$	$C_{-n,0}$	$C_{-n,1}$	$C_{-n,2}$	$\cdots$	$C_{-n,n-1}$	
$\vdots$	$\vdots$	$\vdots$				
-2	$C_{-2,0}$	$C_{-2,1}$				
-1	$C_{-1,0}$					
0						

	Development year					
Origin	0	1	2	$\cdots$	$n-1$	n
$-n-1$						
$-n$						$C_{-n,n}$
$-n+1$					$C_{-n+1,n-1}$	$C_{-n+1,n}$
$\vdots$					$\vdots$	$\vdots$
-1		$C_{-1,1}$	$C_{-1,2}$	$\cdots$	$C_{-1,n-1}$	$C_{-1,n}$
0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$	$\cdots$	$C_{0,n-1}$	$C_{0,n}$

are completely settled, but all claims that occurred at most n years ago are not completely settled. The historical record of paid claims is displayed in the form of a claims triangle. The triangle of paid claims, called the upper triangle, displays on each row the amounts of paid claims for claims incurred with the same origin year. The columns represent the development years—the difference between the year a claim was settled and the year it was incurred. The entry $C_{-k,l}$ represents the amount paid for claims that were incurred in year $-k$ and paid l year later, in year $-k+l$, for $l = 0, \ldots, k-1$. The upper triangle is illustrated in the top half of Table 7.3.

The insurer relies on the assumption that the payment pattern over the development years is repetitive. Even if the number of accidents and claim amounts may differ significantly from year to year, the payment patterns over the development years look similar. Based on this assumption the upper triangle of paid claims can be used to predict unobserved future payments. The unknown future payments are represented in the lower triangle of outstanding claims, with entries $C_{-k,l}$ for $l = k, \ldots, n$. The lower triangle is illustrated in the bottom half of Table 7.3.

To formulate the idea that the payment patterns are repetitive, one possibility, which is the one we follow here, is to assume a multiplicative structure for the cumulative claims. Consider, for $k = 1, \ldots, n+1$ and $l = 0, \ldots, k-1$, the cumulative amounts paid for claims that occurred in year $-k$,

$$D_{-k,l} = \sum_{j=0}^{l} C_{-k,j}\,.$$

Suppose the expected cumulative payments can be written as

$$\mathrm{E}[D_{-k,l+1}] = \mathrm{E}[D_{-k,l}]\, f_l, \quad l = 0, \ldots, n-1,$$

where $f_0, \ldots, f_{n-1}$ are called development factors. Then the expected amounts paid are given by

$$\begin{aligned}
&\mathrm{E}[C_{-k,0}] = \mathrm{E}[D_{-k,0}], \\
&\mathrm{E}[C_{-k,l}] = \mathrm{E}[D_{-k,l} - D_{-k,l-1}] = \mathrm{E}[C_{-k,0}] f_0 \cdots f_{l-2}(f_{l-1} - 1), \quad l = 1, \ldots, n.
\end{aligned}$$

A simple model for $C_{-k,l}$ is obtained by assuming that the observed payments have the representation $C_{-k,l} = \mathrm{E}[C_{-k,l}]R_{-k,l}$, where $\{R_{-k,l}\}_{k,l=0}^{n}$ are independent and identically distributed with $\mathrm{E}[R_{-k,l}] = 1$.

A standard method for predicting the lower triangle (outstanding claims) is called the chain ladder method. In the chain ladder method, the development factors are estimated by

$$\widehat{f}_l = \frac{\sum_{k=l+2}^{n+1} D_{-k,l+1}}{\sum_{k=l+2}^{n+1} D_{-k,l}}, \quad l = 0, \ldots, n-1. \tag{7.13}$$

The expected amounts of paid claims, $\mathrm{E}[C_{-k,l}]$, in the upper triangle can be estimated by

$$\widehat{C}_{-k,0} = \widehat{D}_{-k,0} = \frac{D_{-k,k-1}}{\widehat{f}_0 \cdots \widehat{f}_{k-2}}, \tag{7.14}$$

$$\widehat{C}_{-k,l} = \widehat{D}_{-k,l} - \widehat{D}_{-k,l-1} = \frac{D_{-k,k-1}}{\widehat{f}_l \cdots \widehat{f}_{k-2}} \Big(1 - \frac{1}{\widehat{f}_{l-1}}\Big) \tag{7.15}$$

for $k = 1, \ldots, n+1,\ l = 1, \ldots, k-1$, and the residuals are computed as

$$R_{-k,l} = \frac{C_{-k,l}}{\widehat{C}_{-k,l}}, \quad k = 1, \ldots, n+1,\ l = 0, \ldots, k-1.$$

