MATHEMATICAL METHODS FOR PORTFOLIO MANAGEMENT

by

GUY-ROGER ABESSOLO ONDO

submitted in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

in the subject

OPERATIONS RESEARCH

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: PROF PETRUS H. POTGIETER

AUGUST 2002

ABSTRACT

Portfolio Management is the process of allocating an investor's wealth to investment opportunities over a given planning period. Not only should ***Portfolio Management*** be treated within a multi-period framework, but one should also take into consideration the stochastic nature of related parameters.

After a short review of key concepts from Finance Theory, e.g. utility function, risk attitude, Value-at-Risk estimation methods, and mean-variance efficiency, this work describes a framework for the formulation of the ***Portfolio Management*** problem in a Stochastic Programming setting. Classical solution techniques for the resolution of the resulting Stochastic Programs (e.g. L-shaped Decomposition, Approximation of the probability function) are presented. These are discussed within both the two-stage and the multi-stage case with a special emphasis on the former. A description of how Importance Sampling and EVPI are used to improve the efficiency of classical methods is presented. Postoptimality Analysis, a sensitivity analysis method, is also described.

Keywords: Approximation Schemes; Extreme Value Theory; Importance Sampling; Nested Decomposition; Portfolio Management; Postoptimality Analysis; Progressive Hedging; Scenario Aggregation; Stochastic Programming; Stochastic Quasi-gradient; Value-at-Risk

Contents

Chapter 1

Introduction

Simply speaking, *portfolio management* is the process of allocating an investor's wealth to investment opportunities over a given planning period. Hence, it is a dynamic decision making process, one that is aimed at maximising the expected terminal wealth, or its utility, while allowing the investor to meet future obligations and an investment objective.
There are 6 main categories of investors:

- banks,
- individuals,
- insurance companies,
- mutual funds,
- non-profit organisations, and
- pension funds.

Different investors have different investment objectives and are subjected to different sets of constraints. However, the investment opportunities do have a common characteristic: their future returns are not always known with certainty, and are, therefore, subject to risk. This requires a periodic revision of the portfolio. To model inter-temporal decision making processes subject to risky conditions, an interesting tool is available: Stochastic linear programs with recourse; that is, linear programs whose solutions are corrective actions after the random events have taken place.

In some instances, it is very simple to solve the resulting stochastic program. This is the case when the random variables assume discrete values and the sample space is finite and small; in this case, the integral which appears in the objective function of the program is replaced by a finite sum.

It can also happen that the objective function is separable with respect to the components of the random elements; in this case, the integral becomes a sum of m(m the number of random variables) one-dimensional integrals.

In both cases, the deterministic equivalent of the stochastic program can be solved directly using standard techniques, e.g. simplex method, interior point. However, for more complicated cases, one has to make use of more sophisticated solution techniques. These techniques fall into 3 main categories:

- **Approximation schemes**
  These schemes approximate the integrand or the probability function. Under certain conditions, the solution of the resulting approximate problems will converge to the solution of the original problem.

- **Stochastic approximations**
  Stochastic approximation methods are iterative methods. At each iteration, the expected recourse function is replaced with a randomly generated gradient. The sequence of feasible solutions obtained by solving the approximate problems will, under appropriate conditions, converge in probability to the solution of the true problem.

- **Mathematical programming techniques**
  These techniques exploit the structure of the original problem to devise efficient solution algorithms.

The objective of this thesis is to discuss representatives of these solution techniques.

The discussion is articulated around three main parts. In chapter 2, the financial theory behind *portfolio management* is introduced. First, we discuss the existence of a utility function, and explain its meaning in terms of the investor's willingness to take a risk. We also discuss important concepts such as certainty equivalent, risk and mean-variance efficiency and related selection models. Some value-at-risk estimation methods are discussed as well. A framework for multi-period modelling is described at the end of that chapter.

In chapter 3, we discuss two-stage stochastic programs. After a brief definition of a linear stochastic program with recourse and a brief description of its properties, we present some solution techniques. First, we present the L-shaped decomposition method, a representative of the mathematical programming techniques. Next, we discuss approximation schemes i.e., approximation of the probability function. In the last part of this chapter, we present Monte Carlo sampling techniques and discuss how they can be used in conjunction with the L-shaped method to secure the efficient resolution of large-scale problems.

In chapter 4, we discuss multi-stage problems. We give an extension of the L-shaped decomposition method to the multi-stage case, and discuss how Monte

Carlo techniques can be combined with the L-shaped method. We also discuss how the expected value of perfect information, EVPI, can be used in conjunction with Monte Carlo sampling to enhanced the effectiveness of the L-shaped decomposition method. Scenario analysis and related techniques are also discussed. The last solution technique (which is a stochastic approximation scheme) is the stochastic quasi-gradient method, SQG, which will be applied to the two-stage case. Note that, unlike the other method, the SQG can also be used to solve nonlinear and nonconvex problems.

Chapter 2

Financial background

Portfolio management is a very versatile discipline which encompasses all the different facets of Finance (see [2], [22], [60]). It includes activities as diverse as security analysis, pricing, assessment of risk, risk control as well as financial engineering. In spite of this diversity, one should, however, note that these activities are very often integrated and interdependent. The discussion will be restricted to the portfolio manager's primary concern; that is, the construction and maintenance of portfolios with or without specific risk-return characteristics.

As discussed in the introduction, there are different types of investors. Due to the nature of their activities, some investors prefer current income over high (even exceptionally high) future returns, while others are more interested in always staying fully funded. There are also investors who are prepared to take a bet on their market's views, while others, less "risk seeking", hedge their positions to protect their investment. In other cases, the investor wants the future stream of cash flows to match her liabilities.

The portfolio manager's problem may, therefore, be summarised by the following question:

How do I build my client's portfolio (i.e. what assets should be included in the portfolio and in what quantity), in order to match her profile?

In order to solve this problem, the portfolio manager will need, among other things, to:

- assess the client risk profile,
- assess her needs,
- find out the options that are available, and

- specify how to choose among those options.

Once this analysis is completed, a mathematical model is formulated and solved. In what follows, tools that will assist the portfolio manager in solving the mentioned problem will be presented. We start with an introduction of important concepts such as "utility function", "risk", "certainty equivalent", "risk aversion" and "mean-variance efficiency" . Next, we discuss portfolio selection models. To conclude the chapter, we present some strategies for asset allocation.

2.1 The investor's utility function

Let us consider that N assets are traded. The j^{th} asset, X_j, is characterised by its price p_j, its future cash flow $\tilde{x}_j$, the corresponding return $\tilde{r}_j$ and the probability density function of the future cash flow.
Let us assume that the set $\mathbf{X} = \{X_1, \ldots, X_N\}$, i.e. the set of traded assets, is a mixture set; that is, it satisfies the following four postulates (see [41]:70-72):

Postulate 1 *If $X_i, X_j \in \mathbf{X}$, then the portfolio Z, composed of pX_i and $(1-p)X_j$ with $p \in [0,1]$ and $i \neq j$, is also a traded asset.*

Postulate 2 *If $X_i, X_j \in \mathbf{X}$ and $p \in [0,1]$, then holding the portfolio Z, made up of pX_i and $(1-p)X_j$, or holding the portfolio Y, composed of $1X_i$ and $0X_j$, is equivalent to holding only asset X_i.*

Postulate 3 *If $X_i, X_j \in \mathbf{X}$ and $p, q, r \in [0,1]$, then holding the portfolio Z, composed of pY and $(1-p)V$ (where Y is the portfolio containing rX_i and $(1-r)X_j$, and V is the portfolio containing $qX_i$and $(1-q)X_j$) is equivalent to holding a portfolio containing $(pr + (1-p)q)X_i$ and $(1 - pr - (1-p)q)X_j$.*

We further assume that the investor is endowed with a preference relation π having the following properties:

Property 1 (Reflexivity) *If $X_i \in \mathbf{X}$, then the investor is indifferent between holding a portfolio containing only asset X_i and holding a different portfolio having the same composition.*

Property 2 (Comparability) *If $X_i, X_j \in \mathbf{X}$ $(i \neq j)$ the investor will either prefer asset X_i or prefer asset X_j; she is otherwise indifferent to them.*

Property 3 (Transitivity) *For any $X_i, X_j, X_k \in \mathbf{X}$, if asset X_i is preferred (or indifferent) to asset X_j and asset X_j is preferred (or indifferent) to asset X_k, then asset X_i is preferred (or indifferent) to asset X_k.*

Property 4 (Order Preserving) *For any $X_i, X_j \in \mathbf{X}$ $(i \neq j)$ where asset X_i is preferred to asset X_j and $\alpha, \beta \in [0,1]$, the portfolio composed of αX_i and $(1-\alpha)X_j$ is preferred to the portfolio composed of βX_i and $(1-\beta)X_j$ if and only if $\alpha > \beta$.*

Property 5 (Intermediate Value) *For any $X_i, X_j, X_k \in \mathbf{X}$, if asset X_i is preferred to asset X_j and asset X_j is preferred to asset X_k, then there exists a unique $p \in [0,1]$ such that the portfolio composed of pX_i and $(1-p)X_k$ is equivalent to asset X_j.*

Property 6 (Boundedness) *It is possible to find the investor's least preferred and the most preferred assets among the traded assets.*

Property 7 (Strong Independence) *For any $X_i, X_j \in \mathbf{X} \quad (i \neq j)$, if asset X_i is preferred to asset X_j, then for any $X_k \in \mathbf{X}$, the portfolio composed of pX_i and $(1-p)X_k$ is also preferred to the portfolio composed of pX_j and $(1-p)X_k$; that is, the introduction of further uncertainty does not change preferences.*

Then the investor will be characterised by a utility function, $U(\cdot)$, such that every asset X_j (i.e. investment opportunity)can be represented by a real value corresponding to the utility of its expected future cash flow (for a proof, see [41]:75). For all j, we have

$$U_j \quad = \quad E[U(\tilde{x}_j)]. \tag{2.1}$$

Remark 2.1 Note that, though an investor can compare the utility level of different assets, the utility of an asset cannot be compared across investors.

2.2 Characterisation of risk attitudes

Consider the asset X_k; its expected utility level to the investor is

$$U_k = E[U(\tilde{x}_k)]. \tag{2.2}$$

We would like to know the minimum amount of money, C, that will make the investor indifferent, in terms of utility, between holding asset X_k and having C; that is, we are looking for the real value C such that

$$U(C) = E[U(\tilde{x}_k)]. \tag{2.3}$$

The amount C, if it exists, is called the certainty-equivalent to X_k.

We will, now, use the notion of certainty-equivalent to characterise the investor's risk attitude. An investor is said to be

risk averse : **if** $C < E[\tilde{x}_k]$,

risk neutral : **if** $C = E[\tilde{x}_k]$, **or**

risk seeking : **if** $C > E[\tilde{x}_k]$.

Remark 2.2 The certainty-equivalent can also be used to compare the attitude toward risk of two different investors. Thus, investor A is said to be more risk averse than investor B if, for every portfolio, the certainty-equivalent to the portfolio for investor A is less than or equal to the certainty-equivalent to the corresponding portfolio for investor B.

Alternatively, the investor's attitude toward risk may be found using the second derivative of her utility function; that is, the investor is:

$$\textbf{risk averse} \quad : \quad \textbf{if} \quad U''(x) < 0,$$

$$\textbf{risk neutral} \quad : \quad \textbf{if} \quad U''(x) = 0, \textbf{or}$$

$$\textbf{risk seeking} \quad : \quad \textbf{if} \quad U''(x) > 0.$$

Note that this set of definitions and the previous one (coupled with the definition of C) establish a correspondence between the investor's risk attitude and the shape of the utility function.

The Arrow-Pratt measure

$$A(x) = -\frac{U''(x)}{U'(x)}, \tag{2.4}$$

where x is the investor's level of wealth,

can be used to measure the investor's absolute risk aversion.

The Arrow-Pratt measure's first derivative $A'(x)$ will describe the evolution of the investor's absolute risk aversion with respect to changes in wealth; that is,

$$A'(x) > 0 \quad : \quad \textbf{increasing absolute risk aversion,}$$

$$A'(x) = 0 \quad : \quad \textbf{constant absolute risk aversion, or}$$

$$A'(x) < 0 \quad : \quad \textbf{decreasing absolute risk aversion.}$$

2.3 Standard risk measures

Risk may be perceived in different ways depending on the investor's preoccupation. Different risk measures can be used; below, we give two examples of risk measures:

The Standard deviation

$$\sigma_p = \Big(\int_{-\infty}^{+\infty} (\tilde{r}_p - E[\tilde{r}_p])^2 f(\tilde{r}_p) d\tilde{r}_p \Big)^{1/2}, \tag{2.5}$$

where $\tilde{r}_p$ is the return of the portfolio, and $f(\cdot)$ is the probability density function of the return,

measures the deviation about the mean. It is used to measure the total risk of a portfolio by opposition to the risk relative to a benchmark (e.g. the market).

The semi-standard deviation

$$ssd_p = \Big(\int_{-\infty}^{\tau} (\tilde{r}_p - \tau)^2 f(\tilde{r}_p) d\tilde{r}_p \Big)^{1/2}, \tag{2.6}$$

where τ is a given target return,

measures the deviation below the target return, τ.

Already in the early years of modern finance, Roy (see [59]) claimed that an investor would prefer safety of principal first, and would choose the investment with the smallest probability of going below the principal. The return corresponding to the principal was called the Minimum Acceptable Return (MAR). Markowitz (see [49]) agreed that:

1- only downside risk or safety first is relevant to an investor, and

2- in the case of non-normal asset return distributions, a downside risk measure would be more appropriate.

Indeed, in the case of non-normal distributions (e.g. asymmetric distributions), a negative skewness of the probability function will result in downside returns with larger magnitude of returns than the upside returns. A positive skewness will have upside returns with larger magnitude of returns than the downside returns. The deviation below the MAR appears, therefore, to be a more suitable measure. However, the implicit assumption of risk aversion embedded in the risk measures presented above limits their application to a particular type of investors .

A more general family of downside risk measures was introduced by Bawa (see [3]); i.e. the Lower Partial Moment (LPM) defined as:

$$LPM(a, \tau) = \frac{1}{N} \sum_{i=1}^{N} max\Big(0, (\tau - R_i)\Big)^a, \tag{2.7}$$

where N is the number of observations, τ is the target return, a ($a > 0$) is the degree of the lower partial moment, and R_i is the return on the asset during the period i.

The LPM is directly related to the risk tolerance of the investor through the parameter a; we have

$a < 1$: **risk seeking investor,**
$a = 1$: **risk neutral investor, or**
$a > 1$: **risk averse investor.**

For $a = 2$, the $LPM(2, \tau)$ corresponds to the semi-variance presented earlier. The parameter a can take any positive value; i.e. the measure includes all types of investor behaviour (see [26]).

2.4 Value-at-risk

Let $\Delta S = S_{t+1} - S_t$ be the change in the value of a portfolio between time t and time $t + 1$. S_{t+1} and S_t are the value of the portfolio at $t + 1$ and t respectively. Let $f_{\Delta S}$ be the probability density function of ΔS. Then the value-at-risk (VaR), the maximum loss for a given level of confidence (i.e. α) incurred by holding a position during a determined time interval, is the number which satisfies the following equation:

$$1 - \alpha = \int_{-\infty}^{-VaR} f_{\Delta S}(x)dx, \tag{2.8}$$

that is,

$$P(\Delta S \leq -VaR) = 1 - \alpha. \tag{2.9}$$

In the remaining part of the section we will present some of the methods used to estimate VaR.

2.4.1 Analytical method

The standard variance-covariance method

Let the price of the shares follow the process (see [39])

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma B_t}, \tag{2.10}$$

where $B_t \sim N(0, \sqrt{t})$, μ and σ are the mean and the standard deviation of the rate of return on the shares.

The logarithmic return, from time t to time T, of a portfolio containing one share is

$$\begin{aligned} r_{t,T} &= log(\frac{S_T}{S_t}) \\ &= \left((\mu - \frac{1}{2}\sigma^2)T + \sigma B_T\right) - \left((\mu - \frac{1}{2}\sigma^2)t + \sigma B_t\right) \\ &= (\mu - \frac{1}{2}\sigma^2)(T - t) + \sigma(B_T - B_t), \end{aligned} \tag{2.11}$$

and equation (2.11) can be written

$$r_{t,T} = (\mu - \frac{1}{2}\sigma^2)(T-t) + \sigma\epsilon\sqrt{T-t}, \tag{2.12}$$

where $\epsilon \sim N(0,1)$.

Since the mean of the logarithmic return is very small compared to its variance for short horizons (a few days), it is common to consider that $\mu = \frac{1}{2}\sigma^2$; equation (2.12) becomes

$$r_{t,T} = \sigma\epsilon\sqrt{T-t}; \tag{2.13}$$

that is,

$$\begin{aligned} \Delta S &= S_t(e^{r_{t,T}} - 1) \\ &= S_t(e^{\sigma\epsilon\sqrt{T-t}} - 1). \end{aligned} \tag{2.14}$$

Using the approximation $e^x = 1 + x + o(x)$, the previous equality becomes

$$\Delta S = \sigma\epsilon\sqrt{T-t}S_t, \tag{2.15}$$

and the α confidence level VaR is

$$VaR = \sigma z_\alpha\sqrt{T-t}S_t, \tag{2.16}$$

where z_α is the $\alpha-$quantile of a standard normal distribution.

If we consider a portfolio of N shares with covariance matrix V containing ω_j shares of type j, we have

$$VaR = z_\alpha\sqrt{<\omega S_t, V\omega S_t>}\cdot\sqrt{T-t}, \tag{2.17}$$

where $<\omega S_t, V\omega S_t> = \sigma_p^2 S_t^2$; the variance of the portfolio is

$$\sigma_p^2 = \sum_{i=1}^{N}\omega_i^2\sigma_i^2 + 2\sum_{i=1}^{N}\sum_{j=1,j<i}\omega_i\omega_j\sigma_{ij}. \tag{2.18}$$

Note that for a multi-period return, e.g.

$$\begin{aligned} r_{t,t+2} &= log(\frac{S_{t+2}}{S_t}) \\ &= log(\frac{S_{t+1}}{S_t}) + log(\frac{S_{t+2}}{S_{t+1}}) \\ &= r_{t,t+1} + r_{t+1,t+2}, \end{aligned} \tag{2.19}$$

the time varying standard deviation of the returns, σ_t, can be obtained using any of the following methods (see [9], [36]):

- *Equally weighted moving average approaches*

$$\sigma_t = \sqrt{\frac{1}{(k-1)} \sum_{i=t-k}^{t-1} (x_i - \mu)^2}, \tag{2.20}$$

where σ_t denotes the estimated standard deviation at the beginning of day t. The parameter k is the number of days included in the moving average (i.e. the observation window); x_i is the rate of return of the share at time i, and μ is the mean of the rate of return.

- *Exponentially weighted moving average approaches*

$$\sigma_t = \sqrt{(1-\lambda) \sum_{i=t-k}^{t-1} \lambda^{t-i-1} (x_i - \mu)^2}, \tag{2.21}$$

where λ is a decay factor.

- *Generalised autoregressive heteroscedastic (GARCH) approaches*

$$\sigma_{t+1}^2 = \frac{\alpha_0}{1-\beta} + \alpha_1 \sum_{k=0}^{t} \beta^k r_{t-k}^2, \tag{2.22}$$

where the parameters α_0, α_1 and β can be obtained using the maximum likelihood method.

For nonlinear instruments (instruments whose payoff function is not linear, e.g. options), one can use a Taylor based approximation of the portfolio value function (e.g. delta approximation), or consider the portfolio value corresponding to the VaR to be the price of the underlying asset at the maturity date and calculate the current price of the instrument using the appropriate formula.

Cornish-Fisher approximation

In the previous discussion the return is assumed to follow a normal distribution. However, this assumption is true only for very large samples. The Cornish-Fisher expression, w_α, approximates the $\alpha-$quantile of an empirical distribution using its moments (see [31], [39]), i.e.

$$w_\alpha = z_\alpha + \frac{1}{6}(z_\alpha^2 - 1)m_3 + \frac{1}{24}(z_\alpha^3 - 3z_\alpha)m_4 + \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)m_3^2, \tag{2.23}$$

where:

$$m_3 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}, \text{and} \tag{2.24}$$

$$m_4 = \frac{\mu_4}{\mu_2^2} - 3. \tag{2.25}$$

The parameters μ_3 and μ_4 are the third and fourth moment of the rate of return .

In this case, VaR is

$$VaR = \sigma w_\alpha \sqrt{T-t} S_t. \tag{2.26}$$

2.4.2 Monte Carlo simulation

From the previous derivation of VaR, we know that

$$r_{t,t+1} = \sigma\epsilon, \tag{2.27}$$

where $\epsilon \sim N(0,1)$.

The simulation proceeds as follows:

1- Use equation (2.27) to simulate n returns, i.e. $r^1, r^2, \ldots, r^n$; where

$$r^i = \sigma\epsilon_i. \tag{2.28}$$

2- For each simulated return r^i, we calculate the corresponding change in the share value, $\Delta S^i = S_i(e^{r^i} - 1)$.

3- Sort the changes in the share value from the lowest value to the highest value and determine VaR based on the desired confidence level.

2.4.3 Historical simulation

The historical simulation is a non-parametric approach; no assumptions about the distribution are needed. It uses the empirical distribution of past data to find an estimate of VaR. One uses the actual changes in the market factors (risk factors, e.g. market prices and interest rates) that occurred between past consecutive pairs of dates to construct the distribution. In order to understand the method, let $S_0, \ldots, S_{N-1}$ be the value of the share for N consecutive days starting at $t = 0$; the price change is

$$\Delta S = S_i - S_{i-1}. \tag{2.29}$$

Once the change for every period has been calculated, the value-at-risk is read from a list containing all the observed changes. The resulting VaR reflects the dynamics of the past history of the risk factor. Hence, its accuracy will depend on how well past values are representative of future ones. The method is simple and, therefore, easy to implement. Unfortunately, depending on its nature, the historical sample data may not contain extreme market movements; the consequence is an underestimation of VaR. If the sample contains several extreme movements, VaR can be overestimated. The same problems can be experienced in the case of a conjunction of a trend in the sample and a reversal of the trend in the prediction period.

2.4.4 VaR methods based on extreme value theory

The distribution of extreme values

From extreme value theory, it is known that (see [23]), under certain conditions, the distribution of the extreme values of a non overlapping series of observations, F, converges asymptotically to

$$H_{\xi,\mu,\sigma}(x) = \begin{cases} exp[-(1+\xi\frac{(x-\mu)}{\sigma})]^{\frac{-1}{\xi}} & \text{if} \quad \xi \neq 0 \\ exp[-e^{\frac{(x-\mu)}{\sigma}}] & \text{if} \quad \xi = 0, \end{cases} \tag{2.30}$$

where $1+\xi\frac{(x-\mu)}{\sigma} > 0$.

μ and σ represent the location and scale parameters while ξ, the shape parameter or tail index, indicates the thickness of the tail of the distribution.

Another important result from extreme value theory is that the limiting distribution of excess over a given (high) threshold is a generalised Pareto distribution; i.e. the conditional distribution of excesses over a high threshold u

$$F_u(x) = p(X - u \leq x \backslash X > u), \tag{2.31}$$

can be approximated by

$$G_{\xi,\beta(u)}(x) = \begin{cases} exp[1-(1+\frac{\xi x}{\beta})]^{\frac{1}{\xi}} & \text{if} \quad \xi \neq 0 \\ exp[1-e^{\frac{-x}{\beta}}] & \text{if} \quad \xi = 0, \end{cases} \tag{2.32}$$

where X is the random variable representing the extreme value of a series;

that is,

$$F_u(x) \simeq G_{\xi,\beta(u)}(x), \quad \text{as} \quad x \to \infty. \tag{2.33}$$

VaR estimates

- The parametric case

$H_{\xi,\mu,\sigma}(x)$ described above is assumed to be the exact distribution of the extreme observations; the p-quantile is given by

$$\begin{aligned} \hat{x}_p &= H^{-1}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(x) \\ &= \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}[(-log(p))^{-\hat{\xi}} - 1], \end{aligned} \tag{2.34}$$

where $\hat{\xi}, \hat{\mu}$ and $\hat{\sigma}$ are obtained using the maximum likelihood estimation method.