The predictions for the unobserved cumulative claim amounts in the lower triangle are given by

$$\widehat{D}_{-k,l} = D_{-k,k-1} \widehat{f}_{k-1} \cdots \widehat{f}_{l-1} \tag{7.16}$$

for $k = 1, \ldots, n$ and $l = k, \ldots, n$. The corresponding predictions for the unobserved future payments in the lower triangle are

$$\widehat{C}_{-k,l} = \widehat{D}_{-k,l+1} - \widehat{D}_{-k,l}. \tag{7.17}$$

The last row in the lower triangle of outstanding payments, corresponding to $k = 0$, represents amounts for claims that will occur during the next year. Therefore, it does not contain any observations in the upper triangle and cannot be predicted with the standard chain ladder. We will predict this row by predicting the initial payment

$D_{0,0}$ by the mean of the predictions for the previous years and then apply the chain ladder method for the predictions of $D_{0,l}$ for $l \geq 1$. More precisely, the predictions for the last row may be constructed as follows:

$$\widehat{D}_{0,0} = \frac{1}{n+1} \sum_{k=1}^{n+1} \frac{D_{-k,k-1}}{\widehat{f}_0 \cdots \widehat{f}_{k-2}}, \qquad \widehat{D}_{0,l} = \widehat{D}_{0,0} \widehat{f}_0 \cdots \widehat{f}_{l-1},$$

$$\widehat{C}_{0,0} = \widehat{D}_{0,0}, \qquad \widehat{C}_{0,l} = \widehat{D}_{0,l+1} - \widehat{D}_{0,l}.$$

When the prediction of all future payments $\widehat{C}_{-k,l}$, $k = 0, \ldots, n$, $l = k, \ldots, n$, is completed, the present value L_0 of the outstanding claims is computed as

$$L_0 = \sum_{k=0}^{n} \sum_{l=k}^{n} \widehat{C}_{-k,l} e^{-r_{l-k+1}(l-k+1)},$$

where $\mathbf{r}^{\mathrm{T}} = (r_1, \ldots, r_n)$ is the vector of current zero rates.

At time 1, new information is available as the values of the diagonal entries $C_{-k,k}$, for $k = 0, \ldots, n$, are observed. The value L_1 of the liabilities at time 1 will be computed similarly to L_0. First, the new observations $C_{-k,k}$, for $k = 0, \ldots, n$, are entered into the upper triangle of observed payments. Then, the development factors are updated by

$$\widehat{f}_l^{(1)} = \frac{\sum_{k=l+1}^{n+1} D_{-k,l+1}}{\sum_{k=l+1}^{n+1} D_{-k,l}} \quad \text{for } l = 0, \ldots, n-1,$$

and the predictions, denoted by $\widehat{C}_{-k,l}^{(1)}$, are updated accordingly by entering the updated development factors in (7.16) and (7.17). Assuming the zero rates at time 1 are $\mathbf{r} + \Delta\mathbf{r}$, the value of the liabilities at time 1 can be expressed as

$$L_1 = \sum_{k=0}^{n} C_{-k,k} + \sum_{k=0}^{n-1} \sum_{l=k+1}^{n} \widehat{C}_{-k,l}^{(1)} e^{-(r_{l-k} + \Delta r_{l-k})(l-k)}.$$

Note that L_1 is completely determined by the random variables $C_{-k,k}$, for $k = 0, \ldots, n$, and $\Delta\mathbf{r}$, all observed at time 1.

To protect the value of the liabilities against changes in the zero rates, it is assumed that the insurer has purchased a bond portfolio. Rewrite the current value of the liabilities by summing along the diagonals as

$$L_0 = \sum_{m=1}^{n+1} \left(\sum_{j=m-1}^{n-m+1} \widehat{C}_{-(m-j-1),j} \right) e^{-r_m m}.$$

Then we see that a good choice of the bond portfolio is obtained by buying $\sum_{j=m-1}^{n-m+1} \widehat{C}_{-(m-j-1),j}$ zero-coupon bonds with maturity m years from now. If the zero rate changes are independent of the claim amounts $C_{-k,k}$, for $k = 0, \ldots, n$, then this bond portfolio is the quadratic hedge of the value of the liabilities at time 1.