- The semi-parametric case

It is now assumed that the extreme events are only roughly distributed like

$H_{\xi,\mu,\sigma}(x)$; in this case, we have:

* *In-sample estimate*

$$\hat{x}_p = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}[(n(1-p))^{-\hat{\xi}} - 1]. \tag{2.35}$$

* *Out-of-sample estimate*

$$\hat{x}_p = (\frac{n}{k}(1-p))^{-\hat{\xi}} X_k, \tag{2.36}$$

where X_k is a very high value from the ordered data set, and the estimate, $\hat{\xi}$, of the tail index is obtained using the Hill estimation method (see [37]).

* *Fitting excesses over a high threshold*

Assuming a very high threshold, u, and using the fact that the limiting distribution of excesses over this threshold is a generalised Pareto distribution, we have

$$\hat{x}_p = u + \frac{\hat{\sigma}}{\hat{\xi}}[(\frac{n}{N_u}(1-p))^{-\hat{\xi}} - 1], \tag{2.37}$$

where N_u is the random number of exceedances of u and n is the number of observations.

These results can be combined with other methods (e.g. historical simulation, GARCH models) to produce new VaR estimation methods.

2.4.5 Extreme value theory and historical simulation

This methods integrates the distribution of the tail, $F(x)$, obtained using extreme value theory into the empirical distribution of the data producing a new distribution (see [14]). This is described in the following algorithm:

Draw x^i from the data set $\{x_t\}_{t=1}^T$ with replacement
if $x^i < X^{lower}$ then
    draw x^i from $F(x)$ for the lower tail
else
    if $x^i < X^{upper}$ then
        draw x^i from $F(x)$ for the upper tail
    else
        keep x^i
    end if
end if

where X^{upper} and X^{lower} are the thresholds for the upper and lower tail respectively. $\{x_t\}_{t=1}^{T}$ is the data from time $t = 1$ to time T.

The resulting data set is then treated within a standard historical simulation framework.

2.4.6 Extreme value theory and GARCH models

It is assumed that the stationary time series $(X_t, t \in Z)$ representing the daily negative logarithmic returns satisfies

$$X_t = \mu_t + \sigma_t Z_t, \tag{2.38}$$

where the Z_t's are *iid* with zero mean, variance 1 and distribution $F_Z(z)$. It is also assumed that μ_t and σ_t are measurable with respect to $\mathcal{G}_{t-1}$, the information about the return process available up to time $t-1$.

Consider $F_X(x)$ the marginal distribution of X_t, and $F_{X_{t+1}\backslash\mathcal{G}_t}(x)$ the predictive distribution of the return for next day given the information available up to time t. Hence, the conditional q-quantile is defined by

$$x_q^t = \inf\{x \in R : F_{X_{t+1}\backslash\mathcal{G}_t}(x) \geq q\}, \tag{2.39}$$

and since

$$\begin{aligned} F_{X_{t+1}\backslash\mathcal{G}_t}(x) &= P(\sigma_{t+1}Z_{t+1} + \mu_{t+1} \leq x\backslash\mathcal{G}_t) \\ &= F_Z(\frac{x-\mu_{t+1}}{\sigma_{t+1}}), \end{aligned} \tag{2.40}$$

we have

$$x_q^t = \mu_{t+1} + \sigma_{t+1} z_q, \tag{2.41}$$

where z_q is the upper qth quantile of the marginal distribution of Z_t which, by definition, does not depend on t.

This method uses extreme value theory to find an estimate of z_q (see [50]). The method is summarised below:

1. Estimate μ_{t+1} and σ_{t+1} using a GARCH-type model and calculate the implied residuals z.
2. Use extreme value theory to model the tail of $Z_Z(z)$ and estimate z_q.

The VaR at q level confidence is then

$$\begin{aligned} VaR &= x_q^t \\ &= \hat{\mu}_{t+1} + \hat{\sigma}_{t+1}\hat{z}_q, \end{aligned} \tag{2.42}$$

where the "$\hat{}$" refers to an estimated value.

2.4.7 L-moments based approach

This method, proposed by Gabriel Bonti, Jonathan Hosking and Dirk Siegel (see [10]), is based on the logarithmic daily returns

$$r_t = log\left(\frac{S_t}{S_{t-1}}\right), \tag{2.43}$$

where S_t is the market variable (e.g. closing price) on trading day t. The stochastic nature of the volatility is introduced by scaling the daily returns. The distribution of the scaled returns

$$s_t = \frac{r_t}{v_{t/t-1}}, \tag{2.44}$$

where $v_{t/t-1}$ is the volatility of trading day t estimated using data up to day $t-1$.

is obtained using the L-moments methods (see [38]).

The resulting distribution is then transformed to obtained a Gaussian distribution using the probability integral transform, i.e.

$$z_t = \Phi^{-1}(F(s_t)), \tag{2.45}$$

where F is the cumulative distribution function of s_t and Φ^{-1} is the inverse cumulative distribution function of the standard Gaussian distribution.
In the multi-factor case, the joint distribution of the transformed variables is multivariate Gaussian (or Meta-Gaussian, see [45]).
The resulting simulation procedure is

1. Generate z from the Gaussian distribution.
2. Transform it to have the required marginal distribution, $s = F^{-1}(\Phi(z))$.
3. Form the rescaled return $r = sv_{t+1/t}$.
4. Convert to the original variable $S_{t+1} = S_t e^r$.

The simulation procedure can be repeated to generate as many scenarios as required. The VaR is obtained using an approach similar to the one described for the Monte Carlo simulation method.

2.5 Mean-variance efficient portfolio

The concept of mean-variance efficiency was introduced by Markowitz in his pioneering article(see [48]). He argues that **"the investor does (or should) consider expected return a desirable thing and variance of the return**

an undesirable thing". That is, the knowledge of both the return and its variance is sufficient to choose between two assets.

In essence, a portfolio $(\alpha_1, \ldots, \alpha_N)$ of risky assets with expected return δ is said to be mean-variance efficient if

$$var\Big(\sum_{j=1}^{N} \alpha_j \tilde{r}_j\Big) \leq var\Big(\sum_{j=1}^{N} \eta_j \tilde{r}_j\Big) \tag{2.46}$$

for every portfolio $(\eta_1, \ldots, \eta_N)$ such that

$$\delta = E[\sum_{j=1}^{N} \alpha_j \tilde{r}_j] = E[\sum_{j=1}^{N} \eta_j \tilde{r}_j], \tag{2.47}$$

where α_j and η_j represent the fraction of the portfolio's value allocated to asset X_j.

In other words, a mean-variance efficient portfolio for a given expected return δ, is the portfolio, among all the portfolios having expected return δ, which has the smallest variance.

Alternatively, the portfolio $(\alpha_1, \ldots, \alpha_N)$ with variance γ, is said to be mean-variance efficient if

$$\sum_{j=1}^{N} \alpha_j \tilde{r}_j \geq \sum_{j=1}^{N} \eta_j \tilde{r}_j \tag{2.48}$$

for every portfolio $(\eta_1, \ldots, \eta_N)$ such that

$$\gamma = var\Big(\sum_{j=1}^{N} \alpha_j \tilde{r}_j\Big) = var\Big(\sum_{j=1}^{N} \eta_j \tilde{r}_j\Big). \tag{2.49}$$

That is, a mean-variance efficient portfolio for a given variance γ, is the portfolio, among all the portfolios having variance γ, which has the highest return.

Remark 2.3 Note that in the case where the investor has to choose between two assets such that $E[\tilde{r}_1] = E[\tilde{r}_2]$, $var(\tilde{r}_1) = var(\tilde{r}_2)$ but with different third moments (i.e. one is skewed to the right while the other one has a symmetric density function), she will probably choose the asset whose density function is skewed. This observation suggests a limitation of the mean-variance efficiency concept (when the variance, var is assumed to be the measure of risk).

The mean-variance efficiency is compatible with utility maximisation when the investor has a quadratic utility function (a quadratic utility function also implies that the investor is risk averse). This is also true if the returns are normally distributed.

2.6 Models based on mean-variance efficiency

2.6.1 Single period models

We will again consider the N assets introduced earlier, and we let

- $\mathbf{V}$ be the variance-covariance matrix ($\mathbf{V}$ is assumed to be positive definite),
- r_j be the expected return of the j-th assets,
- $\bar{r}$ be a vector whose components represent the returns on individual assets; i.e., $\bar{r} = (r_1, \ldots, r_N)^T$, and
- α be a vector whose components represent the fraction of the portfolio total value allocated to individual assets, i.e. $\alpha = (\alpha_1, \ldots, \alpha_N)^T$.

The mean-variance efficient portfolio(for a given variance γ) is obtained by solving the problem (see [54])

$$\begin{aligned} &\max \quad \bar{r}^T\alpha \\ &\text{s.t.} \\ &\qquad (\alpha^T\mathbf{V}\alpha)^{1/2} = \gamma, \\ &\qquad \sum_{j=1}^{N}\alpha_j = 1, \end{aligned} \tag{2.50}$$

where the abbreviation "s.t." stands for "subject to"; this convention will be used in the rest of the document as well.

The derivation of the previous model assumes that all the available assets are Risky. It is also assumed that short-selling is allowed. What happens when the conditions are no longer the same? We describe below variants of model (2.50) and the conditions under which they can be used (see [48], [48], [54], [12]):

- Same assumptions as above, but short-selling is not allowed

$$\begin{aligned} &\max \quad \bar{r}^T\alpha \\ &\text{s.t.} \\ &\qquad (\alpha^T\mathbf{V}\alpha)^{1/2} = \gamma, \\ &\qquad \sum_{j=1}^{N}\alpha_j = 1, \\ &\qquad \alpha_j \geq 0, \qquad \forall j. \end{aligned} \tag{2.51}$$

- Short-selling and risk free assets are permitted

$$\max \quad \frac{\bar{r}^T\alpha}{(\alpha^T\mathbf{V}\alpha)^{1/2}}$$

s.t.

$$\sum_{j=1}^{N} \alpha_j = 1. \tag{2.52}$$

- No short sales but risk free assets are permitted

$$\begin{aligned} \max \quad & \frac{\bar{r}^T \alpha}{(\alpha^T \mathbf{V} \alpha)^{1/2}} \\ \text{s.t.} \quad & \\ & \sum_{j=1}^{N} \alpha_j = 1, \\ & \alpha_j \geq 0, \qquad \forall j. \end{aligned} \tag{2.53}$$

Remark 2.4 Several critiques have been formulated against these models; they are related to: 1- The unrealistic assumptions (e.g. normal distribution of returns) attached to their derivation. 2- Their single period nature. 3- Expensive computation resulting from the use of the variance-covariance matrix, $\mathbf{V}$. 4- These models very often result in "error-maximised" and "investment-irrelevant" portfolios (see [52]).

Modifications can, however, be made to the original formulations in order to improve their attractiveness. Indeed, index models (i.e. single index and multi-index models) can be used instead of the variance-covariance based models. One can also make use of alternative definitions of risk (e.g. semi-variance, mean absolute deviation) in order to capture the investor's perception of risk. The use of scenarios have been useful to extend single period model to multi-period models, and to reduce the "error-maximised" and "investment-irrelevant" tendency of the resulting portfolios.

Additional constraints may be used to fit the investor's circumstances.

2.6.2 Multi-period models

We consider the investments made from time $t = 1$ up to time $t = T$. At each time t an information is revealed to us. A scenario s is a description of a possible future state of nature. It includes the information uncovered at each time t. We assume that there are S scenarios. For each scenario s ($s = 1, \ldots, S$), let:

r_{stk} be the return earned on asset k at time t,

p_s be the probability of occurrence of scenario s,

e_s be the total value of the portfolio at time T,

w_{stk} be the decision variable representing the amount of cash allocated to asset k at time t,

$\bar{W}$ be the expected value of the investor's wealth at time T, and

e_{stk} be the contribution of the return on asset k to the value of the portfolio at time t. We have $e_{stk} = 1 + r_{stk}$.

If we assume that the initial wealth of the investor is 1 and that the portfolio allocation can be changed at the beginning of each time period, then the multi-period mean-variance problem can be formulated as follows:

$$\min \quad \sum_{s=1}^{S} p_s (e_s - \bar{W})^2 \tag{2.54}$$

s.t.

$$\sum_{k=1}^{N} w_{s,t+1,k} - \sum_{l=1}^{N} e_{stl} w_{stl} = 0, \qquad \forall t, s, \tag{2.55}$$

$$\sum_{k=1}^{N} e_{sTk} w_{sTk} = e_s, \tag{2.56}$$

$$\sum_{s=1}^{S} p_s e_s = \bar{W}, \tag{2.57}$$

$$\sum_{k=1}^{N} w_{s1k} = 1, \qquad \forall s, \tag{2.58}$$

$$w_{stk} \geq 0, \qquad \forall t, k, s. \tag{2.59}$$

- The objective of the investor is to minimise the variance of his final wealth; this is expressed by (2.54).
- Constraints (2.55) ensure the cash flow conservation; that is, the money that will be spent at time $(t+1)$ must be equal to the value of the portfolio at time t plus the gain resulting from the investment at time t.
- Constraint (2.56) determines the value of the portfolio at time T for scenario s.
- Constraint (2.57) gives the expected value of the investor's wealth at T.
- Constraint (2.58) is the time $t = 1$ budget constraint; it forces the initial investment to be equal to the initial wealth of the investor.
- Constraints (2.59) imply that short sales are not allowed.

Remark 2.5 This model is a stochastic programming problem. Solution techniques used to address problems of that nature are discussed in the remaining

chapters of the document.

The above model is slightly different from those discussed earlier: Here, we seek to minimise the variance of the portfolio's return. The advantage of the new form is that we do not have non-linear constraints. Therefore, the problem becomes easier to solve.

2.7 Beyond the classical mean-variance framework

One of the strongest limitations of the classical mean variance framework, beside the normality assumption, is the instability due to the non-uniqueness of the efficient frontier.

Indeed, given any point on the assumed unique frontier, there is a neighbourhood of the point (on and below the frontier) that includes an infinite number of statistically equivalent points. This fact was illustrated by Jobson and Korkie (see [42]). They used Monte Carlo methods and an historical set of means, standard deviations and correlations of monthly returns for 20 shares to simulate new sets of returns, and for each simulation (i.e. each new set of returns), a new efficient frontier was constructed. This experiment resulted in the construction of a collection of efficient frontiers located below the assumed unique frontier.

All the points located in the resampled region are efficient, and correspond to the statistical equivalence region's efficient portfolios. Two assets, with return $\tilde{r}_1$ and $\tilde{r}_2$, are said to be statistically equivalent if the returns have almost identical sample spaces (i.e. they can be superposed) when we take into account the estimation errors. A portfolio located within this region does not need to be rebalanced though from the point of view of the classical mean-variance framework it should.

Any of the following methods may be used to reduce the effect of the problem mentioned above (see [52]):

- statistical inference,
- the resampled efficient frontier,
- improved input estimation, and
- defining priors and benchmark.

Statistical inference

Before rebalancing a portfolio, one has to determine whether the portfolio needs revision. The portfolio manager has to find out if the portfolio (i.e. elements

of the portfolio) belongs to the statistical equivalence region; that is, if the portfolio is statistically equivalent to a portfolio located on the efficient frontier. If it belongs to the region, then the portfolio does not need to be rebalanced.

The resampled efficient frontier

This method, introduced by R. O. Michaud (see [51]), is inspired from the work of J. D. Jobson and B. Korkie (see [42], [43]). The resampled efficient portfolios can be obtained using the following procedure:

1- Find the monthly expected excess return for each individual asset, and the covariance matrix.

2- Simulate n monthly returns from a multivariate normal distribution with parameters obtained from the previous step.

3- Calculate the new monthly expected excess returns and covariance matrix from the simulated data.

4- Determine the optimal portfolio for a given risk target.

The steps are repeated N times, and N portfolios are obtained. The resampled efficient portfolio is obtained by taking the average of the weights of each asset in the N trials.

In practice the use of the resampled efficient frontier has deliver superior results (see, for example, Fletcher and Hillier [27]). They compare the January 1983 to May 2000 out-of-sample performance of the monthly mean-variance efficient strategy in international asset allocation and the monthly resampled efficient strategy. Their work investigates also the influence of different procedures for the estimation of the expected returns on the performance of the two methods. After an intensive examination, they conclude that the use of the resampled efficient frontier results in an increase of the Sharpe performance ratio and higher returns than those obtained using the traditional mean-variance framework.

These results are essentially due to the fact that the traditional approach considers the data as being a true representation of the state of nature in both the past and the future. Furthermore, the optimisation algorithm used to solve the resulting mathematical programming problem gives more weight to extreme points. Therefore, a small difference between the input data and the realisation of the random elements in the future will produce extremely bad results. When they are not counter-intuitive, they will recommend unnecessary changes of the portfolio's composition. This will result in an increase of the transaction's costs and a low performance.

The resampled efficient paradigm, however, explicitly takes into account the fact that the input data have estimation error. The data do not have to represent the future. R. O. Michaud and R. Michaud have developed this paradigm

around this unfortunate truth. As a matter of fact, they have used recent advances in statistics and this paradigm to design their patented online portfolio optimisation optimiser.

Improved input estimation

Another way to enhance the quality of the results obtained using the traditional framework is to improve the accuracy of the input. This is done by using Bayes-Stein estimators or share valuation models instead of historical data (see [52]:83-99).

Defining priors and benchmarks

The benchmark optimisation framework is based on residual return or the return of an asset minus the return of the benchmark (see [51]). The value of the procedure depends on the benchmark. This method will enhance the investment value for a good benchmark choice.

2.8 A framework for multi-period modelling

In the introduction, we mentioned that, the *portfolio management* process covers a given period of time. Within that time window, key parameters related to the value of the portfolio will quasi-continuously change; for example:

- The price of shares, $X(t)$, whose dynamics are given by the following stochastic differential equation

$$\mathrm{d}X(t) = \alpha(X(t),t)\,\mathrm{d}t + \sigma(X(t),t)\,\mathrm{d}Z(t), \tag{2.60}$$

  where:
  - $\alpha(X(t),t)$ is the instantaneous return on the asset at time t.
  - $\sigma(X(t),t)$ is the volatility of the price at time t.
  - $Z_t \sim \mathbf{N}(0,\sqrt{t})$.

- The time t price of a default-free zero-coupon bond maturing at time T using the Heath-Jarrow-Morton methodology (see [35]) is

$$P(t,T) = \exp\left(-\int_t^T f(t,s)\,\mathrm{d}s\right), \tag{2.61}$$

  with the forward rate, $f(t,T)$, following the multi-factor risk neutral process

$$\mathrm{d}f(t,T) = \mu(t,T,f(t,T))\,\mathrm{d}t + \sum_{i=1}^{n} \sigma_i(t,T,f(t,T))\,\mathrm{d}W_i(t), \tag{2.62}$$

  where σ_i is the volatility of the i-th factor, and μ is the drift of the process. The W_i's are independent brownian motion.

In order to avoid arbitrage, we must have the following relation between the drift and the volatility of the process:

$$\mu(t,T,f(t,T)) = \sum_{i=1}^{n} \sigma_i(t,T,f(t,T)) \int_t^T \sigma_i(t,u,f(t,u))\,du. \qquad (2.63)$$

This suggests the necessity of a multi-period framework; multi-period models are the appropriate tools for *portfolio management.* In what follows, we describe important elements related to multi-period models (see [46], [53], [64]).

2.8.1 Determination of periods and stages

The stages should coincide with dates at which we expect something to happen; e.g. liability repayment dates, portfolio expected revision and rebalancing dates. The time intervals between the stages are called periods. One should note that, the periods do not necessarily have the same length. When using a multi-period approach, one would like to make a decision for the first period taking into account what might happen in the future.

2.8.2 Forecasts of the parameters

The forecast of the parameters or scenario generation is a very important part of the portfolio management process. The quality of the solution obtained using the models will depend on the accuracy of that information.

A scenario is built using the stage-period structure; that is, a scenario is a sequence of joint outcomes of future parameters obtained for each stage t. The objective of the scenario generation process is to construct a realistic set of scenarios containing both optimistic and pessimistic cases. The portfolio manager may use outside sources or specially designed software to determine these parameters.

In general, the scenario generation process can be described as follows:

- Collect the historical data from the relevant time series.
- Use the data to find a suitable model for the dynamics of the asset under consideration. One may also make use of the opinion of an expert.
- Use the model to generate the scenarios.

2.8.3 The objective function

As mentioned in the introduction, the objective of the portfolio management process is to maximise the expected terminal wealth or its utility. Note, however, that the investors do not need to agree on the definition of the word "wealth".

2.8.4 The constraints

Although the constraints will vary from one investor to the next one, there are some very common ones:

- bounds and nonnegativity constraints on assets allocation,
- budget constraints at all stages,
- description of the Asset accumulation and revision process,
- wealth definition,
- restrictions on asset class allocation, and
- cardinality constraints.

2.9 Strategies for asset allocation

As mentioned earlier, key parameters related to the value of the assets fluctuate over time. These fluctuations will induce changes in the value and the composition of the portfolio. Indeed, if the risky assets increase in value, the proportion of the portfolio they represent is likely to increase as well. The resulting increase in the proportion of risky assets will ,of course, increase the exposure of the portfolio to risk. Therefore, one must devise explicit rules or strategies for rebalancing the portfolio in response to such changes. We present 3 such strategies:

- buy–and–hold,
- constant-mix, and
- constant–proportion.

2.9.1 Buy-and-hold strategy

When using a buy-and-hold strategy, an initial mix is determined, e.g. 60/40 shares/bills, assuming that the portfolio contains only these two assets classes. Assets from both classes are bought according to the given proportions at the beginning of the planning period and then held. No matter what happens to relative values, no rebalancing is required. Such a strategy corresponds to investors with zero tolerance to risk when the value of the assets has reached the amount invested in bills.

The value of the portfolio changes linearly with respect to the value of the share. Given the initial mix, every dollar of additional share value increases the value of the portfolio by 60 cents. The value of the portfolio will never fall below the value of the initial investment in bills (this value is referred to as : floor). There is a downside protection while the upside potential is unlimited.

The performance of this strategy is directly proportional to the initial mix; that is, the greater the initial percentage invested in bills, the better the performance of the strategy when bills outperform shares, and the worse the performance when shares outperform bills.

2.9.2 Constant-mix

Unlike the buy-and-hold strategy, the constant-mix strategy is dynamic. The idea is to maintain an exposure to shares that is a constant proportion of wealth; that is, shares are held at all wealth levels. Therefore, a frequent portfolio rebalancing is required. This strategy is consistent with an investor whose risk attitude varies proportionally with her wealth.

A decrease in the value of shares will result in a decrease of the proportion of the portfolio value invested in shares. This change in the mix of the portfolio will necessitate a rebalancing of the portfolio; i.e. more shares have to be bought to keep the mix at the required level. Similarly, an increase in the value of shares will increase the proportion of the portfolio value invested in shares. In this case, rebalancing to constant-mix requires the sale of shares and a purchase of bills. This strategy buys shares as their price falls, and sells shares as their price rises.

2.9.3 Constant-proportion portfolio insurance

The amount of money invested in shares when using this strategy is

$$\text{Money invested in shares} = \alpha\cdot(\text{total value of portfolio - floor}),$$

where α is a fixed multiplier.

Here, the floor should be understood as earlier; it is arbitrarily chosen by the investor, but it is expected that

$$\text{floor} < \text{total value of the portfolio.}$$

Several variants of this strategy may be obtained by changing the value of α. For example, when $\alpha > 1$ we have the constant-proportion portfolio insurance (CPPI). The strategy, here, is to keep the exposure to equities a constant multiple of the cushion; i.e.

$$\text{cushion} = \text{total value of the portfolio at time } t \text{ - floor at time } t.$$

This is achieved by selling shares as their price falls and buying shares as their price rises. Note, however, that it can be difficult to implement this strategy during periods of high volatility. This was the case during the 1987 US equity crash, for example.
The value of the portfolio will never fall below the floor except if the market

falls strongly before the rebalancing takes place. This strategy will do very well in a bull market and poorly in a flat market. The constant-proportion strategy can accommodate different risk's profiles. This is done by adjusting the value of the multiplier, α, to a suitable level.

Remark 2.6 In general, a constant-mix strategy will outperform a comparable (i.e. having identical mix) buy-and-hold strategy if markets are characterised by reversals rather than by trends (see [55]).

Buy-and-hold strategies are constant-proportion strategies with $\alpha = 1$ and a floor equal to the amount of money invested in bills.

Constant-mix strategies are special cases of constant-proportion strategies with the floor equal to zero and $0 < \alpha < 1$.

More about dynamic strategies can be found in Perold (see [55]) and Arnott (see [2]).

Chapter 3

Two-stage stochastic programming

In the previous chapter we introduced the portfolio management problem. We presented the rationale behind the investor's decisions; that is, she will choose the assets that will maximise the expected utility of her final wealth. From the discussion related to the dynamics of the parameters determining the value of the portfolio (e.g. the interest rate and the return), it became clear that this problem should be formulated in a multi-period framework. It was also established that stochastic programming was the appropriate tool.

We put those ideas to work in the following section. They are applied to the case of an investor who has to manage a portfolio of bonds.

3.1 The bond portfolio management problem

We illustrate this problem with an example inspired by Golub, Holmer, McKendall and Pohlman (see [34], [21]). Consider a planning period starting now, time $t = 0$, and ending at time T. The composition of the portfolio can be changed at any time t ($t \leq T$). The uncertain future is represented by a set of scenarios S, and each element $s \in S$ has probability of occurrence p_s. The objective of the portfolio manager is to maximise the expected utility of the wealth at time T. She must also make sure that the future liability payments are made. This is done in two stages:

1. She makes an initial decision about the composition of her portfolio; that is, she decides which of her bonds (from her initial portfolio) to sell and in what quantity (i.e. for what value). She must also decide which of the available bonds (available in the market) to buy and in what quantity. There is no restriction on which bond to buy or to sell. She chooses the investment that will maximise the expected utility value of her final wealth

under the set of constraints. Note that her utility function is $U(\cdot)$.

2. She compares the total cash flow to the value of the liabilities. Depending on whether the total cash flow allows the payment of the liabilities or not, she will have to make additional changes to the portfolio holdings.

In order to formulate this problem, we consider:

- b_0: the initial holding (in face value) of the risk free asset.
- b_j: the initial holdings of bond j ($j = 1, \ldots, J$).
- r_t^s: the short term interest rate during the period $(t, t+1)$ under scenario s.
- f_{jt}^s: the cash flow generated from bond j at time t under scenario s expressed as a fraction of the face value.
- ξ_{jt}^s: the selling price of bond j at time t under scenario s. This price corresponds to the actual price of the bond (i.e. sum of the discounted cash flows) minus the transaction cost. The price at time $t = 0$, ξ_{j0}, is not scenario dependent.
- ζ_{jt}^s: the purchasing price of bond j at time t under scenario s. This price corresponds to the actual price of the bond (i.e. sum of the discounted cash flows) and the transaction cost. The price at time $t = 0$, ζ_{j0}, is not scenario dependent.
- L_t: the liability due at time t.

Next, we consider the following first-stage variables:

- x_j: the face value of bond j purchased at the beginning of the planning period. This variable is assumed to be a non-negative real number.
- y_j: the face value of bond j sold at the beginning of the planning period. This is also a non-negative real number.
- z_{j0}: the face value of bond j held in the portfolio after the initial decisions, x_j and y_j, have been made.
- y_0^+: the initial surplus. It is also a non-negative real number.

We also consider the second-stage variables:

- x_{jt}^s: the face value of bond j purchased at time t under scenario s.
- y_{jt}^s: the face value of bond j sold at time t under scenario s.
- z_{jt}^s: the face value of bond j held after bonds have been bought and sold at time t under scenario s.
- y_t^{-s}: the amount borrowed at time t under scenario s.

- y_t^{+s}: the amount invested in the risk free asset (lending) at time t under scenario s.
- W_T^s: the total wealth at time T under scenario s.

The resulting two-stage stochastic problem is

$$\max \quad \sum_{s \in S} p_s \cdot U(W_T^s), \tag{3.1}$$

under the following constraints:

1. The first-stage constraints:
   - The conservation of the initial holdings; that is, the face value of bond j sold plus the value of the bond left, after the implementation of the initial decision, is equal to the original holdings of the bond plus the value of the bond purchased.

$$y_j + z_{j0} = b_j + x_j, \qquad \forall j. \tag{3.2}$$

   - The cash flow conservation constraint

$$y_0^+ + \sum_{j=1}^{J} \zeta_{j0} x_j = b_0 + \sum_{j=1}^{J} \xi_{j0} y_j. \tag{3.3}$$

2. The second-stage constraints:
   - The conservation of holdings for every bond j at each time t and under each scenario s

$$z_{jt}^s + y_{jt}^s = z_{j,t-1}^s + x_{jt}^s, \qquad \forall j, t, s. \tag{3.4}$$

   - The cash flow conservation constraints at each time t under every scenario s. These constraints take into account the cash generated by buy and sell decisions, borrowing and reinvestment decisions as well as liability payments.

$$\sum_{j=1}^{J} \xi_{jt}^s y_{jt}^s + \sum_{j=1}^{J} f_{jt}^s z_{j,t-1}^s + (1 + r_{t-1}^s) y_{t-1}^{+s} + y_t^{-s} =$$
$$L_t + \sum_{j=1}^{J} \zeta_{jt}^s x_{jt}^s + (1 + r_{t-1}^s) y_{t-1}^{-s} + y_t^{+s}, \qquad \forall s, t. \tag{3.5}$$

   - The final wealth is obtained by valuing the holding of the portfolio at time T and paying off the outstanding debt.

$$W_T^s = \sum_{j=1}^{J} \xi_{jT}^s z_{jT}^s + y_T^{+s} - y_T^{-s}, \qquad \forall s. \tag{3.6}$$

3. The non-negativity constraints of all the variables. Additional constraints have to be introduced; they are: $y_0^{-s} = 0, \ \forall s$, and $y_0^{+s} = y_0^+ \ , \ \forall s$.

The problem formulated above is a two-stage stochastic programming problem. The next section gives a formal definition of this type of problems.

3.2 Definition of the problem

In order to give a formal definition of the two-stage problem, let us consider a decision maker faced with the following problem:

- At time t_1 she has to make a decision, x_1, assumed to be a vector whose components are non-negative real numbers. She makes the decision in order to minimise a given objective function, $f_1(\cdot)$ under a set of constraints. The function $f_1(\cdot)$ is the cost associated with the decision x_1. The decision is made without any concern about the future; that is, $f_1(\cdot)$ is not a random variable, and the constraints are deterministic. The information that will be revealed is represented by the random vector ξ.
- At time t_2 she observes the realisation of an outcome, $\xi(\omega)$ of the random vector ξ. The outcome observed may force the problem to lose its feasibility. In order to adjust to any discrepancies caused by the outcome of the random vector, she is allowed to make a further decision x_2. This decision is made under a new set of constraints. The new decision is also assumed to be a vector whose components are non-negative real numbers. The constraints are described by the elements of the vector ξ. The corrective action is made at an additional cost, $Q(x_1, \xi)$.

In other words, the problem is to find x_1 and x_2 in order to minimise the expected total cost. These decisions are made under two sets of constraints (i.e. there is a set of constraints for each time period).This process may be translated in the following decision-observation scheme:

- **at time t_1: make a decision x_1**
- **at time t_2: observe an outcome $\xi(\omega)$ and make a 2^{nd} decision x_2**

The decision problem above can be formulated as a stochastic programming problem with recourse; that is,

$$\begin{array}{lrcl} \min & f_1(x_1) & + & E[Q(x_1, \xi)] \\ \text{s.t.} & & & \\ & A_{11}x_1 & = & b_1, \\ & x_1 & \geq & 0, \end{array} \tag{3.7}$$

with $Q(x_1, \xi(\omega))$, also called the recourse function, defined as follows:

$$Q(x_1, \xi(\omega)) = \min_{x_2} \quad f_2(x_2(\omega))$$

$$\begin{aligned} &\text{s.t.} \\ &A_{21}(\omega)x_1 + A_{22}(\omega)x_2(\omega) = b_2(\omega), \\ &x_2(\omega) \geq 0, \end{aligned} \tag{3.8}$$

where, $x_1 \in R^{n_1}$, $x_2 \in R^{n_2}$, $A_{11} \in R^{m_1 * n_1}$, $A_{21} \in R^{m_2 * n_1}$, $A_{22} \in R^{m_2 * n_2}$ $b_1 \in R^{m_1}$, $b_2 \in R^{m_2}$, $f_1 : R^{n_1} \to R$, and $f_2 : R^{n_2} \to R$. The random vector ξ maps the probability space space (Ω, P) on to (Ξ, F) with F the distribution function and $\Xi \subseteq R^m$, and we have

$$\xi(\omega) = (f_2(\omega), a_{21}^{11}(\omega), \ldots, a_{21}^{m_2,n_1}(\omega), a_{22}^{11}(\omega), \ldots, a_{22}^{m_2,n_2}(\omega), b_2^1(\omega), \ldots, b_2^{m_2}(\omega))^T.$$

That is, $\xi(\omega)$ is a vector whose components are the components of $A_{21}(\omega)$, $A_{22}(\omega)$, $b_2(\omega)$ and elements describing the function $f_2(\cdot)$ (if $f_2(\cdot)$ is linear, then these elements are the components of a vector $f \in R^{n_2}$). We will assume that $f_2(\cdot)$ has nice properties; that is, for example, it takes only finite values.

Note that x_1 is used as a parameter in problem (3.8); this problem is referred to as second stage problem or *recourse problem.*

An alternative form for problem (3.7) is

$$\begin{aligned} &\min \quad f_1(x_1) + E[\min_{x_2} \quad \mathbf{f_2}(x_2)] \\ &\text{s.t.} \\ &\qquad A_{11}x_1 = b_1, \\ &\mathbf{A_{21}}x_1 + \mathbf{A_{22}}x_2 = \mathbf{b_2}, \qquad a.s. \\ &\mathbf{x_2} \geq 0 \quad a.s., \quad \text{and} \quad x_1 \geq 0, \end{aligned} \tag{3.9}$$

where $\mathbf{f_2}$, $\mathbf{A_{21}}$, $\mathbf{A_{21}}$ and $\mathbf{b_2}$ have random components; these are given by the components of ξ as described above;
that is,

$$\begin{aligned} &\min \quad f_1(x_1) + \int_\Omega Q(x_1, \xi(\omega))P(d\omega) \\ &\text{s.t.} \\ &\qquad A_{11}x_1 = b_1, \\ &\qquad x_1 \geq 0. \end{aligned} \tag{3.10}$$

Assuming that Ω is a finite sample space and that the random elements take discrete values, i.e. $\Omega = \{\omega^1, \omega^2, \ldots, \omega^K\}$, problem (3.7) becomes

$$\begin{aligned} &\min \quad f_1(x_1) + \sum_{j=1}^{K} Q(x_1, \xi(\omega^j))P(\omega^j) \\ &\text{s.t.} \\ &\qquad A_{11}x_1 = b_1, \\ &\qquad x_1 \geq 0, \end{aligned} \tag{3.11}$$

where $P(\omega^j)$ is the probability associated with the realised outcome ω^j, and for all j, we have

$$\begin{aligned} Q(x_1, \xi(\omega^j)) = \min_{x_2} \quad & f_2(x_2(\omega^j)) \\ \text{s.t.} \quad & \\ & A_{21}(\omega^j)x_1 + A_{22}(\omega^j)x_2(\omega^j) = b_2(\omega^j), \quad (3.12) \\ & x_2(\omega^j) \geq 0. \end{aligned}$$

3.3 Properties of recourse problems

We will assume that the function $f_2(\cdot, \omega) \equiv f_2(x_2(\omega))$ is continuous for any fixed ω, and measurable with respect to (Ω, P) for any fixed first argument. Given these assumptions, $Q(x_1, \xi(\omega))$ is also measurable with respect to (Ω, P).

Before to review the different properties of the recourse problem, let us give some definitions:

- When the recourse matrix A_{22}, is known with certainty, the second problem is said to have *fixed recourse*; otherwise, it is said to have *random recourse.*
- The two-stage stochastic problem is said to have *simple recourse* when $A_{22} = [I, -I]$ (I is the identity matrix with m_2 rows such $x_2 = (x_2^+, x_2^-)^T$), $b_2 = \omega$ and, A_{21} is not random.
- The stochastic programming problem is said to have *complete recourse* if $Q(x_1, \xi) < +\infty$ with probability 1 for all $x_1 \in R^{n_1}$.
- When $Q(x_1, \xi) < +\infty$ is true only for $x_1 \in X_1 = \{x_1 \in R^{n_1} : A_{11}x_1 = b_1, x_1 \geq 0\}$, the problem is said to have *relative recourse.*

Now let, $X_2 = \{x_2 \in R^{n_2} : Q(x_1, \xi) < +\infty \quad \text{with} \quad \text{probability} \quad 1\}$.
The sets X_1, X_2 and their intersection $X = X_1 \cap X_2$, are convex and closed. Assuming that the recourse function is bounded from below, that is $Q(x_1, \xi(\omega)) > -\infty$, it has the following properties (see [25], [29], [56], [62]):

- For fixed $x_1 \in R^{n_1}$ and $f_2 \in R^{n_2}$, the function $\xi(\omega) \rightarrow Q(x_1, \xi(\omega))$ is convex and piecewise linear.
- For fixed $x_1 \in R^{n_1}$, $A_{21} \in R^{m_2 * n_2}$ and $b_2 \in R^{m_2}$, the function $f_2 \rightarrow Q(x_1, \xi(\omega))$ is concave and piecewise linear.
- For fixed $\xi(\omega)$, the function $x_1 \rightarrow Q(x_1, \xi(\omega))$ is convex and piecewise linear on X.
- If ξ has finite second moment then the function $Q(x_1) = \int_\Omega Q(x_1, \xi(\omega))P(d\omega)$ is finite and convex on X.

In order to ensure that our problem is well defined, it is preferable to assume that ξ has finite second moment.

In the first part of this chapter, we have given a formal definition of the two-stage stochastic programming problem. We have also reviewed its properties. In the remaining half of the chapter, we will discuss some of the solution techniques used to solve this class of problems.

3.4 L-shaped decomposition method

Let us consider the following stochastic programming problem with recourse:

$$\begin{aligned} \min \quad & f_1(x_1) + \sum_{j=1}^{K} Q(x_1, \xi(\omega^j))P(\omega^j) \\ \text{s.t.} \quad & \\ & A_{11}x_1 = b_1, \\ & x_1 \geq 0, \end{aligned} \tag{3.13}$$

and for all j, we have

$$\begin{aligned} Q(x_1, \xi(\omega^j)) = \min_{x_2} \quad & f_2(x_2(\omega^j)) \\ \text{s.t.} \quad & \\ & A_{21}(\omega^j)x_1 + A_{22}(\omega^j)x_2(\omega^j) = b_2(\omega^j), \\ & x_2(\omega^j) \geq 0. \end{aligned} \tag{3.14}$$

We will assume that the functions $f_1(\cdot)$ and $f_2(\cdot)$ are both linear. We further assume that X_1 is bounded and that problem (3.13) is feasible; that is, the problem, where $\sum_{j=1}^{K} Q(x_1, \xi(\omega^j))P(\omega^j) = 0$, has an optimal solution $\bar{x}$. Furthermore, there exists $x_2(\omega^j) \geq 0$ such that

$$A_{21}(\omega^j)\bar{x} + A_{22}(\omega^j)x_2(\omega^j) = b_2(\omega^j), \quad \forall j. \tag{3.15}$$

We will also assume that the recourse matrix A_{22} is known with certainty; i.e. the second stage problem has fixed recourse.
Problem (3.13) becomes

$$\begin{aligned} \min \quad & f_1^T x_1 + \sum_{j=1}^{K} f^T x_2(\omega^j)P(\omega^j) \\ \text{s.t.} \quad & \\ & A_{11}x_1 = b_1, \\ & A_{21}(\omega^j)x_1 + A_{22}x_2(\omega^j) = b_2(\omega^j), \quad \forall j, \\ & x_2(\omega^j) \geq 0, \quad \forall j. \end{aligned} \tag{3.16}$$

Problem (3.16) is a linear programming problem, and can be solved using straight-forward methods (e.g. simplex, interior point). Depending on K, it may be enormous and difficult to solve. However, a special structure exhibited by the recourse problem can be exploited to solve problem (3.16); that is,

$$\begin{aligned} \min \quad Q(x_1) &= \sum_{j=1}^{K} f^T x_2(\omega^j) P(\omega^j) \\ \text{s.t.} \\ A_{22} x_2(\omega^j) &= b_2(\omega^j) - A_{21}(\omega^j) x_1, \quad \forall j, \\ x_2(\omega^j) &\geq 0, \quad \forall j. \end{aligned} \tag{3.17}$$

Since x_1 is a parameter in the above problem, the right-hand side of the constraints (i.e. $[b_2(\omega^j) - A_{21}(\omega^j)x_1]$) is constant for all j; furthermore, the matrix resulting from the left-hand side of the constraints is diagonal with an identical element (i.e. A_{22} which is repeated for all j).

The feasibility assumption of problem (3.13) implicitly assumes that problem (3.17) is feasible as well. This is true since the $x_2(\omega^j)$'s which satisfy the system of equations (3.15) are the feasible solution of problem (3.17).

The main idea behind the L-shaped method (see [40], [44], [61]) is to replace the expected future cost in (3.13) by a scalar, say Θ, whose value is subsequently restricted by additional constraints. Hence, problem (3.13) becomes

$$\begin{aligned} \min & f_1^T x_1 + \Theta \\ \text{s.t.} \\ & \Theta \geq Q(x_1), \\ & x_1 \in X_1, \end{aligned} \tag{3.18}$$

where Θ is a real number.

In order to solve problem (3.18), the L-shaped method operates by dropping the constraint $\Theta \geq Q(x_1)$ and sequentially adding new constraints; these new constraints are of 2 types: feasibility and optimality constraints.

Let us now explain how the method works.
For a given optimal solution of problem (3.18), $\bar{x}$, $Q(\bar{x})$ can be found by solving problem (3.17) or its dual

$$\begin{aligned} \max \quad & \sum_{j=1}^{K} p_j \pi_j^T [b_2(\omega^j) - A_{21}(\omega^j)\bar{x}] \\ \text{s.t.} \\ & \pi_j^T A_{22} \leq f, \quad \forall j, \end{aligned} \tag{3.19}$$

where, for every j, $\pi_j \in R^{n_2}$ and has unrestricted components. Note that the dual problem can be decomposed into K smaller problems of the form:

$$\begin{aligned} \max \quad & p_j \pi_j^T [b_2(\omega^j) - A_{21}(\omega^j)\bar{x}] \\ \text{s.t.} \quad & \\ & \pi_j^T A_{22} \leq f, \end{aligned} \tag{3.20}$$

for every j.

This method, therefore, splits the original problem (i.e. problem (3.13)) into a master problem, problem (3.18) and a subproblem, which in turn decomposed into K subproblems (i.e. problem (3.20)) corresponding each to an outcome of the random vector.

By solving the master problem where cuts are initially absent, and later on, sequentially added, a trial solution, $\bar{x}$, is found. This solution is then used as a parameter in problem (3.20)

3.4.1 The feasibility cuts

If problem (3.17) was infeasible, its dual, problem (3.19) would be unbounded; hence, there would exist, at least, one feasible direction $\bar{\sigma}$ such that

$$\begin{aligned} \bar{\sigma}^T [b_2(\omega^j) - A_{21}(\omega^j)\bar{x}] &> 0, \\ \textit{and} & \\ \bar{\sigma}^T A_{22} &\leq 0. \end{aligned} \tag{3.21}$$

Hence, to restore feasibility of the primal problem (3.17), a constraint must be added to the dual problem (3.19) that will cut off the infeasible part of X_1. The new constraint, the feasibility cut, is

$$\bar{\sigma}^T [b_2(\omega^j) - A_{21}(\omega^j)x_1] \leq 0. \tag{3.22}$$

This constraint must hold for every $x_1 \neq \bar{x}$. The direction $\bar{\sigma}$ is found by solving the following problem

$$\max\{\sigma^T [b_2(\omega^j) - A_{21}(\omega^j)\bar{x}] : \sigma^T A_{22} \leq 0, \|\sigma\| \leq 1\}. \tag{3.23}$$

The constraint $\|\sigma\| \leq 1$ has been introduced to bound the maximal value of the objective function.
Since σ is unrestricted, we replace it by $\sigma^+ - \sigma^-$, where $\sigma^+, \sigma^- \geq 0$; the problem (3.23) then becomes

$$\begin{aligned} \max \quad & (\sigma^+ - \sigma^-)^T [b_2(\omega^j) - A_{21}(\omega^j)\bar{x}] \\ \text{s.t.} \quad & \\ & A_{22}\sigma^+ - A_{22}\sigma^- \leq 0, \\ & e^T\sigma^+ + e^T\sigma^- \leq 1, \\ & \sigma^+, \sigma^- \geq 0, \end{aligned} \tag{3.24}$$

where e is a vector of ones and $e \in R^{n_2}$, and $\bar{x}$ is a paramter.
The dual of problem (3.24) is

$$\begin{aligned} &\min \quad t \\ &\text{s.t.} \\ &\qquad \begin{aligned} A_{22}y + e^T t &\geq b_2(\omega^j) - A_{21}(\omega^j)\bar{x}, \\ -A_{22}y + e^T t &\geq -(b_2(\omega^j) - A_{21}(\omega^j)\bar{x}), \\ y, t &\geq 0, \end{aligned} \end{aligned} \tag{3.25}$$

where y, t are the dual variables.

If $t = 0$, there exists a $y \geq 0$ satisfying

$$\begin{aligned} A_{22}y &= b_2(\omega^j) - A_{21}(\omega^j)\bar{x} \\ &= 0. \end{aligned} \tag{3.26}$$

Hence, the j-th dual subproblem (3.20) has a feasible solution.
If the optimal solution of (3.25) is positive, i.e $t > 0$, the j-th dual subproblem (3.20) is unbounded; we can deduce the value of $\bar{\sigma}$, and we have

$$\bar{\sigma}^T[b_2(\omega^j) - A_{21}(\omega^j)\bar{x}] > 0. \tag{3.27}$$

Hence, we can construct the new feasibility cut:

$$\bar{\sigma}^T[b_2(\omega^j) - A_{21}(\omega^j)x] \leq 0. \tag{3.28}$$

3.4.2 The optimality cuts

If on the other hand, the dual subproblem (3.20) has an optimal solution (we call $\bar{\pi}_j$ the optimal solution of the j-th dual subproblem) its primal problem is bounded. Since its feasibility was assumed, from the theory of linear programming (see the weak duality lemma in, for example, [47] : 89), we have

$$Q(x_1) \geq \sum_{j=1}^{K} p_j \bar{\pi}_j^T [b_2(\omega^j) - A_{21}(\omega^j)\bar{x}], \tag{3.29}$$

for every x_1, with equality holding for $x_1 = \bar{x}$.

The above inequality is an optimality cut. Note that for both types of cuts, x_1 is a parameter.

After N iterations of the L-shaped algorithm, either we have, at each stage, obtained an infeasible master problem, and the number of feasible cuts is $V = N$, or we have experienced a few cases of feasibility. In the second situation, the number of optimality cuts is $J = N - V$.

3.4.3 The L-shaped decomposition: Algorithm

The ideas presented in the previous sections can be implemented using the following algorithm:

step 0:
Initialize $J = V = \nu = 0$.

step 1:
Set $\nu = \nu + 1$ and solve the master problem:

$$\begin{aligned} \min \quad & f_1^T x_1 + \Theta \\ \text{s.t.} \quad & \\ & A_{11} x_1 = b_1, \\ & x_1 \geq 0, \Theta \in R. \end{aligned} \tag{3.30}$$

If infeasible stop.