To compute the SCR, we need to apply the risk measure ρ to the quantity $A_1 - A_0 R_0 - L_1 + L_0 R_0$. Under the assumption that $A_0 = L_0$, it is sufficient to consider the distribution of $A_1 - L_1$. If the true development factors were known, the distribution of L_1 could be sampled by sampling the diagonal elements $C_{-k,k}$, for $k = 0, \ldots, n$, updating the prediction of the lower triangle, sampling the zero rate changes $\Delta\mathbf{r}$, and computing the outcome of L_1 by discounting the future payments in each simulated scenario. When adopting an empirical approach, the diagonal elements are sampled by sampling the residuals $R_{-k,k}$, for $k = 0, \ldots, m$, from the empirical distribution of the residuals and putting $C_{-k,k} = \widehat{C}_{-k,k} R_{-k,k}$. Similarly, the zero rate changes may be sampled from the empirical distribution of historical zero rate changes.

A problem with the empirical approach is that it does not account for parameter uncertainty in the development factors. The development factors, used for predicting the diagonal means $\widehat{C}_{-k,k}$, are not known but merely estimated from the upper triangle. Since the amount of data is rather limited, the parameter uncertainty may be substantial. To account for the parameter uncertainty, a bootstrap procedure can be implemented as outlined below. The algorithm below generates a sample from the so-called predictive distribution of $A_1 - L_1$, in which the parameter uncertainty is taken into account. The input to the algorithm is an upper triangle of amounts paid, as in the left table in Table 7.3. The algorithm proceeds as follows.

1. Compute the estimates $\widehat{f}_0, \ldots, \widehat{f}_{n-1}$ of the development factors by (7.13).
2. Compute the estimates $\widehat{C}_{-k,l}$ of $\mathrm{E}[C_{-k,l}]$ for $k = 1, \ldots, n+1$, $l = 0, \ldots, k-1$, by (7.14) and (7.15).
3. Compute the residuals $R_{-k,l} = C_{-k,l} / \widehat{C}_{-k,l}$, for $k = 1, \ldots, n+1$, $l = 0, \ldots, k-1$.
4. For each bootstrap iteration, $j = 1, \ldots, N$, repeat the following:
   (a) Draw with replacement bootstrapped residuals $R^*_{-k,l}$, for $k = 1, \ldots, n+1$, $l = 0, \ldots, k-1$, from the set $\{R_{-k,l}, k = 1, \ldots, n+1, l = 0, \ldots, k-1\}$.
   (b) Compute a bootstrapped upper triangle with entries $C^*_{-k,l} = \widehat{C}_{-k,l} R^*_{-k,l}$ for $k = 1, \ldots, n+1$, $l = 0, \ldots, k-1$.
   (c) Compute the development factors $f^*_0, \ldots, f^*_{n-1}$ of the bootstrapped upper triangle as in (7.13).
   (d) Compute one-step predictions $\widehat{C}^*_{-k,k}$, for $k = 0, \ldots, n$, using the bootstrapped upper triangle.
   (e) Draw with replacement the outcomes of diagonal residuals $R^{**}_{-k,k}$, for $k = 0, \ldots, n$, from the set of residuals $\{R_{-k,l}, k = 1, \ldots, n+1, l = 0, \ldots, k-1\}$.
   (f) Add the diagonal $C^{**}_{-k,k} = \widehat{C}^*_{-k,k} R^{**}_{-k,k}$, for $k = 0, \ldots, n$, to the bootstrapped upper triangle to form a sample of the upper triangle at time 1.

Table 7.4 Current zero rates used in Example 7.12

Maturity (years)	1	2	3	4	5	6	7	8	9	10
Zero rate (%)	0.82	1.57	2.16	2.54	2.82	3.04	3.23	3.37	3.49	3.58

(g) Compute the development factors $\widehat{f}_0^{**}, \ldots, \widehat{f}_{n-1}^{**}$ of the upper triangle at time 1.
(h) Compute the predictions $\widehat{C}_{-k,l}^{**}$, for $k = 1, \ldots, n, l = k, \ldots, n$, for the lower triangle at time 1.
(i) Draw one outcome $\Delta\mathbf{r}$ of zero rate changes from the set of historical zero rate changes.
(j) Compute the value of the liabilities at time 1 as

$$L_1 = \sum_{k=0}^{n} C_{-k,k}^{**} + \sum_{k=0}^{n-1} \sum_{l=k+1}^{n} \widehat{C}_{-k,l}^{**} e^{-(r_{l-k}+\Delta r_{l-k})(l-k+1)}$$

and the value of the bond portfolio as

$$A_1 = \sum_{k=0}^{n} \widehat{C}_{-k,k} + \sum_{k=0}^{n-1} \sum_{l=k+1}^{n} \widehat{C}_{-k,l} e^{-(r_{l-k}+\Delta r_{l-k})(l-k+1)}$$

and store the difference $A_1 - L_1$.