Let (x^ν, Θ^ν) be the optimal solution of the master problem. For $\nu = 0$, we set $\Theta = 0$ and $\Theta^\nu = -\infty$.

step 2:
For $j = 1, \ldots, K$, solve subproblem (3.25).

If for some j, we have $t > 0$, then set $V = V + 1$ and let $\sigma_V = \bar{\sigma}$,

$$D_V = \sigma_V^T A_{21}(\omega^j), \text{ and} \tag{3.31}$$
$$d_V = \sigma_V^T b_2^j(\omega^j). \tag{3.32}$$

We then add a feasibility cut, i.e. constraint of type: $D_s x_1 \geq d_s$ to the master problem.
Return to **step1**

Else
Let Opt_j be the optimal value of the objective function the dual subproblem (3.20), and π_J its optimal solution,

If $\Theta^\nu \geq \sum_{j=1}^K Opt_j$, stop: x^ν is optimal.
Set $J = J + 1$ and define

$$E_J = \sum_{j=1}^K p_j \pi_J^T A_{21}(\omega^j), \text{ and} \tag{3.33}$$
$$e_J = \sum_{j=1}^K p_j \pi_J^T b_2^j(\omega^j). \tag{3.34}$$

If $\Theta^\nu = \sum_{j=1}^K Opt_j$, then (x^ν, Θ^ν) is the optimal solution of the master problem.

Otherwise we add a new constraint of type $E_s x_1 + \Theta \geq e_s$, i.e. an optimality cut, and return to **step1**.

Remark 3.1 If the dual subproblem (3.20) is feasible, and the set X_1 is bounded, the method will terminate after a finite number of iterations.

The use of the L-shaped method implies having to deal with an enormous amount of work at each iteration. When only the right-hand side of the linear programming problems is random, we have to solve many similar LPs; this problem (i.e. solving a large number of similar LPs) has attracted a lot of attention, as a result, several methods have been introduced (e.g. "bunching", see [63]).

It was assumed that problem (3.13) is bounded; to guarantee its boundedness, we can just add bounds on the decision variables.

3.5 Approximation of the probability function

Consider the problem

$$\begin{array}{lrcl} \min & f(x_1) & + & E[Q(x_1,\xi)] \\ \text{s.t.} & & & \\ & A_{11}x_1 & = & b_1, \\ & x & \geq & 0, \end{array} \tag{3.35}$$

where $x \in R^n$, $A_{11} \in R^{m1*n1}$, $b_1 \in R^{m1}$ and $f : R^{n1} \to R$. The vector ξ is defined as earlier. However, we will assume that only the function $\mathbf{f_2}$ is random. Note that, we have that the expected second stage cost $Q(x)$ can be expressed as follows:

$$\begin{aligned} Q(x) &= E[Q(x,\xi)] \\ &= \int_\Omega Q(x,\xi(\omega))P(d\omega) \\ &= \int\int\ldots\int Q(x,\xi(\omega))\phi(\omega)d\omega_1 d\omega_2 \ldots d\omega_M, \end{aligned} \tag{3.36}$$

where $\phi(\omega)$ is the density function of the random vector $\omega = (\omega_1,\ldots,\omega_2)^T$. Note that we have dropped the subscript of x in the expression of the cost; we will do the same in the remaining part of the chapter.

In other words, in order to solve problem (3.35), we need to calculate the multiple integral (3.36) or a multiple sum; this is a very difficult task.

A way out (see [7], [25]) is to partition Ξ (or Ω) into a finite number of disjoint subsets $\Xi_1,\ldots\Xi_L$ with probabilities $p_1 = p\{\xi_1 \in \Xi_1\},\ldots,p_L = p\{\xi_L \in \Xi_L\}$

respectively. Using this new probability measure (say, P^ν), the expectation is approximated by

$$Q^\nu(x) = \sum_{l=1}^{L} p_l Q(x, \xi_l). \tag{3.37}$$

However, if a good and efficient method for finding P^ν is not used, $Q^\nu(x)$ will be a good approximation to $Q(x)$ only if $L \to +\infty$; i.e., the multiple integral or multiple sum is replaced by an infinite sum. In other words, we are still in trouble. A good approximation scheme should include:

- a proper way for replacing the original random vector with a discrete one,
- a way to estimate the accuracy of the approximation, and
- a technique, if needed, of improving the accuracy of the approximation.

The following convergence results are the basis for the design of such schemes:

Consider $\{P^\nu, \nu = 1, \ldots\}$, a sequence of probability measures converging in distribution to P, and suppose that for $x \in X = \{x : A_1 x = b_1, x \geq 0\}$ the function $Q(x, \xi)$ is uniformly integrable with respect to P^ν; furthermore, we suppose that there exists a bounded set D such that

$$D \cap \operatorname{argmin}\Big[Q^\nu(x) = \int_{\Xi} Q(x, \xi) P^\nu(d\xi) : x \in X\Big] \neq \emptyset, \quad \text{for almost all} \quad \nu,$$

then

$$\inf_X Q = \lim_{\nu \to \infty} (\inf_X Q^\nu), \tag{3.38}$$

and

$$\text{if} \quad x^\nu \in \operatorname{argmin}_X Q^\nu, \quad \text{and} \quad x = \lim_{k \to \infty} x^{\nu_k}, \tag{3.39}$$

then

$$x \in \operatorname{argmin}_X Q \max. \tag{3.40}$$

These results suggest that, P^ν can be chosen such that

$$\inf_X Q^\nu \leq \inf_X Q, \tag{3.41}$$

and $P^{\nu+1}$ such that

$$\inf_X Q \leq \inf_X Q^{\nu+1}. \tag{3.42}$$

An easy way to find such probabilities is to exploit the properties of the function $Q(x, .)$. For example, if $Q(x, .)$ is convex, we can use the Jensen inequality to construct probability measures that yield lower approximates of Q.

3.5.1 Lower bound: The Jensen inequality

If a function $g(\tilde{y})$ is convex over the support of the random variable $\tilde{y}$, then, by the Jensen inequality, we have

$$E[g(\tilde{y})] \geq g(E[\tilde{y}]). \tag{3.43}$$

In other words, for convex function, the Jensen inequality can be used to obtain lower bounds. The direction of the inequality is reversed for a concave function. We know from the discussion of the properties of problems with recourse that the function $\xi(\omega) \to Q(x, \xi(\omega))$ is convex for fixed x. Hence, by Jensen's inequality, we have

$$\begin{aligned} Q(x, E[\xi]) &= Q(x, \int_{\Omega} \xi(\omega) P(d\omega) \\ &\leq E[Q(x, \xi)] \\ &= \int_{\Omega} Q(x, \xi(\omega)) P(d\omega). \end{aligned} \tag{3.44}$$

The previous bound can be improved by using Ξ_ℓ $(\ell = 1, \ldots, L)$, a partition of Ξ.

Let $\xi^\ell = E[\xi|\Xi_\ell]$ and $p_\ell = p(\xi \in \Xi_\ell)$ so that $Q(x, E[\xi|\Xi_\ell]) = Q(x, \xi^\ell)$.
Then applying inequality (3.44) to each Ξ_ℓ and summing over Ξ, we get

$$E[Q(x, \xi)] \geq \sum_{\ell} p_\ell Q(x, \xi^\ell). \tag{3.45}$$

3.5.2 The Edmundson-Madansky inequality

The main idea here is to find a new probability measure, $\hat{p}$, with support ext $\Xi = \{\xi^1, \xi^2, \ldots, \xi^J\}$ (the set of extreme points of the convex hull of Ξ) such that

$$\sum_{j=1}^{J} \hat{p}_j(\xi) = 1,$$

and

$$\sum_{j=1}^{J} \xi^j \hat{p}_j(\xi) = \xi.$$

Using this idea for our recourse function, we obtain

$$Q(x, \xi) = Q(x, \sum_{j=1}^{J} \xi^j \hat{p}_j(\xi)). \tag{3.46}$$

Since the function $Q(x, .)$ is convex, the following inequality holds

$$Q(x, \xi) \leq \sum_{j=1}^{J} \hat{p}_j(\xi) Q(x, \xi^j). \tag{3.47}$$

Taking the expectation on both sides, we obtain

$$\begin{aligned} Q(x) &\leq \int_{\Xi} \sum_{j=1}^{J} \hat{p}_j(\xi) Q(x, \xi^j) P(d\xi) \\ &= \sum_{j=1}^{J} \Big[\int_{\Xi} \hat{p}_j(\xi) Q(x, \xi^j) P(d\xi) \Big] \\ &= \sum_{j=1}^{J} \bar{p}_j Q(x, \xi^j), \end{aligned} \tag{3.48}$$

where the $\bar{p}_j$'s represent a discrete probability function on ext Ξ, and we have $\bar{p}_j = \int_{\Xi} \hat{p}_j(\xi) P(d\xi)$.

To find the best possible approximation to $Q(x)$, we solve the following problem

$$\min_{\bar{p}_j} \quad \sum_{j=1}^{J} \bar{p}_j Q(x, \xi^j)$$

s.t.

$$\begin{aligned} \sum_{j=1}^{J} \xi^j \bar{p}_j(\xi) &= E_P[\xi], \\ \sum_{j=1}^{J} \bar{p}_j(\xi) &= 1, \\ \bar{p}_j(\xi) &\geq 0, \qquad \forall j. \end{aligned} \tag{3.49}$$

We will first apply these ideas to the case where ξ is a one-dimensional random variable defined on an interval; that is, $\xi = \xi_1$ and $\Xi = [a, b]$.

Consider the convex function $\Phi(\xi) : [a, b] \to R$. Consider, also, the linear function $h(\xi) : [a, b] \to R$ whose equation is of the form: $h(\xi) = c\xi + d$.

We assume that the graph of $h(\xi)$ goes through the points $(a, \Phi(a))$ and $(b, \Phi(b))$. Hence, the equation of $h(\xi)$ is

$$\begin{aligned} h(\xi) &= \frac{\Phi(b) - \Phi(a)}{b - a} \xi + \frac{b}{b - a} \Phi(a) - \frac{a}{b - a} \Phi(b) \\ &= \frac{b - \xi}{b - a} \Phi(a) + \frac{\xi - a}{b - a} \Phi(b). \end{aligned} \tag{3.50}$$

The expected value of $h(\xi)$ is

$$E[h(\xi)] = \frac{b - E[\xi]}{b - a} \Phi(a) + \frac{E[\xi] - a}{b - a} \Phi(b), \tag{3.51}$$

where

$$E[\xi] = \int_a^b \xi P(d\xi).$$

If we let

$$\hat{p}(\xi = a) = \frac{b - \xi}{b - a}, \tag{3.52}$$

and

$$\hat{p}(\xi = b) = \frac{\xi - a}{b - a}, \tag{3.53}$$

then

$$\begin{aligned} h(\xi) &= \hat{p}(\xi = a)\Phi(a) + \hat{p}(\xi = b)\Phi(b) \\ &= E_{\hat{p}}[\Phi(\xi)]. \end{aligned} \tag{3.54}$$

We have found a way to replace an arbitrary probability density function (P) by a discrete probability function ($\hat{p}$). The expectation with respect to the new probability is $E_{\hat{p}}$.

Hence, for any $\xi \in \Xi$, we have

$$\begin{aligned} \xi &= E_{\hat{p}}[\xi] \\ &= \hat{p}(\xi = a)a + \hat{p}(\xi = b)b \\ &= \frac{b - \xi}{b - a}a + \frac{\xi - a}{b - a}b, \end{aligned} \tag{3.55}$$

and

$$\Phi(\xi) = \Phi\Big(\frac{b - \xi}{b - a}a + \frac{\xi - a}{b - a}b\Big). \tag{3.56}$$

Using the convexity of $\Phi(\cdot)$, we can write

$$\Phi(\xi) \leq \frac{b - \xi}{b - a}\Phi(a) + \frac{\xi - a}{b - a}\Phi(b). \tag{3.57}$$

We have just established the Edmundson-Madansky inequality which gives an upper bound to the value of $E[\Phi(\xi)]$.

Hence, since the recourse function of our problem is stochastic, we can apply inequality (3.57) to find an upper bound:

$$\begin{aligned} Q(x) &= E[Q(x, \xi)] \\ &= \int_a^b Q(x, \xi)P(d\xi) \\ &\leq \int_a^b [\frac{b - \xi}{b - a}Q(x, a) + \frac{\xi - a}{b - a}Q(x, b)]P(d\xi) \\ &= \frac{b - E[\xi]}{b - a}Q(x, a) + \frac{E[\xi] - a}{b - a}Q(x, b), \end{aligned} \tag{3.58}$$

that is,

$$\begin{aligned} Q(x) &\leq p_a Q(x,a) + p_b Q(x,b) \\ &= E_{\bar{P}}[Q(x,\xi)], \end{aligned} \tag{3.59}$$

where

$$p_a = \frac{b - E[\xi]}{b-a}, \tag{3.60}$$

and

$$p_b = \frac{E[\xi] - a}{b-a}. \tag{3.61}$$

The best choice for the new probability function is obtained by solving the problem

$$\begin{aligned} \min_{\bar{P}} \; & E_{\bar{P}}[Q(x,\xi)] = p_a Q(x,a) + p_b Q(x,b) \\ \text{s.t.} \; & \\ & p_a a + p_b b = E_P[\xi], \\ & p_a + p_b = 1, \\ & p_a, p_b \geq 0. \end{aligned} \tag{3.62}$$

As was explained earlier, the bound can be tightened by partitioning Ξ. We will first consider the case where the number of subsets is equal to 2 (i.e. L=2); that is,

$$[a,b] = [a,\xi^0] \cup [\xi^0, b]. \tag{3.63}$$

Hence,

$$\begin{aligned} Q(x) &= E[Q(x,\xi)] \\ &= \int_a^b Q(x,\xi) P(d\xi) \\ &= \int_a^{\xi^0} Q(x,\xi) P(d\xi) + \int_{\xi^0}^b Q(x,\xi) P(d\xi). \end{aligned} \tag{3.64}$$

We can, now, apply the Edmunson-Madansky inequality on both intervals, $[a,\xi^0]$, and $[\xi^0,b]$; we obtain the following inequality

$$\begin{aligned} Q(x) \leq p_1 \Big(\frac{\xi^0 - \bar{\xi}_1}{\xi^0 - a} Q(x,a) + \frac{\bar{\xi}_1 - a}{\xi^0 - a} Q(x,\xi^0) \Big) + \\ p_2 \Big(\frac{b - \bar{\xi}_2}{b - \xi^0} Q(x,\xi^0) + \frac{\bar{\xi}_2 - \xi^0}{b - \xi^0} Q(x,b) \Big), \end{aligned} \tag{3.65}$$

where

$$\begin{aligned} \xi^0 &= E\{\xi|[a,b]\}, \\ \bar{\xi}_1 &= E\{\xi|[a,\xi^0]\}, \text{ and} \\ \bar{\xi}_2 &= E\{\xi|[\xi^0,b]\}; \end{aligned}$$

and

$$p_1 = p(\xi \in [a,\xi^0]), \quad p_2 = p(\xi \in [\xi^0,b]).$$

The best choices for p_1 and p_2 are found by solving a problem equivalent to problem (3.62).

The inequality (3.65) can be generalised for $L \geq 2$. In order to do that, let the intervals

$$I_\ell = [\xi^{\ell-1},\xi^\ell] \quad \text{for} \quad \ell = 1,\ldots,L,$$

such that

$$I_\ell \cap I_m = \emptyset \quad \text{for} \quad \ell \neq m, \quad \text{and} \quad p_\ell = p(\xi \in I_\ell),$$

be a partition of $\Xi = [a,b]$, where $\xi^0 = a$ and $\xi^L = b$.

Hence, we obtain the following generalisation of inequality (3.65):

$$Q(x) \leq \sum_{\ell=1}^{L} \left(\frac{p_\ell}{\xi^\ell - \xi^{\ell-1}} \left[(\xi^\ell - \bar{\xi}_\ell) Q(x,\xi^{\ell-1}) + (\bar{\xi}_\ell - \xi^{\ell-1}) Q(x,\xi^\ell) \right] \right), \quad (3.66)$$

where $\bar{\xi}_\ell = E[\xi|I_\ell]$.

In the above discussion, ω was assumed to be of dimension one and we had $\xi = \xi_1$; we will, now, derive the inequality for the case where ω is a K-dimensional vector with independent components and we have $\xi = (\xi_1,\ldots,\xi_K)^T$.

If $K = 2$, we have

$$\xi = (\xi_1,\xi_2)^T \quad \text{such that}$$

$$\xi_k \in [a_k,b_k] \quad \text{for} \quad k = 1,2 \quad \text{and} \quad \Xi = \prod_{k=1}^{2} [a_k,b_k].$$

Then

$$Q(x) = E[Q(x,\xi_1,\xi_2)], \quad (3.67)$$

and since ξ_1 and ξ_2 are independent, we can write

$$Q(x) = \int_{a_2}^{b_2} dF_2(\xi_2) \int_{a_1}^{b_1} dF_1(\xi_1) Q(x,\xi_1,\xi_2), \quad (3.68)$$

where, $F_k(\xi_k) : [a_k, b_k] \to R[0,1]$ is the distribution function of ξ_k for $k = 1, 2$.

Now, we introduce a discrete probability function on $[a_1, b_1]$ and $[a_2, b_2]$:

$$p_1^1 = p(\xi_1 = a_1) = \frac{b_1 - \bar{\xi}_1}{b_1 - a_1}, \quad p_2^1 = p(\xi_1 = b_1) = \frac{\bar{\xi}_1 - a_1}{b_1 - a_1},$$
$$p_1^2 = p(\xi_2 = a_2) = \frac{b_2 - \bar{\xi}_2}{b_2 - a_2}, \quad p_2^2 = p(\xi_2 = b_2) = \frac{\bar{\xi}_2 - a_2}{b_2 - a_2},$$

where, for $k = 1, 2$, we have

$$\bar{\xi}_k = E\{\xi_k | [a_k, b_k]\},$$

and

$$p_1^k + p_2^k = 1.$$

Hence, using the discrete density functions introduced above, the possible outcomes of $\tilde{\xi}$ are:

$$\xi^1 = (a_1, a_2)^T, \quad \xi^2 = (b_1, a_2)^T, \quad \xi^3 = (a_1, b_2)^T, \quad \xi^4 = (b_1, b_2)^T,$$

with probabilities

$$\bar{p}_1 = p(\xi = \xi^1) = \left(\frac{b_2 - \bar{\xi}_2}{b_2 - a_2}\right)\left(\frac{b_1 - \bar{\xi}_1}{b_1 - a_1}\right), \quad \bar{p}_2 = p(\xi = \xi^2) = \left(\frac{b_2 - \bar{\xi}_2}{b_2 - a_2}\right)\left(\frac{\bar{\xi}_1 - a_1}{b_1 - a_1}\right)$$
$$\bar{p}_3 = p(\xi = \xi^3) = \left(\frac{b_1 - \bar{\xi}_1}{b_1 - a_1}\right)\left(\frac{b_2 - \bar{\xi}_2}{b_2 - a_2}\right), \quad \bar{p}_4 = p(\xi = \xi^4) = \left(\frac{\bar{\xi}_1 - a_1}{b_1 - a_1}\right)\left(\frac{\bar{\xi}_2 - a_2}{b_2 - a_2}\right)$$

Hence, we obtain the following inequality

$$Q(x) \leq \bar{p}_1 Q(x, \xi^1) + \bar{p}_2 Q(x, \xi^2) + \bar{p}_3 Q(x, \xi^3) + \bar{p}_4 Q(x, \xi^4). \tag{3.69}$$

Suppose, now, that $K \geq 2$; that is

$$\xi = (\xi_1, \ldots, \xi_K)^T \quad \text{such that}$$
$$\xi_k \in [a_k, b_k] \quad \text{for every} \quad k = 1, \ldots, K \quad \text{and} \quad \Xi = \prod_{k=1}^{K} [a_k, b_k].$$

If the function $F_k(\xi_k) : [a_k, b_k] \to R[0,1]$ is the distribution of ξ_k, then

$$\begin{aligned} Q(x) &= E[Q(x, \xi_1, \ldots, \xi_K)] \\ &= \int_{a_K}^{b_K} dF_K(\xi_K) \ldots \int_{a_1}^{b_1} dF_1(\xi_1) Q(x, \xi_1, \ldots, \xi_K). \end{aligned} \tag{3.70}$$

As earlier, we apply the Edmunson-Madansky inequality to the each interval $[a_k, b_k]$ while leaving the other intervals and the corresponding components unattended:

First, we apply the inequality to the interval $[a_1, b_1]$ while leaving $\xi_2, \ldots, \xi_K$ and their corresponding intervals unattended; we obtain

$$\begin{aligned} Q(x, \xi_1, \ldots, \xi_K) &\leq \Big(\frac{b_1 - \xi_1}{b_1 - a_1}\Big) Q(x, a_1, \xi_2, \ldots, \xi_K) \\ &+ \Big(\frac{\xi_1 - a_1}{b_1 - a_1}\Big) Q(x, b_1, \xi_2, \ldots, \xi_K). \end{aligned} \tag{3.71}$$

Integrating with respect to ξ yields

$$\begin{aligned} \int_{a_1}^{b_1} dF_1 Q(x, \xi_1, \ldots, \xi_K) &\leq \Big(\frac{b_1 - \bar{\xi}_1}{b_1 - a_1}\Big) Q(x, a_1, \xi_2, \ldots, \xi_K) \\ &+ \Big(\frac{\bar{\xi}_1 - a_1}{b_1 - a_1}\Big) Q(x, b_1, \xi_2, \ldots, \xi_K). \end{aligned} \tag{3.72}$$

Next, we leave $\xi_3, \ldots, \xi_K$ unattended, and repeat the above procedure for $\xi_2 \in [a_2, b_2]$ on expression (3.72) to obtain

$$\int_{a_2}^{b_2} dF_1 \int_{a_1}^{b_1} dF_1 Q(x, \xi_1, \ldots, \xi_K) \leq Q_1 + Q_2 + Q_3 + Q_4, \tag{3.73}$$

where

$$\begin{aligned} Q_1 &= \Big(\frac{b_1 - \bar{\xi}_1}{b_1 - a_1}\Big)\Big(\frac{b_2 - \bar{\xi}_2}{b_2 - a_2}\Big) Q(x, a_1, a_2, \xi_3, \ldots, \xi_K), \\ Q_2 &= \Big(\frac{b_1 - \bar{\xi}_1}{b_1 - a_1}\Big)\Big(\frac{\bar{\xi}_2 - a_2}{b_2 - a_2}\Big) Q(x, a_1, b_2, \xi_3, \ldots, \xi_K), \\ Q_3 &= \Big(\frac{\bar{\xi}_1 - a_1}{b_1 - a_1}\Big)\Big(\frac{b_2 - \bar{\xi}_2}{b_2 - a_2}\Big) Q(x, b_1, a_2, \xi_3, \ldots, \xi_K), \text{ and} \\ Q_4 &= \Big(\frac{\bar{\xi}_1 - a_1}{b_1 - a_1}\Big)\Big(\frac{\bar{\xi}_2 - a_2}{b_2 - a_2}\Big) Q(x, b_1, b_2, \xi_3, \ldots, \xi_K). \end{aligned}$$

The same procedure is repeated for $\xi_3, \ldots, \xi_K$; finally, we obtain

$$Q(x) \leq \Big(\prod_{k=1}^{K} (b_k - a_k)^{-1}\Big) \sum_{\gamma \in \Gamma} \Big(\Big(\prod_{k=1}^{K} |\bar{\xi}_k - \gamma_k|\Big) Q(x, \gamma_1, \ldots, \gamma_K)\Big), \tag{3.74}$$

where, $\Gamma = \{\gamma = (\gamma_1, \ldots, \gamma_K) : \quad \gamma_i = a_i \quad \text{or } b_i, i = 1, \ldots, K\}$.
Note that, $\Gamma = \text{ext } \Xi$.

As for the one dimensional case, we can tighten this inequality. It is done by partitioning each interval $[b_k, a_k]$; therefore, we have

$$\Xi_\ell = \prod_{k=1}^{K} [a_k^\ell, b_k^\ell] \quad \text{for every} \quad \ell = 1, \ldots, L.$$

In this case, we have

$$\begin{aligned} Q(x) &= E[Q(x,\xi)] \\ &= \sum_{\ell=1}^{L} \int_{\Xi_\ell} Q(x,\xi) P(d\xi) \\ &= \sum_{\ell=1}^{L} p_\ell \int_{\Xi_\ell} Q(x,\xi) P_\ell(d\xi) \\ &= \sum_{\ell=1}^{L} p_\ell E_\ell[Q(x,\xi)], \end{aligned} \tag{3.75}$$

where $P = p_\ell P_\ell$, and E_ℓ is the expectation with respect to P_ℓ.
The expectation, $E_\ell[Q(x,\xi)]$, is dealt with separately using the corresponding intervals; hence, we obtain

$$Q(x) \leq \sum_{\ell=1}^{L} p_\ell E_\ell[Q(x,\xi^\ell)], \tag{3.76}$$

where, ξ^ℓ is the random element associated with the probability function defined on Ξ_ℓ by

$$\xi^\ell = \begin{cases} a_k^\ell & \text{with probability } \dfrac{b_k^\ell - \bar{\xi}_k^\ell}{b_k^\ell - a_k^l} \\ \\ b_k^\ell & \text{with probability } \dfrac{\bar{\xi}_k^\ell - b_k^\ell}{b_k^\ell - a_k^\ell}, \end{cases} \tag{3.77}$$

where $\bar{\xi}_k^\ell = E[\xi|\Xi_\ell]$, and $p_\ell = p(\xi \in \Xi_\ell)$ for every $\ell = 1, \ldots, L^k$ and $k = 1, \ldots, K$.

Remark 3.2 The inequalities were derived for the case where a random variable is defined on an interval; the derivation for the case where the random variables are defined on simplices can be found in Frauendorfer (see [29]).

If we have randomness in both the objective and the right-hand side of the constraints, and the corresponding random variables are different and independent, then we get a lower bound by applying the Jensen inequality on the right-hand side random variables and the Edmunson-Madansky inequality in the objective. An upper bound is obtained if we apply these inequalities the other way around.

The inequalities resulting from the approximation schemes will give us a lower (LB) and an upper (UB) bound for the expected recourse function, i.e.

$$LB \leq E[Q(x,\xi)] \leq UB.$$

If (UB − LB $\leq$ Tolerance level), then we have found a solution; otherwise we need to tighten the bounds.

The L-shaped method can be be used within the approximation schemes.

The approximation of the probability function is not the only approximation scheme available; it is also possible to approximate the objective function itself. However, this technique works only for certain types of functions.

One should also note that there are several other approximation schemes. One such scheme is the Barycentric approximation which exploits the saddle property of the recourse function with respect to the random variables. This scheme is a powerful alternative to the Jensen and the Edmunson-Madansky inequalities when the number of random elements is large. Note that both these approximation schemes are special cases of the Barycentric approximation. Useful information on this scheme can be found in Frauendorfer (see [29]).

The discussion was focused on the case where the components of ξ are independent; the derivation of the inequalities for the case where these components are dependent can be found in Frauendorfer (see [28]). Note, however, that in practice these components are always assumed to be independent since it is extremely difficult to find an accurate description of their relationship.

The extension to the multi-stage case can be found in Frauendorfer (see [30]) and Birge (see [8]).

3.6 Monte Carlo sampling method

3.6.1 Monte Carlo method: A simple scheme

Approximation schemes have been used to tackle the difficulties resulting from the existence of multiple integrals, multiple sums and continuous distributions. However, when the number of independent variables is large, these schemes become extremely difficult to implement. In this case, Monte Carlo sampling method appears to be the only effective way to calculate the integral.

Suppose that we want to calculate the following integral:

$$\begin{aligned} I &= \int_{\Omega} f(\omega)\phi(\omega)\mathrm{d}\omega \\ &= \int\int\ldots\int f(\omega)\phi(\omega)\mathrm{d}\omega_1\mathrm{d}\omega_2\ldots\mathrm{d}\omega_K, \end{aligned} \tag{3.78}$$

where $\omega = (\omega_1, \omega_2, \ldots, \omega_K)^T$ has probability density function $\phi(\omega_1, \omega_2, \ldots, \omega_K)$.

The main idea behind Monte Carlo simulation is to find $\bar{I}$ an estimate of, I using a randomly and independently generated sample of vectors of size N, i.e. $(\omega^1, \omega^2, \ldots, \omega^N)$.

Since $\bar{I}$ is unlikely to be equal to I, two questions should be answered during the process:

- What should the sample size be in order to ensure a specific statistical accuracy?
- Given a sample of size N, how accurate is the estimated solution?

Different Monte Carlo techniques can be used to find the estimate, and for each of them, there is a different answer to the above questions.

Using the simple Monte Carlo technique, we generate N scenarios $(\omega^1, \omega^2, \ldots, \omega^N)$ independently from their joint probability density function, and the estimate of I is

$$\bar{I} = \frac{1}{N} \sum_{n=1}^{N} f(\omega^n). \tag{3.79}$$

Note that $\bar{I}$ is a random variable, and $E[\bar{I}] = I$; its accuracy is given by its variance

$$\sigma_{\bar{I}}^2 = \frac{var(I)}{N}, \tag{3.80}$$

where $var(I)$ is the variance of I.

We see that, though $\sigma_{\bar{I}}^2$ does not depend on K, the number of components of ω, it does, however, depend on the sample size. If we want to obtain a good accuracy using this technique, we must use a very large sample. Unfortunately, this is a very expensive option. Therefore, we will have to rely on a different technique.

3.6.2 Importance sampling

The idea here is (see [15], [16], [33], [40]) to find a probability density function $g(\cdot)$ such that,
firstly we have

$$\begin{aligned} I &= \int_{\Omega} f(\omega)\phi(\omega)\mathrm{d}\omega \\ &= \int_{\Omega} \frac{f(\omega)\phi(\omega)g(\omega)}{g(\omega)}\mathrm{d}\omega \\ &= E_g[\frac{f(\tilde{y})\phi(\tilde{y})}{g(\tilde{y})}], \end{aligned} \tag{3.81}$$

where $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_K)^T$ is a random vector with outcome $y \in Y$, and whose probability density function is also $g(\cdot)$.

Once we have generated N scenarios $y^1, y^2, \ldots, y^N$, we can calculate the estimate

$$\bar{I} = \frac{1}{N} \sum_{n=1}^{N} \frac{f(y^n)\phi(y^n)}{g(y^n)}. \tag{3.82}$$

Its variance is

$$\sigma_{\bar{I}}^2 = \frac{1}{N}\Big(E_g[\frac{f^2(\tilde{y})\phi^2(\tilde{y})}{g^2(\tilde{y})}] - I^2\Big). \tag{3.83}$$

Secondly, the probability density function $g(\cdot)$ should be chosen in such a way that the objective of the improved technique is achieved, i.e. the variance of the estimate must be reduced. In this case, the best result possible would be: $\sigma_{\bar{I}}^2 = 0$, that is we have the correct value of I in just one observation !!!
This implies that

$$g^*(y) = \frac{f(y)\phi(y)}{\sum_{y \in Y} f(y)\phi(y)}. \tag{3.84}$$

Unfortunately, the expression of g^* can not be used since its denominator, $\sum_{y \in Y} f(y)\phi(y)$, is what we are looking for, i.e. I. The way out is to replace $\sum_{y \in Y} f(y)\phi(y)$ by an approximation. A simple approximation is obtained by assuming that $f(y)$ is roughly additive (see [15]), i.e. $f(y) \approx \sum_{k=1}^{K} f_k(y_k)$; where f_k is the marginal contribution due to the k^{th} component of y, y_k. Using this approximation, we obtain a new expression for $g(\cdot)$; that is,

$$g(y) = \frac{\phi(y) \sum_{k=1}^{K} f_k(y_k)}{\sum_{y \in Y} \phi(y) \sum_{i=1}^{K} f_i(y_i)}. \tag{3.85}$$

Note that, the quality of this approximation will directly affect the variance-reduction capability of the importance sampling technique for this particular problem; i.e. if the additivity assumption is a reasonable one, we will be able to reduce the variance of $\bar{I}$ using a small sample. A further treatment of the previous expression gives

$$\begin{aligned}
g(y) &= \frac{\phi(y) \sum_{k=1}^{K} f_k(y_k)}{\sum_{l=1}^{K} \sum_{y \in Y} f_l(y_k)\phi(y)} \\
&= \frac{\phi(y) \sum_{k=1}^{K} f_k(y_k)}{\sum_{l=1}^{K} E[f_l(\tilde{y}_l)]} \\
&= \prod_{k=1}^{K} \phi_k(y_k) \sum_{l=1}^{K} \Big(\frac{E[f_l(\tilde{y}_l)]}{\sum_{m=1}^{K} E[f_m(\tilde{y}_m)]}\Big)\Big(\frac{f_l(y_l)}{E[f_l(\tilde{y}_l)]}\Big) \\
&= \sum_{k=1}^{K} \Big(\frac{E[f_k(\tilde{y}_k)]}{\sum_{l=1}^{K} E[f_l(\tilde{y}_l)]}\Big)\Big(\frac{\phi_k(\tilde{y}_k) f_k(y_k)}{E[f_k(\tilde{y}_k)]} \prod_{m \neq k} \phi_m(y_m)\Big),
\end{aligned} \tag{3.86}$$

where $E[f_k(\tilde{y}_k)]$ is estimated by sampling the marginal distribution of $\tilde{\omega}_k$.

Using this result, we can derive a new expression for I, i.e.

$$\begin{aligned}
I &= E_g[\frac{f(\tilde{y})\phi(\tilde{y})}{g(\tilde{y})}] \\
&= \sum_{y\in Y} \frac{f(y)\phi(y)g(y)}{g(y)} \\
&= \Big(\sum_{y\in Y} \frac{f(y)\phi(y)}{g(y)}\Big)\Big(\sum_{k=1}^{K}\Big(\frac{E[f_k(\tilde{y}_k)]}{\sum_{l=1}^{K} E[f_l(\tilde{y}_l)]}\Big)\Big(\frac{\phi_k(\tilde{y}_k)f_k(y_k)}{E[f_k(\tilde{y}_k)]}\prod_{m\neq k}\phi_m(y_m)\Big)\Big) \\
&= \sum_{k=1}^{K}\Big(\frac{E[f_k(\tilde{y}_k)]}{\sum_{l=1}^{K} E[f_l(\tilde{y}_l)]}\Big)\Big(\sum_{y\in Y}\frac{f(y)\phi(y)}{g(y)}\Big(\frac{\phi_k(\tilde{y}_k)f_k(y_k)}{E[f_k(\tilde{y}_k)]}\Big)\prod_{m\neq k}\phi_m(y_m)\Big) \\
&= \sum_{k=1}^{K}\Big(\frac{E[f_k(\tilde{y}_k)]}{\sum_{l=1}^{K} E[f_l(\tilde{y}_l)]}\Big)E_k[\frac{f(\tilde{y})\phi(\tilde{y})}{g(\tilde{y})}], \qquad (3.87)
\end{aligned}$$

where the use of E_k means that the component y_k of the sample vector y^n is independently sampled using the marginal distribution,

$$g_k(y_k) = \frac{\phi_k(\tilde{y}_k)f_k(y_k)}{E[f_k(\tilde{y}_k)]}. \qquad (3.88)$$

Note that, the above calculations make use of the $f_k(y_k)$'s; however, they are not given, and need to be calculated. In order to do that, we consider the cost in the $(f, t_1, \ldots, t_K)$-space so that we have

$$f_k(y_k) \approx f(\tau_1, \ldots, \tau_{i-1}, y_i, \tau_{i+1}, \ldots \tau_K) - f(\tau_1, \ldots, \tau_i, \ldots \tau_K), \qquad (3.89)$$

where the point $\tau = (\tau_1, \ldots \tau_K)^T$ is arbitrarily chosen from the set of possible outcomes of $\tilde{y}$.

If F is the additive approximation of f, then we have

$$F(y) = f(\tau) + \sum_{k=1}^{K} f_k(y_k). \qquad (3.90)$$

Hence,

$$\begin{aligned}
I &= f(\tau) + E_g[f(y) - f(\tau)] \\
&= f(\tau) + \sum_{y\in Y}(f(y) - f(\tau))g(y) \\
&= f(\tau) + \sum_{k=1}^{K} E[f_k(\tilde{y}_k)]\sum_{y\in Y}\Delta(y)\frac{f_k(y_k)}{E[f_k(\tilde{y}_k)]}\prod_{m\neq k}\phi_m(y_m), \qquad (3.91)
\end{aligned}$$

where

$$\Delta(y) = \frac{f(y) - f(\tau)}{\sum_{l=1}^{K} E[f_l(\tilde{y}_l)]}.$$

To find the estimate of I, we partition the sample space into K non-empty subsets such that N_k the size of the subset k is approximately proportional to $E[f_k(\tilde{y}_k)]$.
Using

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N^k} \Delta(y^n), \tag{3.92}$$

the estimate of the k-th sum,

$$\sum_{y \in Y} \Delta(y) \frac{f_k(y_k)}{E[f_k(\tilde{y}_k)]} \prod_{m \neq k} \phi_m(y_m), \tag{3.93}$$

we obtain

$$\bar{I} = f(\tau) + \sum_{k=1}^{K} E[f_k(\tilde{y}_k)]\mu_k, \tag{3.94}$$

and its variance is given by

$$\sigma_{\bar{I}}^2 = \sum_{k=1}^{K} \frac{E[f_k(\tilde{y}_k)]\bar{\sigma}_k^2}{N_k}, \tag{3.95}$$

where $\bar{\sigma}_k^2$ is the sample variance of the k-th expectation (note that, $\bar{\sigma}_k^2 = 0$ if $N_k = 1$).

Now that we have found, $g(\cdot)$, the probability function that has the potential to reduce the variance of the estimate when a small sample is used, we can go back to our original problem.

3.6.3 Importance sampling within the L-shaped method

After N iterations of the L-shaped methods, we have the following master problem:

min $f_1^T x_1 + \Theta$
s.t.

$$\begin{aligned}
A_{11}x_1 &= b_1, \\
\sigma_s^T A_{21}(\omega^j)x_1 &\geq \sigma_s^T b_2(\omega^j) \qquad (s = 1, \ldots, V), \\
\sum_{j=1}^{K} p_j \pi_s^T A_{21}(\omega^j)x_1 + \Theta &\geq \sum_{j=1}^{K} p_j \pi_s^T b_2^j(\omega^j) \qquad (s = 1, \ldots, J), \\
x_1 &\geq 0, \quad \Theta \in R,
\end{aligned} \tag{3.96}$$

which can assume the new form:

$\min f_1^T x_1 + \Theta$
s.t.

$$\begin{aligned} A_{11}x_1 &= b_1, \\ \sum_{j=1}^{K} \sigma_s^T A_{21}(\omega^j)x_1 &\geq \sum_{j=1}^{K} \sigma_s^T b_2^j(\omega^j) \qquad (s = 1, \ldots, V), \\ \sum_{j=1}^{K} p_j \pi_s^T A_{21}(\omega^j)x_1 + \Theta &\geq \sum_{j=1}^{K} p_j \pi_s^T b_2^j(\omega^j) \qquad (s = 1, \ldots, J), \\ x_1 &\geq 0, \Theta \in R, \end{aligned} \tag{3.97}$$

that is,

$\min f_1^T x_1 + \Theta$
s.t.

$$\begin{aligned} A_{11}x_1 &= b_1, \\ E[\sigma_s^T A_{21}(\tilde{\omega})]x_1 &\geq E[\sigma_s^T b_2(\tilde{\omega})] \qquad (s = 1, \ldots, V), \\ E[p_j \pi_s^T A_{21}(\tilde{\omega})]x_1 + \Theta &\geq E[p_j \pi_s^T b_2(\tilde{\omega})] \qquad (s = 1, \ldots, J), \\ x_1 &\geq 0, \Theta \in R, \end{aligned} \tag{3.98}$$

where the terminology is the same as the one introduced earlier (see section 3.4).

The importance sampling methodology will be applied to estimate the expected values embedded in the constraints of problem (3.98), and to find the expected second-stage objective value (see [40]: 22-43); i.e. the objective value of the subproblems.

Using $\hat{x}$ as a parameter in the expression of $\Delta(\cdot)$, the terminology and the ideas presented in the previous section, we can write

$$\Delta^D(\omega, \hat{x}) = \frac{\sigma_s^T(\omega, \hat{x})A_{21}(\omega) - \sigma_s^T(\tau, \hat{x})A_{21}(\tau)}{\sum_{j=1}^{m_2} E[\tilde{M}_j^D]}, \tag{3.99}$$

$$\Delta^d(\omega, \hat{x}) = \frac{\sigma_s^T(\omega, \hat{x})b_2(\omega) - \sigma_s^T(\tau, \hat{x})b_2(\tau)}{\sum_{j=1}^{m_2} E[\tilde{M}_j^d]}, \tag{3.100}$$

$$\Delta^E(\omega, \hat{x}) = \frac{\pi_s^T(\omega, \hat{x})A_{21}(\omega) - \pi_s^T(\tau, \hat{x})A_{21}(\tau)}{\sum_{j=1}^{m_2} E[\tilde{M}_j^E]}, \tag{3.101}$$

$$\Delta^d(\omega, \hat{x}) = \frac{\pi_s^T(\omega, \hat{x})b_2(\omega) - \pi_s^T(\tau, \hat{x})b_2(\tau)}{\sum_{j=1}^{m_2} E[\tilde{M}_j^e]}, \tag{3.102}$$

where

which can assume the new form:

$\min f_1^T x_1 + \Theta$
s.t.

$$\begin{aligned} A_{11}x_1 &= b_1, \\ \sum_{j=1}^{K} \sigma_s^T A_{21}(\omega^j)x_1 &\geq \sum_{j=1}^{K} \sigma_s^T b_2^j(\omega^j) \qquad (s = 1, \ldots, V), \\ \sum_{j=1}^{K} p_j \pi_s^T A_{21}(\omega^j)x_1 + \Theta &\geq \sum_{j=1}^{K} p_j \pi_s^T b_2^j(\omega^j) \qquad (s = 1, \ldots, J), \\ x_1 &\geq 0, \Theta \in R, \end{aligned} \tag{3.97}$$

that is,

$\min f_1^T x_1 + \Theta$
s.t.

$$\begin{aligned} A_{11}x_1 &= b_1, \\ E[\sigma_s^T A_{21}(\tilde{\omega})]x_1 &\geq E[\sigma_s^T b_2(\tilde{\omega})] \qquad (s = 1, \ldots, V), \\ E[p_j \pi_s^T A_{21}(\tilde{\omega})]x_1 + \Theta &\geq E[p_j \pi_s^T b_2(\tilde{\omega})] \qquad (s = 1, \ldots, J), \\ x_1 &\geq 0, \Theta \in R, \end{aligned} \tag{3.98}$$

where the terminology is the same as the one introduced earlier (see section 3.4).

The importance sampling methodology will be applied to estimate the expected values embedded in the constraints of problem (3.98), and to find the expected second-stage objective value (see [40]: 22-43); i.e. the objective value of the subproblems.

Using $\hat{x}$ as a parameter in the expression of $\Delta(\cdot)$, the terminology and the ideas presented in the previous section, we can write

$$\Delta^D(\omega, \hat{x}) = \frac{\sigma_s^T(\omega, \hat{x})A_{21}(\omega) - \sigma_s^T(\tau, \hat{x})A_{21}(\tau)}{\sum_{j=1}^{m_2} E[\tilde{M}_j^D]}, \tag{3.99}$$

$$\Delta^d(\omega, \hat{x}) = \frac{\sigma_s^T(\omega, \hat{x})b_2(\omega) - \sigma_s^T(\tau, \hat{x})b_2(\tau)}{\sum_{j=1}^{m_2} E[\tilde{M}_j^d]}, \tag{3.100}$$

$$\Delta^E(\omega, \hat{x}) = \frac{\pi_s^T(\omega, \hat{x})A_{21}(\omega) - \pi_s^T(\tau, \hat{x})A_{21}(\tau)}{\sum_{j=1}^{m_2} E[\tilde{M}_j^E]}, \tag{3.101}$$

$$\Delta^d(\omega, \hat{x}) = \frac{\pi_s^T(\omega, \hat{x})b_2(\omega) - \pi_s^T(\tau, \hat{x})b_2(\tau)}{\sum_{j=1}^{m_2} E[\tilde{M}_j^e]}, \tag{3.102}$$

where

- $M_j^D = \sigma_s^T(\omega_j, \hat{x})A_{21}(\omega_j)$, is the marginal contribution of ω_j.
- $M_j^d = \sigma_s^T(\omega_j, \hat{x})b_2(\omega_j)$, is the marginal contribution of ω_j.
- $M_j^E = \pi_s^T(\omega_j, \hat{x})A_{21}(\omega_j)$, is the marginal contribution of ω_j.
- $M_j^e = \pi_s^T(\omega_j, \hat{x})b_2(\omega_j)$, is the marginal contribution of ω_j.

The expected values used in the constraints are

$$\begin{aligned} I_1 &= E[\sigma_s^T A_{21}(\tilde{\omega})] \\ &= \sigma_s^T(\tau, \hat{x})A_{21}(\tau) + \\ &\quad \sum_{j=1}^{K} E[\tilde{M}_j^D] \sum_{\omega \in \Omega} \Delta^D(\omega, \hat{x}) \frac{M_j^D}{E[\tilde{M}_j^D]} \prod_i^K \phi_i(\omega_i) \end{aligned} \tag{3.103}$$

$$\begin{aligned} I_2 &= E[\sigma_s^T b_2(\tilde{\omega})] \\ &= \sigma_s^T(\tau, \hat{x})b_2(\tau) + \sum_{j=1}^{K} E[\tilde{M}_j^d] \sum_{\omega \in \Omega} \Delta^d(\omega, \hat{x}) \frac{M_j^d}{E[\tilde{M}_j^d]} \prod_i^K \phi_i(\omega_i) \end{aligned} \tag{3.104}$$

$$\begin{aligned} I_3 &= E[\pi_s^T A_{21}(\tilde{\omega})] \\ &= \pi_s^T(\tau, \hat{x})A_{21}(\tau) + \\ &\quad \sum_{j=1}^{K} E[\tilde{M}_j^E] \sum_{\omega \in \Omega} \Delta^E(\omega, \hat{x}) \frac{M_j^E}{E[\tilde{M}_j^E]} \prod_i^K \phi_i(\omega_i) \end{aligned} \tag{3.105}$$

$$\begin{aligned} I_4 &= E[\pi_s^T b_2(\tilde{\omega})] \\ &= \pi_s^T(\tau, \hat{x})b_2(\tau) + \sum_{j=1}^{K} E[\tilde{M}_j^e] \sum_{\omega \in \Omega} \Delta^e(\omega, \hat{x}) \frac{M_j^e}{E[\tilde{M}_j^e]} \prod_i^K \phi_i(\omega_i) \end{aligned} \tag{3.106}$$

Their estimates are

$$\bar{I}_1 = \sigma_s^T(\tau, \hat{x})A_{21}(\tau) + \sum_{j=1}^{K} E[\tilde{M}_j^D]\mu_j^D, \tag{3.107}$$

$$\bar{I}_2 = \sigma_s^T(\tau, \hat{x})b_2(\tau) + \sum_{j=1}^{K} E[\tilde{M}_j^d]\mu_j^d, \tag{3.108}$$

$$\bar{I}_3 = \pi_s^T(\tau, \hat{x})A_{21}(\tau) + \sum_{j=1}^{K} E[\tilde{M}_j^E]\mu_j^E, \tag{3.109}$$

$$\bar{I}_4 = \pi_s^T(\tau, \hat{x})b_2(\tau) + \sum_{j=1}^{K} E[\tilde{M}_j^e]\mu_j^e, \tag{3.110}$$

where

$\mu_D^j = \frac{1}{N_j}\Delta^D(\omega^j, \hat{x})$, $\mu_d^j = \frac{1}{N_j}\Delta^d(\omega^j, \hat{x})$

$\mu_E^j = \frac{1}{N_j}\Delta^E(\omega^j, \hat{x})$, $\mu_e^j = \frac{1}{N_j}\Delta^e(\omega^j, \hat{x})$,

and $\sum_{j=1}^{m_2} N_j = N$.