Example 7.12 (Sampling from the predictive distribution). Consider a nonlife insurer with upper triangle of paid claim amounts as in Table 7.5. The claims are assumed to be completely settled 9 years after the incident year. The objective is to determine the predictive distribution of the value of the assets minus the value of the liabilities, $A_1 - L_1$, 1 year from now. The bootstrapping algorithm outlined above is run. A historical sample of quarterly zero rate changes serves as the basis for generating annual zero rate changes $\Delta\mathbf{r}$. Each annual zero rate scenario is constructed by sampling four quarterly scenarios, with replacement, and adding them up. The current zero rates are given in Table 7.4.

A histogram of $N = 10{,}000$ samples from the predictive distribution of $A_1 - L_1$ is given in Fig. 7.7.

7.7 Notes and Comments

An introduction to the bootstrap and related resampling procedures, including statistical applications, is given in the classic book [11] by Bradley Efron and Robert Tibshirani.

Stochastic claims reserving techniques, extending the chain ladder, have been developed in the actuarial literature in recent decades by Thomas Mack [28] and

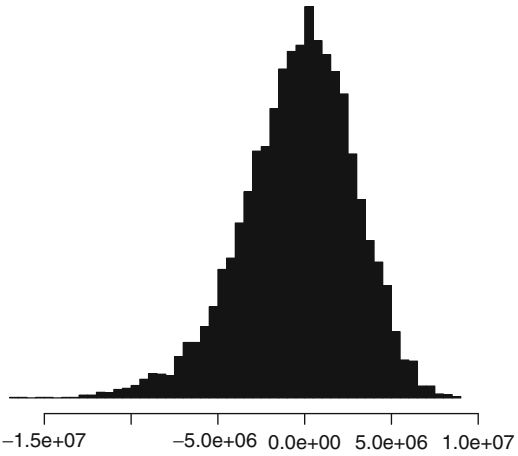

Fig. 7.7 Histogram in Example 7.12 for predictive distribution of the value of assets minus the value of the liabilities in 1 year from now

many others. A comprehensive treatment of such techniques is the book [46] by Mario Wütrich and Michael Merz. Our approach to bootstrapping the chain ladder method is a slight variation of the method presented by Peter England and Richard Verrall [14]. The upper claims triangle in Table 7.5 used in Example 7.12 originates from a paper by G.C. Taylor and F.R. Ashe [45].

7.8 Exercises

Exercise 7.1 (Empirical VaR). A unit within a bank is required to report an empirical estimate of $\mathrm{VaR}_{0.01}(X)$, where X is the portfolio value the next day from its trading activities. The empirical estimate $\widehat{\mathrm{VaR}}_{0.01}(X)$ is based on market prices from the previous $n + 1$ days that are transformed into a sample of size n from the distribution of X and the sample points are assumed to be independent and identically distributed. Compute the probability

$$\mathrm{P}\left(\widehat{\mathrm{VaR}}_{0.01}(X) > \mathrm{VaR}_{0.01}(X)\right)$$

as a function of n and determine its minimum and maximum for n in $\{100, \ldots, 300\}$.

Exercise 7.2 (Empirical tail conditional median). The tail conditional median $\mathrm{TCM}_p(X) = \mathrm{median}[L \mid L \geq \mathrm{VaR}_p(X)]$, where $L = -X/R_0$, has been proposed as a more robust alternative to $\mathrm{ES}_p(X)$ since $\mathrm{TCM}_p(X)$ is not as sensitive as $\mathrm{ES}_p(X)$ to the behavior of the left tail of the distribution of X.

Let Y have a standard Student's t distribution with ν degrees of freedom, and set $X = e^{0.01Y} - 1$. Consider the empirical estimators $\widehat{\mathrm{TCM}}_{0.05}(X)$ and $\widehat{\mathrm{ES}}_{0.05}(X)$ based on a sample of size 200 from the distribution of $L = -X$. Generate histograms based on samples of size 10^5 from the distributions of $\widehat{\mathrm{TCM}}_{0.05}(X)$ and $\widehat{\mathrm{ES}}_{0.05}(X)$ for $\nu = 2$ and $\nu = 10$.