Using these estimates, problem(3.98) becomes

$$\begin{aligned} \min \quad & f_1^T x_1 + \Theta \\ \text{s.t.} \quad & \\ & A_{11}x_1 = b_1, \\ & \bar{I}_1 x_1 \geq \bar{I}_2 \qquad (s = 1, \ldots, V), \\ & \bar{I}_3 x_1 + \Theta \geq \bar{I}_4 \qquad (s = 1, \ldots, J), \\ & x_1 \geq 0, \Theta \in R, \end{aligned} \tag{3.111}$$

- **The Algorithm**(see [40] : 42-43)

Step 0
Initialise: $\nu = 0$, $\bar{UB}^0 = \infty$.

Step 1
Solve the relaxed master problem:

$$\begin{aligned} \min \quad & f_1^T x_1 + \Theta \\ \text{s.t.} \quad & \\ & A_{11}x_1 = b_1, \\ & x_1 \geq 0, \quad \Theta \in R, \end{aligned} \tag{3.112}$$

and let a lower bound be, $\bar{LB}^\nu = f^T\hat{x}^\nu + \Theta^\nu$.

Step 2
Let $\nu = \nu + 1$
Solve the dual subproblems and obtain an upper bound:
$\bar{UB}^\nu = \min\{\bar{UB}^{\nu-1}, f_1^T\hat{x}^\nu + E[f_2(\hat{x}^\nu)]\}$.
Add a new cut, computed using importance sampling, to the master problem.

Step 3
Solve the master problem, and let a lower bound be, $\bar{LB}^\nu = f^T\hat{x}^\nu + \Theta^\nu$.

Step 4
If $\bar{LB}^{\nu}$ and $\bar{UB}^{\nu}$ (note that both are random variables) are indistinguishable in distribution, then go to **Step 2**.
Otherwise calculate a new solution $(\hat{x}^{\nu}, \hat{\Theta}^{\nu})$.

Step 5
If the solution is satisfactory, stop.
Otherwise go to Step 6.

Step 6
Increase the sample size, and go to Step 2.

Remark 3.3 To find the estimate of the expected cost of the second-stage problem, it suffices to follow the same steps using the appropriate integral.

The use of Monte Carlo techniques implies not solving the whole problem; the solution is obtained for the sample set but not for all possible outcomes, $\omega \in \Omega$. However, it is reasonable to assume that the error made while estimating an integral is normally distributed with variance zero (see [40] : 32-34).

In general, for a reasonably large sample set, the variance of the estimate is small.

Cuts obtained using Monte Carlo techniques are not necessarily valid cuts; that is, they do sometimes cut off the feasible region of the second-stage problem (see [40] : 34-35).

Chapter 4

Multi-stage stochastic programming

4.1 The bond portfolio management problem

We illustrate this problem with an extension of the example introduced in the previous. Consider a planning period starting now, time $t = 0$, and ending at time T. The portfolio manager has to make a decision regarding the composition of her portfolio at each point in time. The decisions, at time t ($t \geq 0$), are made taking into account the history up to date, $\omega^{t-1} \in \Omega^t$. This history represents the realisations, $\omega_t \in \Omega_t$, of the random parameters observed at each time point from $t = 0$ up to and including the previous date (i.e. $\omega^t = (\omega_0, \ldots, \omega_{t-1})^T$ where ω_0 is the value of the parameters at $t = 0$). The manager must also consider the possible state of nature, ω_t, at this time point. The objective of the portfolio manager is to maximise the expected utility of the wealth at time T. Note that her utility function is $U(\cdot)$. She must also make sure that the future liability payments are made. This is done in $T + 1$ stages:

1. In the first stage (i.e. time $t = 0$), she starts with an initial portfolio, and makes the first stage decision about the composition of her portfolio; that is, she chooses the investment that will maximise the expected utility value of her final wealth under the set of constraints.

2. In the $(t + 1)$-th stage ($1 \leq t \leq T$), she investigates the possible scenarios (i.e. ω^t) and has to assess the value of her portfolio. She compares the total cash flow to the value of the liabilities. Depending on whether the total cash flow allows the payment of the liabilities or not, she will have to make additional changes to the portfolio holdings.

In order to formulate this problem,we consider:

- b_0: the initial holding (in face value) of the risk free asset.

- b_j: the initial holdings of bond j ($j = 1, \ldots, J$).
- $r_t(\omega^t)$: the short term interest rate during the period $(t, t+1)$ given the history ω^t.
- $f_{jt}(\omega^t)$: the cash flow generated from bond j at time t given the history ω^t; the cash flow is expressed as a fraction of the face value.
- $\xi^s_{jt}(\omega^t)$: the selling price of bond j at time t given the history ω^t. This price corresponds to the actual price of the bond (i.e. sum of the discounted cash flows) minus the transaction cost. The price at time $t = 0$, ξ_{j0}, is not scenario dependent.
- $\zeta^s_{jt}(\omega^t)$: the purchasing price of bond j at time t given the history ω^t. This price corresponds to the actual price of the bond (i.e. sum of the discounted cash flows) and the transaction cost. The price at time $t = 0$, ζ_{j0}, is not scenario dependent.
- $L_t(\omega^t)$: the liability due at time t given the history ω^t.

Next, we consider the following first-stage variables:

- x_j: the face value of bond j purchased at the beginning of the planning period. This variable is assumed to be a non-negative real number.
- y_j: the face value of bond j sold at the beginning of the planning period. This is also a non-negative real number.
- z_{j0}: the face value of bond j held in the portfolio after the initial decisions, x_j and y_j, have been made.
- y_0^+: the initial surplus. It is also a non-negative real number.

We also consider the $(t+1)$-th stage variables ($1 \leq t \leq T-1$):

- $x_{jt}(\omega^t)$: the face value of bond j purchased at time t given the history ω^t.
- $y_{jt(\omega^t)}$: the face value of bond j sold at time t given the history ω^t.
- $z_{jt}(\omega^t)$: the face value of bond j held after bonds have been bought and sold at time t given the history ω^t.
- $y_t^-(\omega^t)$: the amount borrowed at time t given the history ω^t.
- $y_t^+(\omega^t)$: the amount invested in the risk free asset (lending) at time t given the history ω^t.

Finally, we consider the $(T+1)$-th stage variables:

- $x_{jT}(\omega^T)$: the face value of bond j purchased at time t given the history ω^T.
- $y_{jT}(\omega^T)$: the face value of bond j sold at time T given the history ω^T.

- $z_{jT}(\omega^T)$: the face value of bond j held after bonds have been bought and sold at time T given the history ω^T.
- $y_T^-(\omega^T)$: the amount borrowed at time T given the history ω^T.
- $y_T^+(\omega^T)$: the amount invested in the risk free asset (lending) at time T given the history ω^T.
- $W_T(\omega^T)$: the total wealth at time T given the history ω^T.

The resulting multi-stage stochastic problem is

$$\max \sum_{\omega^T \in \Omega^T} p(\omega^T) \cdot U\left(W_T(\omega^T)\right), \tag{4.1}$$

under the following constraints:

1. The first-stage constraints:
   - The conservation of the initial holdings; that is, the face value of bond j sold plus the value of the bond left, after the implementation of the initial decision, is equal to the original holdings of the bond plus the value of the bond purchased.
   $$y_j + z_{j0} = b_j + x_j, \qquad \forall j. \tag{4.2}$$
   - The cash flow conservation constraint.
   $$y_0^+ + \sum_{j=1}^{J} \zeta_{j0} x_j = b_0 + \sum_{j=1}^{J} \xi_{j0} y_j. \tag{4.3}$$
2. The $(t+1)$-th stage constraints $(1 \le t \le T-2)$:
   - The conservation of holdings for every bond j given every possible outcome of the random elements.
   $$z_{jt}(\omega^t) + y_{jt}(\omega^t) = z_{j,t-1}(\omega^t) + x_{jt}(\omega^t), \ \forall j, \text{ and } \forall \omega^t \in \Omega^t. \tag{4.4}$$
   - The cash flow conservation constraints at each time t for every outcome of the random elements. These constraints take into account the cash generated by buy and sell decisions, borrowing and reinvestment decisions as well as liability payments.
   $$\begin{aligned} \sum_{j=1}^{J} \xi_{jt}(\omega^t) y_{jt}(\omega^t) + \sum_{j=1}^{J} f_{jt}(\omega^t) z_{j,t-1}(\omega^t) + y_t^-(\omega^t) \\ + \left(1 + r_{t-1}(\omega^t)\right) y_{t-1}^+(\omega^t) \\ = L_t(\omega^t) + \sum_{j=1}^{J} \zeta_{jt}(\omega^t) x_{jt}(\omega^t) + \left(1 + r_{t-1}(\omega^t)\right) y_{t-1}^-(\omega^t) \\ + y_t^+(\omega^t), \quad \forall \omega^t \in \Omega^t. \end{aligned} \tag{4.5}$$

3. The $(T+1)$-th stage constraints:

   - The conservation of holdings for every bond j given every possible outcome of the random elements.

$$z_{jT}(\omega^T) + y_{jT}(\omega^T) = z_{j,T-1}(\omega^T) + x_{jT}(\omega^T), \qquad (4.6)$$
$$\forall j \quad \text{and} \quad \forall \omega^T \in \Omega^T.$$

   - The cash flow conservation constraints for every outcome of the random elements. These constraints take into account the cash generated by buy and sell decisions, borrowing and reinvestment decisions as well as liability payments.

$$\sum_{j=1}^{J} \xi_{jT}(\omega^T) y_{jT}(\omega^T) + \sum_{j=1}^{J} f_{jT}(\omega^T) z_{j,T-1}(\omega^T) + y_T^-(\omega^T) + \left(1 + r_{T-1}(\omega^T)\right) y_{T-1}^+(\omega^T)$$
$$= L_T(\omega^T) + \sum_{j=1}^{J} \zeta_{jT}(\omega^T) x_{jT}(\omega^T) + \left(1 + r_{T-1}(\omega^T)\right) y_{T-1}^-(\omega^T) + y_T^+(\omega^T), \quad \forall \omega^T \in \Omega^T. \qquad (4.7)$$

   - The final wealth is obtained by valuing the holding of the portfolio at time T and paying off the outstanding debt.

$$W_T(\omega^T) = \sum_{j=1}^{J} \xi_{jT}(\omega^T) z_{jT}(\omega^T) + y_T^+(\omega^T) - y_T^-(\omega^T), \qquad (4.8)$$
$$\forall \omega^T \in \Omega^T.$$

4. The non-negativity constraints of all the variables.

   - All the variables are non-negative real numbers.

5. The non-anticipativity constraints.

4.2 Definition of the problem

The portfolio manager has to make a series of decisions over a given planning period; there are T time stages, $t = 1, \ldots, T$.

At $t = 1$, a decision, x_1, is made. At every time $t = 2, \ldots, T$, an outcome, ω_t, is observed and a decision x_t is made. As was explained for the two-stage case, the decision, x_1, is made without any knowledge of future outcomes (in the best cases, only their probability density functions are known).

At $t = 2$, a realisation of the random element, $\tilde{\omega}_2$, is observed and a decision, x_2, is made. The decision, x_2, is made in such a way that the objective function is minimised (maximised) with respect to the decision x_1, on the one hand, and the outcome of the random elements $\tilde{\omega}_t$, for $t = 3, \ldots, T$ on the other hand.

The decision, x_3, is made in such a way that the objective function is minimised (maximised) with respect to the decisions x_1, and x_2, on the one hand, and the outcome of the random elements $\tilde{\omega}_t$, for $t = 4, \ldots, T$, on the other hand.

The process is repeated until $t = T$.

This observation-decision process is best captured in the following representation:

$$x_3 \rightarrow \omega_2 \rightarrow x_2 \ldots \rightarrow \omega_T \rightarrow x_T,$$

and is modelled as follows:

$$\min_{x_1} f_1(x^1) + E_{\omega^2}\Big\{ \min_{x_2} f_2(x^2(\tilde{\omega}^2)) + \ldots + E_{\omega_T \setminus \omega^T}\Big[\min_{x_T} f_T(x^T(\tilde{\omega}^T))\Big]\Big\}$$

s.t.

$$\begin{array}{lcll} A_{11}x_1 & = & b_1, & \\ \sum_{i=1}^{t} A_{ti}(\tilde{\omega}^t)x_i(\tilde{\omega}^t) & = & b_t(\tilde{\omega}^t) & a.s. \quad \forall t, \\ x_t(\tilde{\omega}^t) & \geq & 0 & a.s. \quad \forall t, \end{array} \tag{4.9}$$

where

- $x_1(\tilde{\omega}^1) = x_1$.
- Each $x_t(\tilde{\omega}^t)$ is a vector of dimension n_t.
- $A_{11} \in R^{m_1 * n_1}$ and $b_1 \in R^{m_1}$.
- For each $t = 2, \ldots, T$ we have $A_{t,t-1} \in R^{m_t * n_{t-1}}, A_{tt} \in R^{m_t * n_t}$ and $b_t \in R^{m_t}$.
- For each $t = 1, \ldots, T$, the decision variable $x_t \in R^{n_t}, x^t = (x_1, \ldots, x_t)^T$, and we have, $x = (x_1, \ldots, x_T)^T \in R^n (n = \sum_{t=1}^{T} n_t)$.
- For each $t = 2, \ldots, T$, the random vectors $\tilde{\omega}_t (t \geq 2)$, are defined in the probability space (Ω_t, P_t), and $\tilde{\omega}^t = (\tilde{\omega}_1, \ldots, \tilde{\omega}_{t-1})^T$ and the sequence $\{\tilde{\omega}_t\}_{t=1}^{t}$ is assumed to be auto-correlated, i.e. they are dependent between stages.
- $E_{\omega_t \setminus \omega^t}$ is the conditional expectation with respect to the random variable $\tilde{\omega}_t$ given the history process ω^t.

- For each $t = 2, \ldots, T$, the decision process $x^t = (x_1, \ldots, x_t)^T$, is assumed to be adapted or nonanticipative, i.e. for a given history, ω^t, we must have a unique solution x^t.

If, for each, $t = 2, \ldots, T$, we let $B_t(x^{t-1}, \omega^t) = \sum_{\tau=1}^{t-1} A_{t\tau}(\omega^t)x_\tau$, and $A_t(x_t) = A_{tt}(\tilde{\omega}^t)x_t$, the previous model becomes:

$$\min_{x_1} f_1(x^1) + E_{\omega^2}\Big\{ \min_{x_2} f_2(x^2(\tilde{\omega}^2)) + \ldots + E_{\omega_T \setminus \omega^T}\Big[\min_{x_T} f_T(x^T(\tilde{\omega}^T))\Big]\Big\}$$

s.t.

$$\begin{array}{lcl} A_1 x_1 & = & b_1, \\ B_t(\tilde{\omega}^t)x_{t-1}(\tilde{\omega}^{t-1}) + A_t(\tilde{\omega}^t)x_t(\tilde{\omega}^t) & = & b_t(\tilde{\omega}^t) \quad a.s. \quad \forall t, \\ x_t(\tilde{\omega}^t) & \geq & 0 \quad a.s. \quad \forall t. \end{array} \tag{4.10}$$

Problem (4.10) can also be expressed in the same form as the one used for the two-stage case; that is,

$$\begin{array}{lcl} \min \quad f_1(x_1) + E_{\omega_2}[Q_2(x_1, \tilde{\omega}^2)] & & \\ \text{s.t.} & & \\ A_1 x_1 & = & b_1, \\ x_1 & \geq & 0, \end{array} \tag{4.11}$$

where, for every $t = 2, \ldots, T$, we have:

$$\begin{array}{l} Q_t(x_{t-1}(\omega^{t-1}), \omega^t) = \min f_t(x_t(\omega^t)) + E_{\omega_{t+1} \setminus \omega^{t+1}}[Q_{t+1}(x_t, \tilde{\omega}^{t+1})] \\ \qquad \text{s.t.} \end{array}$$

$$\begin{array}{lcl} B_t(\omega^t)x_{t-1}(\omega^{t-1}) + A_t(\omega^t)x_t(\omega^t) & = & b_t(\omega^t)\text{a.s.} \\ x_t(\omega^t) & \geq & 0. \end{array} \tag{4.12}$$

4.3 The L-shaped method applied to multi-stage problems

The L-shaped method introduced earlier is extended to the multi-stage case; This is achieved by means of the nested Benders' decomposition method (see [6], [32]). This methods splits the multi-stage problem into a collection of two-stage relations that are connected by a nesting scheme.

We assume that the function $f_t(x_t)$ is linear, i.e. $f_t(x_t) = f_t^T x_t$, and that the second stage problems have fixed recourse. We also assume that the random variables take discrete values and that their respective sample spaces are finite. At stage $t = 2, \ldots, T-1$, for each outcome of the random vector $\tilde{\omega}_t$, we have the subproblem

$$\min \quad f_t^T x_t(\omega^t) + E_{\omega_{t+1} \setminus \omega^{t+1}}[Q_{t+1}(x_t, \tilde{\omega}^{t+1})]$$

s.t.

$$\begin{aligned} B_t(\omega^t)x_{t-1}(\omega^{t-1}) + A_t(\omega^t)x_t(\omega^t) &= b_t(\omega^t) \quad \text{a.s.} \\ x_t(\omega^t) &\geq 0, \end{aligned} \tag{4.13}$$

where

$$Q_{t+1}(x_t, \omega^{t+1}) = \min_{\text{s.t.}} f_{t+1}^T x_{t+1}(\omega^{t+1}) + E_{\omega_{t+2}|\omega^{t+2}}[Q_{t+2}(x_{t+1}, \tilde{\omega}^{t+2})]$$

$$\begin{aligned} B_{t+1}(\omega^{t+1})x_t(\omega^t) + A_{t+1}(\omega^{t+1})x_{t+1}(\omega^{t+1}) &= b_{t+1}(\omega^{t+1}) \text{a.s.} \\ x_{t+1}(\omega^{t+1}) &\geq 0. \end{aligned} \tag{4.14}$$

At stage T, for each outcome of the random vector $\tilde{\omega}_T$, we have

$$\begin{aligned} \min \quad & f_T^T x_T(\omega^T) \\ \text{s.t.} \quad & \\ & B_T(\omega^T)x_{T-1}(\omega^{T-1}) + A_T(\omega^T)x_T(\omega^T) = b_T(\omega^T) \quad \text{a.s.} \\ & x_T(\omega^T) \geq 0. \end{aligned} \tag{4.15}$$

Since we have assumed that the random vector $\tilde{\omega}_t$, assumes discrete values and that the sample space Ω_t is finite, then for the k_2-th outcome of $\tilde{\omega}_t$, the problem (4.13) is

$$\begin{aligned} \min\ & f_2^T x_2^{k_2} + \textstyle\sum_{k_3=K_2+1}^{K_3} \bar{p}_{k_3} f_3^T x_3^{k_3} + \ldots + \sum_{k_T=K_{T-1}+1}^{K_T} \bar{p}_{k_T} f_T^T x_T^{k_T} \\ \text{s.t.} & \end{aligned}$$

$$\begin{aligned} A_2 x_2^{k_2} &= b_2^{k_2} - B_2^{k_2} x_0, \\ A_3 x_3^{k_3} &= b_3^{k_3} - B_3^{k_3} x_2^{k_2}, \quad k_3 \in d_3(k_2), \\ &\vdots \\ A_T x_T^{k_T} &= b_T^{k_T} - B_T^{k_T} x_{a(k_T)}, \quad k_T \in d_T(k_2), \end{aligned} \tag{4.16}$$

where, for each t and each node k, $d_t(k_i)$ represents the set of all nodes in period t which are descendants of node k_i, and $a(k_t)$ refers to the immediate ancestor of node k_t. The probabilities $\bar{p}_{k_t} = p_{k_t}/p_{k_2}$ are the conditional probabilities given that the realisation k_2 has been observed. The following figure (where

$T = 3$) will help to understand the notation used:

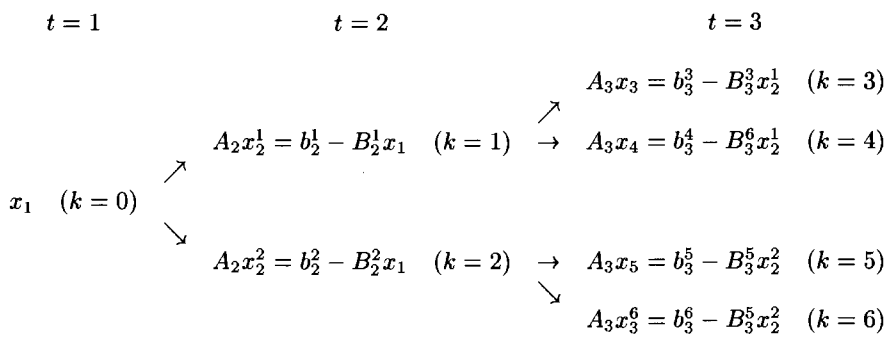

Figure 1

We have: $d_3(2) = \{5, 6\}$, $d_3(1) = \{3, 4\}$, $a(6) = 2$ $\quad x_{a(6)} = x_2^2$.

Since, at each stage t and for each outcome of the random vector $\tilde{\omega}_t$, we have a two-stage stochastic programming problem, the L-shaped method discussed earlier can be used to solve the different subproblems; starting from $t = 1$ and working sequentially towards $t = T$, the resulting overall method proceeds by first obtaining feasible solutions that are passed on as parameters to the next stage (forward pass), and then to moving towards optimality by working sequentially from $t = T$ down to $t = 1$ (backward pass).
That is, each period sends its optimal solution to the following period (except $t = T$), and its optimal cost to the preceding period where it is used to generate cuts.

So, using the L-shaped method, the corresponding master problems are:

- **For** $t = 1$:

$\min f^T x_1 + \Theta_1$
s.t.

$$\begin{aligned} A_1 x_1 &= b_1, \qquad (4.17)\\ \sigma_s^T B_1 x_1 &\geq \sigma_s^T b_s, \qquad (s = 1, \ldots, V(k_1))\\ \sum_{k \in d_2(1)} \bar{p}_k \pi_s^T B_1^k x_1^k + \Theta_1^k &\geq \sum_{k \in d_2(1)} \bar{p}_k \pi_s^T b_1^k, \qquad (s = 1, \ldots, J(k_1)). \end{aligned}$$

- **For** $t = 2, \ldots, T-1$:

$\min f^T x_t^{k_t} + \Theta_t^{k_t}$
s.t.

$$A_t x_t^{k_t} = b_t^{k_t} - B_t^{k_t} x_{a(k_t)}, \qquad (4.18)$$

$$\begin{aligned} \sigma_s^T B_t^{k_t} x_t^{k_t} &\geq \sigma_s^T b_s, \qquad (s = 1, \ldots, V(k_t)) \\ \sum_{k \in d_{t+1}(k_t)} \bar{p}_k \pi_s^T B_t^k x_t^k + \Theta_t^k &\geq \sum_{k \in d_{t+1}(k_t)} \bar{p}_k \pi_s^T b_t^k, \qquad (s = 1, \ldots, J(k_t)). \end{aligned}$$

- **For** $t = T$:

$\min f^T x_T^{k_T}$
s.t.

$$A_T x_T^{k_T} = b_T^{k_T} - B_T^{k_T} x_{a(k_T)}. \tag{4.19}$$

4.3.1 The nested decomposition for stochastic programming algorithm

The ideas presented in the previous sections can be implemented using the following algorithm:

Step 0:
For $t = 1$, set $k_t = \Theta_{k_t} = 0$, and $J(k_t) = V(k_t) = 0$ in the stage 1 master problem (4.17), and solve it.

Step 1:

If the stage 1 master problem is infeasible, then STOP. The multi-stage problem is infeasible.

Otherwise, for $t = 2$ and for all possible outcomes of the random vector, $\tilde{\omega}_2$, i.e. for $k_2 = 1, \ldots, K_2$, solve the corresponding master problem (where the solution obtained previously, is used as a parameter). The master problem is treated as a two-stage problem; a feasibility cut is added to the master problem in the case of infeasibility, and we go back to **Step 1**.

Step 2:
For $t = 2$ up to $t = T$, and for all possible outcomes of the random vector, $\tilde{\omega}_t$, having the same immediate ancestor as the current outcome, k_t, i.e. $\omega_t \in d_t(a(k_t))$, solve the corresponding master problem, where the appropriate ancestor solution, $x_{a(k_t)}$, is used as a parameter. Here, feasibility cuts are added to the master problem using principles similar to those presented in the case of two-stage problems, and $V(k_t)$ is updated accordingly.

Step 3:
For $t = T$ down to $t = 1$, and for all outcomes of the random vector $\tilde{\omega}_t$, having the same immediate ancestor as the current outcome (k_t), that is $\omega_t \in d_t(a(k_t))$, solve the corresponding master problem, where the appropriate ancestor solution, $x_{a(k_t)}$, is used as a parameter. Now, the Θ_{k_t}'s are no longer constrained to be equal to zero. Using the rules introduced

in the two-stage case, we add optimality cuts to the master problem, and update $J(k_t)$.

The optimal solution is found for $t = 1$.

Remark 4.1 At any stage, t, of the procedure and for any outcome, k_t, the optimal solution of the master problem is a lower bound, and the expected cost at this stage gives an upper bound of the expected costs of all scenarios descendant from this stage. When the difference between both values is small enough, the optimal solution of the master problem can be assumed to represent the future expected costs, and the master problem is said to be balanced with its descendants.

The method will yield an optimal solution after a finite number of steps.

Note that, like for the two stage case, techniques like bunching can be used to improve the efficiency of the algorithm.

In the case of unboundedness, we can set bounds on the value of the decision variables.

4.4 Monte Carlo importance sampling

For real life problems, the evaluation of the expected cost by direct summation is not always practical (because, for example, of multiple sums). The ideas contained in the above remarks (see remark 4.1 on this page) are combined with the Monte Carlo importance sampling method to produce a new solution technique (see [40] : 79-120)

4.4.1 The upper bound estimate

An upper bound for the total costs of the multi-stage problem is given by the expected costs of the first-stage problem.

Let $\hat{\omega}^1 = (\omega_2, \omega_3, \ldots, \omega_T)^T$ be a possible outcome of the random vector $\tilde{\omega}$; the vector (path) $\hat{\omega}^1$ may be obtained using the Monte Carlo importance sampling method (this process is called the path-sampling procedure). Using this information, we can find, $\hat{x}_1$, the optimal solution of the first stage subproblem and the resulting total cost. Similarly, we can find $\hat{x}_t$ (for $t = 2, \ldots, T$) using $\hat{x}_{t-1}$ as parameter in the subproblem at stage t and $\hat{w}^t = (\omega_{t+1}\omega_{t+2}, \ldots, \omega_T)^T$ as the future state of nature from stage t. The resulting total cost given scenario $\hat{\omega}^1$ is

$$C(x(\hat{\omega}^1)) = \sum_{t=1}^{T} f_t^T \hat{x}_{t-1}^{\hat{\omega}^1}. \tag{4.20}$$

Hence, the expected total cost is

$$E[C(x(\tilde{\omega}^1))] = \sum_{\omega \in \Omega} C(x(\omega^1)) \tag{4.21}$$

$$= \sum_{\omega \in \Omega} \sum_{t=1}^{T} f_t^T \hat{x}_{t-1}^{\hat{\omega}^1}. \tag{4.22}$$

The importance sampling scheme is used to find $E[C(x(\tilde{\omega}^1))]$ as in section 3.6.2.

4.4.2 The lower bound estimate

The cost of scenario $\hat{\omega}^1$ obtained earlier

$$C(x(\hat{\omega}^1)) = \sum_{t=1}^{T} f_t^T \hat{x}_{t-1}^{\hat{\omega}^1}, \tag{4.23}$$

is a lower bound of the total cost of the multi-stage problem.

4.4.3 Importance sampling for multi-stage problems: Algorithm

The corresponding algorithm is described below (see [40] : 118-119):

Step 1:
Solve stage 1 problem. The solution obtained is the trial solution, $\hat{x}_1$, and the resulting cost is a lower bound estimate of the expected cost of the multi stage problem, $\hat{LB}$.

Step 2:
Using the path-sampling procedure, we sample a number of paths and calculate an upper bound, $\hat{UB}$ of the expected cost of the multi-stage problem.

> **If** $\hat{UB} - \hat{LB} \leq TOL$ (TOL= tolerance level) then $\hat{x}_1$ is the optimal solution to the multi-stage problem. STOP.
>
> **Otherwise** go to **Step 3**.

Step 3:
Starting at stage T-1 and moving backwards until time $t = 1$, and using ω_1 obtained during the path-sampling procedure, and $\hat{x}_t^{\hat{\omega}^t}$ as ancestor solution, we take the corresponding subproblem as master problem and compute cuts by sampling again ω^{t+1} descendant subproblems. This step is repeated until an acceptable solution is obtained for each scenarios ω_t in stage t. Arriving at stage 1, a new solution, $\hat{x}_1$ and a new lower bound are found.
Go to **Step 2**, and repeat the process until $\hat{UB} - \hat{LB} \leq TOL$.

4.5 EVPI-based importance sampling

The expected value of perfect information, EVPI, is (see [5])

$$EVPI = Z_{HN} - Z_{WS}, \tag{4.24}$$

where Z_{HN}, the optimal cost obtained using the "here-and-now" approach, is

$$Z_{HN} = \min_{x_1} \Big\{ f_1(x_1) + E_{\omega_2}\Big[Q_2(x_1, \omega_2)\Big] \Big\}. \tag{4.25}$$

Z_{WS}, the optimal cost obtained using the "wait-and-see" approach ,is

$$Z_{WS} = E_{\omega_2}\Big[\min_{x_1} \Big\{ f_1(x_1) + Q_2(x_1, \omega_2) \Big\} \Big]. \tag{4.26}$$

At any time $t = 2, \ldots, T-1$, the local EVPI is

$$EVPI_t(\omega_t) = Z_{HN_t}(\omega_t) - Z_{WS_t}(\omega_t). \tag{4.27}$$

The EVPI which represents the level of stochasticity of a stochastic programming problem, can be used to decide whether the random variables of the problem may be replaced by their expected values or not (i.e. if the problem is to be considered stochastic or not).
In the multi-stage case, the same idea can be exploited. Indeed, if at each time t and for each possible future state of nature (i.e. each outcome of the random vector at each time t), the EVPI of the resulting stochastic programming problem (i.e. local EVPI) is evaluated, problems with low level of stochasticity will be identified and replaced by a deterministic one (the new problem is obtained by replacing random variables by their expected values). Therefore, more time and resources will be allocated to the other problems (i.e. problems with high level of stochasticity). The result is a refinement of the whole resolution process. To illustrate what has just been explained consider the following scenario tree:

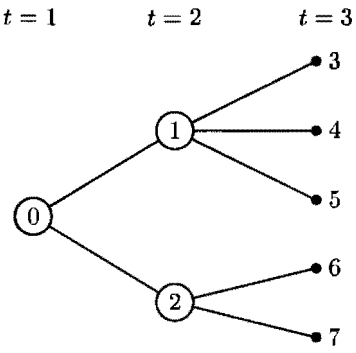

Figure 2

This tree represents all the possible states of nature from $t = 1$ up to time $t = 3$. It is clear that the level of stochasticity at node 1 is higher than the one at node 2; there are more paths originating at one 1 than there are paths leaving node 2. Furthermore, we know (this is also apparent on the tree) that as time passes, more information becomes available; that is, the level of stochasticity decreases with time. These observations lead to the following remarks:

- If the EVPI at a node is zero, all its descendant nodes have zero EVPI.
- The EVPI of a node whose descendants have no branching, is zero. A consequence is that, at time T, all nodes have zero EVPI.
- If the EVPI of a node is non-zero, all its ancestors have non-zero EVPI.

From the above discussion, it appears that, while the second stage realisation of the random variable in the case of a two-stage stochastic programming problem can be sampled using a single joint probability density function, it is not a reasonable choice for sampling in the multi-stage case (see [17]).

In what follows, we will describe how the information on local EVPI can be used as the importance sampling criterion for the sampling procedure (see [17]). The result will be a sampling procedure different from the one presented earlier.

In general, the sampling method employed replaces the scenario tree T, representing the possible data paths, by a sequence of sampled subtrees $T_1, \ldots, T_N$; a subtree T_n is obtained by increasing the number of sampled branches from a node of T_{n-1} with high EVPI, and decreasing or resampling the branches of a node with negligible EVPI. As n increases, the subtrees will become more representative of T.

At iteration n, one of the following may occur:

- At a node whose local EVPI is not negligible (i.e. they are not below a given tolerance level) and there is no branching from its descendants in T_{n-1}, the tree is expanded to a depth of one or more stages. The expansion is done by independent sampling conditional on the appropriate nodal data. The resulting subtree (i.e. T_n) will have more branches.
- At a node where the local EVPI is negligible, since the true local EVPI may not be negligible (though that may be the case), the subtree generated from it is resampled.
- If after a few resampling iterations the local EVPI at a node is fixed at zero (or is negligible), we assume that the true local EVPI is negligible; we will, therefore, assume that it is sufficient to represent the local problem by a single scenario from that node. Note that, as discussed earlier, the random variables of the local problem may also be replaced by their expected values.

Once these points have been investigated and the necessary actions taken, we have a new tree T_n. The process is repeated until the resulting tree is assumed to be reasonably representative of T (or just until we have reached a given number of iterations.)

4.6 Scenario analysis

4.6.1 The scenario aggregation method

An other decomposition method used to solve multi-stage problems is scenario aggregation or scenario analysis.

This method uses a finite number of scenarios (see [56]). Each scenario represents a possible state of nature for the given planning period, and has a known probability of occurrence. Furthermore, each scenario will yield a different multi-stage subproblem. The resulting overall multi-stage problem seeks to optimise the expected objective values of this collection of problems.

As earlier, we consider a planning period with horizon T. At each time $t = 2, \ldots, T$ an outcome of the random vector, $\tilde{\omega}_t$, is observed. A vector $s = (\omega_2^s, \ldots, \omega_T^s)^T$ whose components represent a realisation of $\tilde{\omega}_t$ for $t = 2, \ldots, T$, is called a scenario; s represents a possible state of nature for the whole planning period.
Consider the set of scenarios $S = \{s_1, \ldots, s_N\} \subset \Omega$; for every element of S, the resulting multi-stage subproblem is

$$\min f_s(X(s)) = \left\{ f_1(x_1^s) + f_2(x_2^s) + \ldots + f_T(x_T^s) \right\}$$

s.t.

$$\begin{array}{llll} A_1 x_1 & & = & b_1, \\ B_2(\omega_2^s) x_1 \quad A_2 x_2^s & & = & b_2(\omega_2^s), \\ \qquad\qquad \ddots & & \vdots & \\ \qquad\qquad\qquad B_T(\omega_{T-1}^s) x_T^s & & = & b_T^s(\omega_T^s), \\ x_i^s \quad \geq \quad 0, & & & \end{array} \tag{4.28}$$

or

$$\begin{aligned} &\min f_s(X(s)) \\ &\text{s.t.} \\ &\quad X(s) \in C_s \subset R^n. \end{aligned} \tag{4.29}$$

The solution of problem (4.28), i.e. $X(s) = (x_1^s, \ldots, x_T^s)^T$, is sometimes called an admissible solution.

The solution obtained for each $t = 2, \ldots, T$ must be consistent with the state of nature; i.e. if the two scenarios up to time $t = 3$ are $s^1 = (1,3,6)^T$ and $s^2 = (1,3,8)^T$, then we must have $x_1^1 = x_1^2$ and $x_2^1 = x_2^2$ but $x_3^1 \neq x_3^2$; two scenarios having a common history up to time t, must also have identical solutions for that time interval. This constraint is known as nonanticipativity requirement. The policies (solutions) satisfying the nonanticipativity requirement are said to be implementable.

The purpose of this method is to find an admissible policy which is implementable and optimises the overall multi-stage problem. In order to formalise these ideas, let us consider :

- A_t, the set of bundles of scenarios such that the scenarios in each bundle share a common history from time 1 up to time t. Of course, the elements of A_t, the $A's$, are disjoint and exhaust S; i.e.

$$\cup_{A \in A_t} A = S, \quad \text{for every} \quad t = 1, \ldots, T.$$

- The set of implementable policies,

$$\mathcal{N} = \{X \in R^n | x_t^s \quad \text{is constant on each bundle} \quad A \in A_t, \quad \text{and} \quad \forall s \in S\}.$$

- The set of admissible policies,

$$\mathcal{C} = \{X \in R^n | X(s) \in C_s, \quad \text{for every} \quad s \in S\}.$$

We will assume that the probability of occurrence of scenario s is $p_s > 0$, and we have $\sum_{s \in S} p_s = 1$.

If for all $s \in S$, we introduce the functions

$$F_s(X(s)) = \begin{cases} f_s(X(s)) & \text{if} \quad X(s) \in C_s, \\ +\infty \quad \text{otherwise.} \end{cases}$$

We have the overall multi-stage problem

$$\begin{aligned} &\min F(X) = \sum_{s \in S} p_s F_s(X(s)) \\ &\text{s.t.} \\ &\quad X \in \mathcal{N}. \end{aligned} \tag{4.30}$$

The set $\mathcal{C}$ is known since it is defined by the constraints of the individual scenario subproblem. However, we do not have an explicit definition of $\mathcal{N}$. A way out of this situation is to determine, for every t, the weighted average

$$x_t(A) = \frac{\sum_{s \in A} p_s x_t^s}{\sum_{s \in A} p_s}, \tag{4.31}$$

for every bundle $A \in A_t$.

Using this average, we define an element of $\mathcal{N}, \hat{x}_t^s$ by $\hat{x}_t^s = x_t(A)$, for every $s \in A$. In other words, we have found a transformation, $J : R^n \to \mathcal{N}$ defined by $J(X) = \hat{X}$.

Note that, though $\hat{X}$ is guaranteed to be in $\mathcal{N}$, we are not sure that it will be in $\mathcal{C}$. We, therefore, need to modify the constraint of problem (4.30); the new problem is

$$\begin{aligned} &\min F(X) = \sum_{s \in S} p_s F_s(X(s)) \\ &\text{s.t.} \\ &\quad X \in \mathcal{C} \cap \mathcal{N}. \end{aligned} \tag{4.32}$$

The constraint $X \in \mathcal{N}$ can be reformulated as $X = JX$ (since J is a linear transformation) or $KX = 0$, where $K = I - J$ (I is the identity operator). Problem (4.32) becomes

$$\begin{aligned} &\min F(X) = \sum_{s \in S} p_s F_s(X(s)) \\ &\text{s.t.} \\ &\quad KX = 0, \\ &\quad X \in \mathcal{C}. \end{aligned} \tag{4.33}$$

The transformation $K = I - J$ is a projection of R^n onto the subspace $\mathcal{M}$, defined by

$$\begin{aligned} \mathcal{M} &= \{W \in R^n | JW = 0\} \\ &= \{W \in R^n | E[W_t(s)] = 0 \quad \text{for every} \quad A \in A_t, t = 1, \ldots, T\}. \end{aligned}$$

Note that $\mathcal{M}$ is the complement of $\mathcal{N}$ with respect to the transformation J.

Therefore, if we introduce the operator $< \cdot, \cdot >$ defined by

$$\begin{aligned} < X, Y > &= E[X(s), Y(s)] \\ &= \sum_{s \in S} p_s X(s) Y(s) \end{aligned} \tag{4.34}$$

(this operator is slightly different from the actual inner product),

then the constraint $KX = 0$ is equivalent to $< X, W >= 0$.
Hence, problem (4.33) becomes

$$\begin{aligned} &\min F(X) = \sum_{s \in S} p_s F_s(X(s)) \\ &\text{s.t.} \\ &\quad < X, W >= 0, \\ &\quad X \in \mathcal{C}. \end{aligned} \tag{4.35}$$

This problem is solved using the augmented Lagrangian method (see [47] : 423-454); its augmented Lagrangian is the function

$$L_c(X,W) = F(X) + < X,W > + \frac{1}{2}c\|KX\|^2, \tag{4.36}$$

and the resulting problem is

$$\begin{aligned} &\min L_c(X,W) \\ &\text{s.t.} \\ &\quad X \in \mathcal{C}. \end{aligned} \tag{4.37}$$

Solving problem (4.37) will give us a good approximation to the optimal solution of problem (4.35). However, to solve problem (4.37) may not be an easy task; indeed, while the expressions $F(X)$ and $< X,W >$ are separable (with respect to the scenarios), it is not the case for $\|KX\|^2$.

The difficulty resulting from the inseparability of $\|KX\|^2$ lies in the fact that, using the augmented Lagrangian method, we minimise $L_c(X,W)$ with respect to X for a given W and repeat the process for different values of W. At each iteration we need to compute $L_c(X,W)$. Hence, the simpler its components the better.

The progressive hedging algorithm uses a modification of $L_c(X,W)$ to solve the problem. The idea behind this algorithm is first to replace the expression $\|KX\|^2$ by a separable one. To understand how it is done, suppose that we know an element of $\mathcal{N}$, say $\hat{X}^k$, and that $\hat{X}^k$ is not too far away from JX. If these suppositions are true, it will mean that X is close to $\hat{X}^k$ (since $\hat{X}^k = J\hat{X}^k$). Hence, we can replace KX, which is $X - JX$, by $X - \hat{X}^k$. We, therefore, obtain a new expression for the Lagrangian,

$$L_c(X,W) = F(X) + < X,W > + \frac{1}{2}c|X - \hat{X}^k|^2, \tag{4.38}$$

which is separable, and can be minimised by minimising the Lagrangian of the individual scenario subproblems

$$L_c^s(X,W) = f_s(X(s)) + < X(s),W(s) > + \frac{1}{2}c|X(s) - \hat{X}^k(s)|^2, \quad \forall X(s) \in C_s, \tag{4.39}$$

that is,

$$\begin{aligned} &\min \left\{ f_s(X(s)) + < X(s),W(s) > + \frac{1}{2}c|X(s) - \hat{X}^k(s)|^2 \right\} \\ &\text{s.t.} \\ &\quad X \in C_s. \end{aligned} \tag{4.40}$$

4.6.2 The progressive hedging algorithm

The solution of (4.35) is obtained using the following algorithm (see [56]):

Step 0:
Initialise $\gamma = 0$. For every scenario $s \in S$, solve the scenario subproblem (4.29).
Let $X^0(s)$ be the solution for scenario s, and set $W^0 = 0$.

Step 1:
Compute $\hat{X}^k = JX^k$.

Step 2:
For every scenario $s \in S$, solve the subproblem:

$$\min \left\{ f_s(X(s)) + < X(s), W^k(s) > + \frac{1}{2} c|X(s) - \hat{X}^k(s)|^2 \right\}$$
s.t.
$$X \in C_s. \tag{4.41}$$

If X(s) is an optimal solution to the problem, then set $X^k(s) = X(s)$.

Step 3:
Let $W^{k+1} = W^k + cKX^{k+1}, k = k + 1$, and go to **step 2**.

Remark 4.2 The solution is improved at each iteration of the algorithm, and the sequence $\{X^k, k = 0, 1, 2, \ldots\}$ converges to the solution X^* (for a proof see, for example, [58]).

The rate of convergence will depend on c. Therefore, small values of c will speed up the process, while large values will an opposite effect.

4.6.3 Postoptimality analysis: The contamination technique

Definition of the problem

Consider the problem

$$\min f(x, P)$$
s.t.
$$x \in X \subset R^n. \tag{4.42}$$

It is assumed that, $f(., P)$ is convex and $f(x, .)$ is linear and X is a convex set. P is a discrete probability function assigning probability p_s to the scenario s.

The scenarios and their respective probabilities may be obtained from different sources. However, since the set S does not necessarily represent the universe

of all possible scenarios, one is interested in both the robustness of the solution obtained by solving problem (4.42) and the optimal objective value.

Hence, the scenario generation procedure should be such that small perturbations of both the probability function P and the set S, alter the outcome only slightly so that the new results remain close to the original ones. We also expect larger perturbations not to yield a catastrophe. The contamination technique investigates the effects of these perturbations and gives corresponding error bounds (see [18], [19], [20]).

The contamination technique

Let

- X be a nonempty convex closed set of feasible solutions and X does not depend on P,
- $\phi(P)$ be the minimal value of the objective function,
- $X(P)$ be the set of optimal solutions,

and consider the family of probability functions P_λ, of the form

$$P_\lambda = (1-\lambda)P + \lambda Q. \tag{4.43}$$

where Q is a given fixed probability function, and $\lambda \in [0,1]$.

P_λ is obtained by contamination of the original probability function P by the probability function Q; for fixed density functions P and Q, the contaminated density P_λ depends only on λ.
Hence, using P_λ instead of P in problem (4.42), we obtain a family of optimisation problems parameterised by λ, and

$$f(x, P_\lambda) = f_Q(x, \lambda) \tag{4.44}$$

is the corresponding objective function; $f(., P_\lambda)$ is convex and $f(x, .)$ is concave. The resulting stochastic programming formulation is

$$\begin{aligned} &\min f(x, P_\lambda) = f_Q(x, \lambda) \\ &\text{s.t.} \\ &\quad x \in X \subset R^n. \end{aligned} \tag{4.45}$$

The optimal value function and the set of optimal solutions of the perturbed problem (4.45) are

$$\begin{aligned} \phi(P_\lambda) &= \phi_Q(\lambda) \\ &= \inf_{x \in X} f_Q(x, \lambda), \end{aligned} \tag{4.46}$$

and

$$\begin{aligned} X(P_\lambda) &= X_Q(\lambda) \\ &= \operatorname{argmax}_{x \in X} f_Q(x, \lambda), \end{aligned} \tag{4.47}$$

respectively.

Information about persistence, stability and sensitivity for parametric problems similar to problem (4.45) can be deduced from any of the following claims:

- If we further assume that the set $X(P) = X_Q(0)$ is nonempty and bounded and $X(Q) = X_Q(1) \neq \emptyset$, then the function ϕ_Q is finite and concave on $[0, 1]$, continuous at $\lambda = 0$ and its value at $\lambda = 0$ is such that

$$\begin{aligned} \phi_Q(0) &= \min_{x \in X} f(x, P) \\ &= \phi(P). \end{aligned} \tag{4.48}$$

- If the function f_Q is jointly continuous with respect to x and λ, then its derivative with respect to λ at $\lambda = 0$ for all x from a neighbourhood, say $\tilde{X}$ of $X(P)$ exists, and if the difference quotient

$$\frac{1}{\lambda}[f_Q(x, \lambda) - f_Q(x, 0)] \tag{4.49}$$

converges uniformly in x on $\tilde{X}$ as $\lambda \to 0^+$, then the derivative of the optimal value of the problem (4.45) at $\lambda = 0^+$ is

$$\begin{aligned} \phi'_Q(0^+) &= \frac{d}{d\lambda}\phi_Q(0^+) \\ &= \min_{x \in X(P)} \frac{d}{d\lambda} f_Q(x, 0^+). \end{aligned} \tag{4.50}$$

Since $f(x, .)$ is linear, the function

$$f_Q(x, \lambda) = (1 - \lambda) f(x, P) + \lambda f(x, Q) \tag{4.51}$$

is also linear in λ, and the sequence of difference quotients is stationary for fixed x.