Table 7.5 Upper triangle of paid claims and development factors in Example 7.12

	Development year									
Origin year	0	1	2	3	4	5	6	7	8	9
−10	357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948
−9	352,118	884,021	933,894	1,183,289	445,745	320,996	527,804	266,172	425,046	
−8	290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405		
−7	310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286			
−6	443,160	693,190	991,983	769,488	504,851	470,639				
−5	396,132	937,085	847,498	805,037	705,960					
−4	440,832	847,631	1,131,398	1,063,269						
−3	359,480	1,061,648	1,443,370							
−2	376,686	986,608								
−1	344,014									
−0										
Development factor	3.491	1.747	1.457	1.174	1.104	1.086	1.054	1.077	1.018	1.000

Exercise 7.3 (Empirical expected shortfall). Let $\{Z_1, \dots, Z_n\}$ be a sample of independent and identically distributed historical log returns that are distributed as the log return $\log(S_T/S_0)$ of an asset from today until time $T > 0$. Show that if the risk-free return over the investment period is 1, then the empirical estimator of $\mathrm{ES}_p(S_T - S_0)$ is given by

$$\min_c -c + \frac{1}{np}\sum_{k=1}^{n}(c + S_0 - S_0 e^{Z_k})I\{Z_k \le \log(1 + c/S_0)\}.$$

Exercise 7.4 (Empirical spectral risk measure). Let $\{Z_1, \dots, Z_n\}$ be a sample of independent and identically distributed historical log returns that are distributed as the log return $\log(S_T/S_0)$ of an asset from today until time $T > 0$. Show that if the risk-free return over the investment period is 1 and if ρ_ϕ is a spectral risk measure with risk aversion function ϕ, then the empirical estimator of $\rho_\phi(S_T - S_0)$ is given by

$$S_0\Big(1 - \sum_{k=1}^{n}\phi_k e^{Z_{k,n}}\Big), \quad \text{where } \phi_k = \int_{(n-k)/n}^{(n-k+1)/n}\phi(u)du.$$

Project 7 (Total returns). Consider a 5-year investment in a portfolio of dividend-paying stocks. The yearly portfolio returns S_{t+1}/S_t and dividends D_{t+1} paid at time $t + 1$ are modeled as

$$\frac{S_{t+1}}{S_t} = e^{\mu + 0.2X_{t+1}} \quad \text{and} \quad \frac{D_{t+1}}{S_t} = 0.05e^{-0.05^2/2 + 0.05Y_{t+1}},$$

where $X_1, \dots, X_5, Y_1, \dots, Y_5$ are independent and standard normally distributed.

(a) Consider the value in 5 years of investing \$1 million in a portfolio of stocks and reinvesting the dividends in the portfolio of stocks. Determine the function f such that the value V_5 in 5 years of the investment strategy can be expressed as $V_5 = f(\mu, X_1, \dots, X_5, Y_1, \dots, Y_5)$.

(b) Simulate a sample of suitable size from the distribution of $(X_1, \dots, X_5, Y_1, \dots, Y_5)$ and use this sample to determine the empirical distribution of V_5 for a range of values of the parameter μ. Estimate the smallest value of μ for which the probability that V_5 exceeds the value in 5 years of an investment of \$1 million in a 5-year zero-coupon bond with zero rate 5% per year is 0.75.

Project 8 (Pension savings). Consider a yearly investment of \$1,000 in long positions in a portfolio of stocks and a risk-free, 1-year, zero-coupon bond over a 30-year period. The yearly returns on the portfolio of stocks in year k is modeled as $R_k = e^{\mu + \sigma Z_k}$, where Z_k is standard normally distributed. The yearly returns are assumed to be independent. The yearly return on the risk-free bond is assumed to be $e^{0.01}$. The fraction of the yearly amount invested in the portfolio of stocks at the beginning of year k is $p(1 - c(k-1)/30)$, where $p, c \in [0, 1]$.

(a) Determine a function f such that the value of the pension savings in 30 years can be expressed as $V_{30} = f(\mu, \sigma, p, c, Z_1, \ldots, Z_{30})$.
Simulate a sample of suitable size n from the distribution of $(Z_1, \ldots, Z_{30})$ and use this sample to determine the empirical distribution F_n of V_{30} for a range of values of the parameters μ, σ, p, c.

(b) Set $\mu = 0.06$ and $\sigma = 0.2$ and investigate the effects on the empirical distribution $F_n(p, c)$ of V_{30} of varying p and c. Suggest a suitable criterion for selecting the optimal empirical distribution $F_n(p, c)$ and determine the optimizer (p, c).