Hence, expression (4.50) becomes

$$\begin{aligned} \phi'_Q(0^+) &= \min_{x \in X(P)} [f(x, Q) - f(x, P)] \\ &= \min_{x \in X(P)} f(x, Q) - \phi_Q. \end{aligned} \tag{4.52}$$

Since $\phi_Q(\cdot)$ is concave on $[0, 1]$, we have the following inequalities

$$(1 - \lambda)\phi_Q(0) + \lambda\phi_Q(1) \leq \phi_Q(\lambda) \leq \phi_Q(0) + \lambda\phi'_Q(0^+), \quad \forall \lambda \in [0, 1], \tag{4.53}$$

which gives the bounds (upper and lower) on the perturbed optimal value function, and bounds on the relative change of the perturbed optimal value due to contamination is

$$\phi_Q(1) + \lambda\phi_Q(0) \leq \frac{1}{\lambda}[\phi_Q(\lambda) - \phi_Q(0)] \leq \phi'_Q(0^+), \ \forall\lambda \in [0,1]. \tag{4.54}$$

Both expressions, (4.53) and (4.54), can be expressed in terms of P and Q as follows:

$$(1-\lambda)\phi(P) + \lambda\phi(Q) \leq \phi(P_\lambda) \leq \phi(P) + \lambda\phi'_Q(0^+), \ \forall\lambda \in [0,1], \tag{4.55}$$

and

$$\phi(Q) - \phi(P) \leq \frac{1}{\lambda}[\phi(P_\lambda) - \phi(P)] \leq \phi'_Q(0^+), \ \forall\lambda \in [0,1], \tag{4.56}$$

respectively.
Note that, if problem (4.42) has a unique solution, say $x(P)$, for all $\lambda \in [0,1]$, the expressions (4.52) and(4.55) become

$$\phi'_Q(0^+) = f(x(P), Q) - \phi(P), \tag{4.57}$$

and

$$(1-\lambda)\phi(P) + \lambda\phi(Q) \leq \phi(P_\lambda) \leq (1-\lambda)\phi(P) + \lambda f(x(P), Q), \ \forall\lambda \in [0,1] \tag{4.58}$$

respectively.

Suppose, now, that the set of optimal solutions of problem (4.45)
for $\lambda = 1$, $X(Q) = X_Q(1)$, is nonempty and bounded, then similar bounds can be derived for the optimal value $\phi(P_\lambda)$ using the solution $x(Q)$ and the optimal value $\phi(Q)$ of $\min_{x\in X}$ f(x,Q); that is

$$(1-\lambda)\phi(P) + \lambda\phi(Q) \leq \phi(P_\lambda) \leq \lambda\phi(Q) + (1-\lambda)f(x(Q), P), \ \forall\lambda \in [0,1] \tag{4.59}$$

so that,

$$\begin{aligned}(1-\lambda)\phi(P) + \lambda\phi(Q) \leq \phi(P_\lambda) \leq \min\{&(1-\lambda)\phi(P) + \lambda f(x(P), Q), \lambda\phi(Q) \\ &+(1-\lambda)f(x(Q), P)\}, \ \forall\lambda \in [0,1].\end{aligned}$$

Remark 4.3 The derivation of the bounds relies only on the assumed properties of the function $f(x,.)$.

In the case where problem (4.42) has multiple optimal solutions, the corresponding bounds can be computed at any one of the solutions. The resulting bounds will be valid but not necessarily tight.

The derivatives of the optimal value of problem (4.45) can be used to derive further characteristics of robustness.

If the expectation of the random parameter, $\tilde{\omega}$, is to be preserved, then the contamination is done using a probability function Q such that $E_Q[\tilde{\omega}] = E_P[\tilde{\omega}]$.

The extension of this method to multi-stage problems can be found in Dupačová (see [19]). The case where the objective function is nonlinear in P can be found in Dupačová (see [20]).

The contamination technique (which is derived from parametric programming) discussed here is not the only tool available for optimality analysis; a statistical approach based on statistical properties of $X(P), \phi(P)$ and those of their estimate, can also be used.

4.7 Stochastic quasi-gradient (SQG) methods

The different solution techniques discussed earlier exploit the special structures and properties exhibited either by the objective function or by the set of constraints. For problems having the same features or more complicated ones, the SQG methods can be used to find the solution.

4.7.1 General Idea

Let us consider the following stochastic programming problem

$$\begin{aligned} &\min \quad F^0 \\ &\text{s.t.} \\ &\quad F^i(x) \leq 0, \quad i = 1, \ldots, m, \\ &\quad\quad x \leq 0, \end{aligned} \tag{4.60}$$

where $x \in R^n$, and $F^i(x) = E[f^i(x, \tilde{\omega})]$ is assumed to be convex (for the moment).

The main idea (see [24]) behind the SQG methods is to construct a sequence of approximate solutions $\{x^s\}$ using the following two-relations scheme

$$x^{s+1} = x^s - \rho_s \xi(s), \text{ and} \tag{4.61}$$

$$F^i(x^*) - F^i(x^s) \geq \gamma_i(s) + < F^i_x(x), x^* - x^s >, \tag{4.62}$$

where x^* is an optimal solution, and

$$\begin{aligned} F^i_x(x) &= E[f^i_x(x, \tilde{\omega})] \\ &= E[\frac{\partial}{\partial x} f^i(x, \tilde{\omega})] \end{aligned}$$

is a subgradient. The variable ρ_s is the step size. The vector $\xi(s)$, with components $\xi^i(s)$ (or $\xi^i \quad , i = 1, \ldots, m$), is the direction of the movement and is also

called a stochastic quasi-gradient.

However, since it can be difficult to calculate the functions $f^i(x)$ and their respective subgradients, the SQG will use their statistic estimate; i.e.

$$\eta_i(s) = \frac{1}{N_s}\sum_{k=1}^{N_s} f^i(x^s, \omega_k^s), \tag{4.63}$$

and

$$\xi_i(s) = \frac{1}{N_s}\sum_{k=1}^{N_s} f_x^i(x^s, \omega_k^s), \tag{4.64}$$

where N_s represents the size of a collection of independent samples ω_k^s $k = 1, \ldots, N_s$.

These estimate are obtained using information about the past history of the process (that is, $X^s = (x^0, \ldots, x^s)^T$, and, possibly, some other variables); we have

$$E[\eta_i(s)|X^s] = f^i(x^s) + \epsilon(s), \tag{4.65}$$

and

$$E[\xi_i(s)|X^s] = f^i(x^s) + \epsilon'(s), \tag{4.66}$$

where $\epsilon(s) \in R$, and $\epsilon'(s) \in R^n$.

Since the objective of the procedure is to improve the quality of the solution, x^s, at each step, we expect both $\epsilon(s)$ and $\epsilon'(s)$ to decrease as s goes to infinity, and, at some point, we must have $\epsilon(s) \to 0$ and $\|\epsilon'(s)\| \to 0$.

Hence, we obtain the following new scheme

$$x^{s+1} = x^s - \rho_s \xi(s), \text{ and} \tag{4.67}$$

$$\eta_i(x^*) - \eta_i(x^s) \geq < E[\xi^i|X^s], x^* - x^s > + \gamma_i(s), \tag{4.68}$$

where $\gamma_i(s) \to 0$ as $s \to +\infty$.

The formulas used for the computation of the estimate will depend on the nature of the problem. Suppose that we can calculate the values of $f^i(x, \omega)$; i.e.

$$\eta_i(s) = \frac{1}{N_s}\sum_{k=1}^{N_s} f^i(x^s, \omega^{sk}), \tag{4.69}$$

and we have

$$\begin{aligned} E[\eta_i(s)|X^s] &= \frac{1}{N_s}\sum_{k=1}^{N_s} f^i(x^s, \omega^{sk}) \\ &= F^i(x^s), \end{aligned} \tag{4.70}$$

then

- If the functions $f^i(x)$ have uniformly bounded second derivatives, then

$$\xi_i(s) = \sum_{j=1}^{N} \frac{f^i(x^s + \Delta_s e^j, \omega^{sj}) - f^i(x^s, \omega^{s0})}{\Delta_s} e^j, \tag{4.71}$$

and

$$E[\xi_i(s)|X^s] = F_x^i + \epsilon'(s), \tag{4.72}$$

where e^j is the unit vector on the j^{th}d axis, and Δ_sd is the mesh for the finite difference approximation of the gradient; N is the size of a collection of independent samples of ω^s.

- If the functions $F^i(x)$ are non-differentiable, then

$$\xi_i(s) = \sum_{j=1}^{N} f_x^i(x^s, \omega^{sk}), \quad \text{at} \quad x = x^s. \tag{4.73}$$

4.7.2 Application to the two-stage recourse problem

Let us consider the following stochastic problem:

$$\begin{array}{lrcl} \min & F_0 = f_1^T x_1 & + & E[Q(x_1, \xi(\tilde{\omega}))] \\ \text{s.t.} & & & \\ & A_{11} x_1 & = & b_1, \\ & x_1 & \geq & 0, \end{array} \tag{4.74}$$

where

For every realisation, ω, of the random vector $\tilde{\omega}$. The function $Q(x_1, \tilde{\xi})$, also called the recourse function, is defined as follows:

$$\begin{array}{lrcll} Q(x_1, \xi(\omega)) = \min\limits_{x_2} & f_2(x_2(\omega)) & & & \\ \text{s.t.} & & & & \\ & A_{21}(\omega)\hat{x}_1 + A_{22} x_2 & = & b_2(\omega) & \text{a.s.}, \\ & x_2 & \geq & 0. & \end{array} \tag{4.75}$$

The function F_0 is continuous and convex. The dual of problem (4.75) is

$$\begin{array}{l} \max \quad y(b_2(\omega) - A_{21}(\omega)\hat{x}_1) \\ \text{s.t.} \\ \quad y A_{22} \leq f_2, \end{array} \tag{4.76}$$

where $y \in R^{n_2}$, and $\hat{x}_1$ is the optimal solution of the first stage problem.

Hence,

$$\frac{\partial}{\partial x_1} Q(x_1, \xi(\omega)) = \Big\{ - vA_{21}(\omega) | v \in \text{argmax}_y \Big[y(b_2(\omega) - A_{21}(\omega)\hat{x}_1) | yA_{22} \leq f_2 \Big] \Big\},$$

and the estimate of the gradient of F_0 at x^s is

$$\xi^0(s) = f_1 - y^s A_{21}(\omega^s), \tag{4.77}$$

where ω^s is obtained by random sampling from Ω using P and

$$y^s \in \text{argmax}_y \Big[y(b_2(\omega) - A_{21}(\omega)\hat{x}_1) | yA_{22} \leq f_2 \Big].$$

Note that, we have

$$F^0(x) - F^0(x^s) \quad \geq \quad < E[\xi^0 | X^s], x - x^s > . \tag{4.78}$$

The resulting procedure for solving the two-stage recourse problem is

1. At iteration s, for a given x^s, observe a realisation, ω^s, of the random vector and determine the vector $\xi^s = (f_2^s, A_{21}^s, b_2^s)^T$.
2. Solve the dual problem (4.76) using ξ^s, and $\hat{x}_1 = x^s$ to find y^s.
3. Calculate the estimate

$$\xi^0(x) = f_1 - y^s A_{21}(\omega^s), \tag{4.79}$$

and find a new value for x^s

$$x^{s+1} = \max[0, x^s - \rho_s \xi^0(x)]. \tag{4.80}$$

The procedure is repeated until a satisfactory solution is obtained.

Chapter 5

Conclusion

Stochastic programming has been used to model financial problems for more than 20 years (see [11]).

More recent applications include the award winning model of Russel-Yasuda Kasai (see [13]). This model, designed for a Japanese insurance company, incorporates the random nature of both the liabilities and the investment returns in a multi-stage setting. The implementation of this model resulted in an additional income of $79 million for the company in the first two years of use.

Golub, Holmer, McKendall, Pohlman and Zenios (see [34]) report on a stochastic programming model for money management specially designed for fixed-income securities. This model hedges against related uncertainty (interest rates, risk premia, interest rate volatility and other uncertainty) and outperforms classical techniques (e.g. immunisation). They establish the superiority of multi-period stochastic models for active portfolio management over the dynamic use of a single-period stochastic model as well.

Furthermore, Andrea Beltratti, Andrea Consiglio and S.A. Zenios (see [4]) develop a stochastic programming model for the management of an international bond portfolio which incorporates interest rate risk in the local market and exchange rate volatility across markets. Having these two parameters in the same model is extremely important since the volatilities of the fixed-income markets (especially bond markets) are typically lower than the volatilities of exchange rates, and hence the currency risk of an international bond portfolio can offset any benefits from international diversification.

As a matter of fact, the theoretical advances, the availability of good commercial software and the affordability of parallel computing have resulted in an ever increasing use of stochastic programming in Finance.

However, despite promising advances, the solutions obtained using these models

Chapter 5

Conclusion

Stochastic programming has been used to model financial problems for more than 20 years (see [11]).

More recent applications include the award winning model of Russel-Yasuda Kasai (see [13]). This model, designed for a Japanese insurance company, incorporates the random nature of both the liabilities and the investment returns in a multi-stage setting. The implementation of this model resulted in an additional income of $79 million for the company in the first two years of use.

Golub, Holmer, McKendall, Pohlman and Zenios (see [34]) report on a stochastic programming model for money management specially designed for fixed-income securities. This model hedges against related uncertainty (interest rates, risk premia, interest rate volatility and other uncertainty) and outperforms classical techniques (e.g. immunisation). They establish the superiority of multi-period stochastic models for active portfolio management over the dynamic use of a single-period stochastic model as well.

Furthermore, Andrea Beltratti, Andrea Consiglio and S.A. Zenios (see [4]) develop a stochastic programming model for the management of an international bond portfolio which incorporates interest rate risk in the local market and exchange rate volatility across markets. Having these two parameters in the same model is extremely important since the volatilities of the fixed-income markets (especially bond markets) are typically lower than the volatilities of exchange rates, and hence the currency risk of an international bond portfolio can offset any benefits from international diversification.

As a matter of fact, the theoretical advances, the availability of good commercial software and the affordability of parallel computing have resulted in an ever increasing use of stochastic programming in Finance.

However, despite promising advances, the solutions obtained using these models

should always be considered with caution. The quality of the solutions will depend on the quality of the data. Bad predictions will result in inferior solutions.

References

[1] S. Arnold, K. Ord, and A. Stuart (1999), *Kendall's advanced theory of statistics. Volume 2A: Classical inference and the linear model*, Edward Arnold, London.

[2] Robert D. Arnott and Frank J. Fabozzi (eds) (1988), *Asset Allocation : A handbook of Portfolio Policies, Strategies and Tactics*, Probus Publishing Company, Chicago, Illinois.

[3] Vijay S. Bawa (1975), "Optimal rules for ordering uncertain prospects", *Journal of Financial Economics*, 2(1):95-121.

[4] Andrea Beltratti, Andrea Consiglio and S.A. Zenios (1999), "Scenario modelling for the management of international bond portfolios", *Annals of Operations Research*, 85:227-247.

[5] John R. Birge (1982), "The value of the stochastic solution in stochastic linear programs with fixed recourse", *Mathematical Programming*, 24:314-325.

[6] John R. Birge (1985), "Decomposition and partitioning methods for multistage stochastic linear programs", *Operations Research*, 33:989-1007.

[7] John R. Birge and J. B. Wets (1986), "Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse", *Mathematical Programming Study*, 27:54-102.

[8] John R. Birge, Francois Louveaux (1997), *Introduction to Stochastic Programming*, Springer, New York, NY.

[9] Tim Bollerslev, Ray Y. Chou, Kenneth F. Kroner (1992), "ARCH modeling in finance", *Journal of Econometrics*, 52:5-59.

[10] Gabriel Bonti, Jonathan Hosking and Dirk Siegel (2000), "Beyond the lognormal", *Risk*, 13(5):59-62.

[11] S.P. Bradley and D.B. Crane (1972), "A dynamic model for bond portfolio management", *Management Science*, 19:139-151.

[12] R. Brooks-Hill, C. Parkan, W.T. Ziemba (1974), "Calculation of investment portfolios with risk free borrowing and lending", *Management Science*, 21(2):209-222.

[13] D.R. Cariño, T. Kent , D.H. Meyers, C. Stacy, M. Sylvanus, A.L. Turner, K. Watanabe and W.T. Ziemba (1994), "The Russel-Yasuda Kasai Model: An asset/liability model for a Japanese insurance company using multistage stochastic programming", *Interfaces*, 24(1):29-49.

[14] Jon Danielson and Casper G. de Vries (1997), "Value -at-Risk and extreme returns", London School of economics, Financial Markets group Discussion Paper no. 273.

[15] G.B. Dantzig and Peter W. Glynn (1990), "Parallel processors for planning under uncertainty", *Annals of Operations Research*, 22:1-21.

[16] G.B. Dantzig and Gerd Infanger (1993), "Multi-stage stochastic linear programs for portfolio optimization", *Annals of Operations Research*, 45:59-76.

[17] M.A.H. Dempster and R.T. Thompson (1999),"EVPI-based importance sampling solution procedures for multi-stage stochastic linear programs on parallel MIMD architectures", *Annals of Operations Research*, 90:161-184.

[18] Jitka Dupačová (1990), "On statistical sensitivity analysis in stochastic programming", *Annals of Operations Research*, 30:199-214.

[19] Jitka Dupačová (1995), "Postoptimality for multi-stage stochastic linear programs", *Annals of Operations Research*, 56:65-78.

[20] Jitka Dupačová (1996), "Scenario-based stochastic programs: Resistance with respect to sample", *Annals of Operations Research*, 64:21-38.

[21] Jitka Dupačová, Marida Bertocchi and Vittorio Moriggia (1997), "Postoptimality for scenario based financial planning models with an application to bond portfolio management", in: *Worldwide asset and liability modeling*, William T. Ziemba and John M. Mulvey (eds), Cambridge University Press, Cambridge, 263-285.

[22] E.J. Elton and M.J: Gruber (1995), *Modern portfolio theory and investment analysis*, John Wiley, New York.

[23] P. Embrechts, C. Klüppelberg and T. Mikosch (1997), *Modelling extremal events for insurance and finance*, Springer, Berlin.

[24] Y. Ermoliev (1988), "Stochastic quasi-gradient methods", in: *Numerical techniques for stochastic optimization*, Y. Ermoliev and R. Wets (eds), Springer-Verlag, Berlin, 141-185.

[25] Y. Ermoliev and R. Wets (eds) (1988), "Stochastic Programming, an introduction", in: *Numerical techniques for stochastic optimization*, Springer-Verlag, Berlin, 1-32.

[26] Peter C. Fishburn (1977), "Mean-risk analysis with risk associated with below-target returns", *American Economic Review*, 67(2):116-126.

[27] Jonathan Fletcher and Joe Hillier (2001), "An examination of resampled portfolio efficiency", *Financial Analysts Journal*", 57(5), 66-75.

[28] Karl Frauendorfer (1988), "Solving SLP recourse problems with arbitrary multivariate distributions- The dependent case", *Mathematics of Operations Research*, 13(3):377-394.

[29] Karl Frauendorfer (1992), *Stochastic Two-Stage Programming*, Springer-Verlag, Berlin.

[30] Karl Frauendorfer (1996), "Barycentric scenario trees in convex multistage stochastic programming", *Mathematical Programming*, 75:277-293.

[31] S. Arnold, K. Ord, and A. Stuart(1999), *Kendall's advanced theory of statistics*

[32] Horand I. Gassman (1990), "MSLiP: A computer code for the multistage stochastic linear programming problem", *Mathematical Programing*, 47:407-423.

[33] Peter W. Glynn and Donnald L. Iglehart (1989), "Importance sampling for stochastic simulations", *Management Science*, 35(11):1367-1392.

[34] Bennett Golub, Martin Holmer, Raymond McKendall, Lawrence Pohlman, Stravos A. Zenios(1995), "A stochastic programming model for money management ", *European Journal of Operational Research*, 85:282-296.

[35] David Heath, Robert Jarrow and Andrew Morton (1992), "Bond pricing and the term structure of interest rates : A new methodology for contingent claims valuation", *Econometrica*, 60(1) :77-105.

[36] D. Hendricks (1996), "Evaluation of value-at-risk models using historical data", *Federal Reserve Bank of New York Economic Policy Review*, 2:39-69.

[37] B. M. Hill (1975), "A simple general approach to inference about the tail of a distribution," *Annals of Statistics*, 35:1163-1173.

[38] J. Hosking (1990), "L-moments: analysis and estimation of distributions using linear combinations of order statistics", *Journal of the Royal Statistical Society*, series B, 52:105-124.

[39] J.C. Hull (2000), *Options, futures and other derivatives*, Prentice Hall, New Jersey.

[40] Gerd Infanger (1994), *Planning under uncertainty: Solving large-scale stochastic linear programs*, Boyd Fraser publishing company, Danvers, Massachusetts.

[41] Robert A. Jarrow (1988), *Finance theory*, Prentice-Hall, Englewood Cliffs, NJ.

[42] J.D. Jobson, Bob Korkie (1981), "Putting Markowitz theory to work", *Journal of Portfolio Management* , 7(4):70-74.

[43] J.D. Jobson and Bob Korkie (1981), " Estimation for Markowitz efficient portfolios", *Journal of the American Statistical Association* , 75(371):544-554.

[44] P. Kall and S.W. Wallace (1994), *Stochastic Programming*, John Wiley, Chichester.

[45] K.S. Kelly and R. Krzysztofowicz (1997), "A bivariate meta-gaussian density for use in hydrology", *Stochastic Hydrology and Hydraulics*, 11:17-31.

[46] Yiannis A. Koskosidis and Antonio M. Duarte (1997), "A scenario-based approach to active asset allocation", *Journal of Portfolio Management*, 23(2):74-85.

[47] David G. Luenberger (1989), *Linear and nonlinear programming* (2^{nd} edition), Addison-Wesley Publishing Company, p. 89.

[48] Harry Markowitz (1952), "Portfolio selection", *Journal of Finance*, 7:77-91.

[49] Harry Markowitz (1952), "The optimization of a quadratic function subject to linear constraints", *Naval Research Logistics Quaterly*, 3:111-133.

[50] A. J. McNeil and R. Frey (1999), "Estimation of tail-related risk measures for heteroscedastic financial time series: An Extreme value approach", Mimeo, ETH Zurich.

[51] Richard O. Michaud (1989), "The Markowitz optimisation enigma: Is "optimised" Optimal?", *Financial Analysts Journal*, 45(1):31-42.

[52] Richard O. Michaud (1998), *Efficient asset management: A practical guide to stock portfolio optimization and asset allocation*, Harvard Business School Press, Boston, Massachusetts.

[53] John M. Mulvey, Daniel P. Rosenbaum, Bala Shetty (1997), "Strategic financial risk management and Operations Research", *European Journal of Operational Research*, 97:1-16.

[54] P.M. Pardalos (1997), "Optimization techniques for portfolio selection", in: *New Operational Approaches for Financial Modeling*, Constantin Zopounidis(ed.), Physica-Verlag, 19-33.

[55] A.F. Perold and William F. Sharpe (1988), "Dynamic strategies for asset allocation", *Financial Analysts Journal*, Jan-Feb:16-27.

[56] A. Prekopa (1995), *Stochastic Programming*, Kluwer Academic Publishers.

[57] Stephen M. Robinson (1991), "Extended scenario analysis ", *Annals of Operations Research*, 31:385-398.

[58] R. T. Rockafellar and R. J-B. Wets (1991), "Scenarios and policy aggregation in optimization under uncertainty", *Mathematics of Operations Research*, 16:119-147.

[59] A. D. Roy (1952), "Safety first and the holding of assets", *Econometrica*, 20(3):431-449.

[60] Jaap Spronk and Winfried Hallerbach (1997), "Financial modeling: where to go ? With an illustration for portfolio management", *European Journal of Operational Research*, 99:113-125.

[61] R. Van Slyke and R. Wets (1969), "L-shaped linear programs with application to optimal control and stochastic progamming", *SIAM Journal on Applied Mathematics*, 17:638-663.

[62] R. Wets (1983), "Stochastic programming: Solution techniques and approximation schemes", *Mathematical Programming: The State of the Art* ,A. Bachem,M. Grotschel, B. Koste (eds), Springer-Verlag, Berlin, 566-603.

[63] R. Wets (1988), "Large-scale linear programming techniques in stochastic programming", in: *Numerical techniques for stochastic optimization*, Y. Ermoliev and R. Wets (eds), Springer-Verlag, Berlin.

[64] William T. Ziemba and John M. Mulvey (eds) (1998), *Worldwide Asset and Liability Modeling*, Cambridge University Press, Cambridge, UK.