Chapter 2

Decision Theory

2.1 Introduction

Everyday one faces problems which call for decisions.

Example 2.1. *At a summer evening, a person makes plans for the next day, and chooses between going to the beach or to go shopping. If next day is sunny, the beach is the best choice, but if the weather is less good, it is best to go shopping. How should the person decide what plans to make (if really having to plan ahead, already the evening before)?*

The difficulty, in the above example, is that the weather for tomorrow cannot be known, with complete certainty, so that it is a question of a *decision under uncertainty*.

Decision theory makes science of the art of making good decisions under uncertainty.

2.1.1 Decision Theory and Utility

In decision theory one defines a set a of *actions*, which one can choose to take. The result of the action depends on the value of the *state* x, which usually is random, $a = a(x)$.

Different states x appear with different *probabilities*. The corresponding *probability density function* is denoted $f(x)$, and may be *discrete* or *continuous*, corresponding to a discrete or continuous *state space*, respectively. It follows that the action $a = a(x)$ is a *random variable*.

Example 2.2. *A person has a choice between getting* 4.000.000 *SEK with certainty, or tossing a coin and get* 10.000.000 *SEK if the result is tails, and nothing otherwise.*

In this example, for the two actions a_1 *and* a_2*, we have*

$$a_1(x) = \begin{cases} 4.000.000 & \textit{for} \quad x = \textit{tails} \\ 4.000.000 & \textit{for} \quad x = \textit{heads} \end{cases} \qquad a_2(x) = \begin{cases} 10.000.000 & \textit{for} \quad x = \textit{tails} \\ 0 & \textit{for} \quad x = \textit{heads} \end{cases}.$$

The state space i $\{\textit{tails}, \textit{heads}\}$ *is discrete with probability density function* $f(\textit{tails}) = f(\textit{heads}) = \frac{1}{2}$.

To make scientific the concept of making a *good decision*, about what action to take, one has to specify the *utility* $U(a)$ of different actions a. The utility measures "how

much good" comes out of different actions.

As actions $a = a(x)$ are random variables, also utlities $U(a) = U(a(x))$ are random variables. The best action is the one that maximizes the expected utility

$$\mathbf{E}\{U(a)\} = \begin{cases} \int_x U(a(x))f(x)\,dx & \text{for continuous state space} \\ \sum_x U(a(x))f(x) & \text{for discrete state space} \end{cases}.$$

As common as it is to speak about utility, it is to speak about *risk*: The risk $R(a)$ measures "how much pain" comes out of different actions. Again the risk is a random variable. And the best action is the one that minimizes the expected risk

$$\mathbf{E}\{R(a)\} = \begin{cases} \int_x R(a(x))f(x)\,dx & \text{for continuous state space} \\ \sum_x R(a(x))f(x) & \text{for discrete state space} \end{cases}.$$

Example 2.2. (Continued) *If not very rich already, the person would probably choose to get* 4.000.000 *with certainty. However, if having a lot of money already, the second option could be argued to be a better choice, as the expected value of the gain,* 5.000.000 *SEK, is larger than for the first option.*
If not very rich, the utility could be $U_{poor}(a) = \sqrt{a/10.000.000}$*, while in the second case, the utility could be linear* $U_{rich}(a) = a/10.000.000$*. For the two actions* a_1 *and* a_2*, this give*

$$\mathbf{E}\{U_{poor}(a_1)\} = \sum_{x\in\{tails,heads\}} U(a_1(x)) = \sqrt{\frac{4.000.000}{10.000.000}}\cdot\frac{1}{2} + \sqrt{\frac{4.000.000}{10.000.000}}\cdot\frac{1}{2} = 0.63$$

and

$$\mathbf{E}\{U_{poor}(a_2)\} = \sum_{x\in\{tails,heads\}} U(a_2(x)) = \sqrt{\frac{10.000.000}{10.000.000}}\cdot\frac{1}{2} + \sqrt{\frac{0}{10.000.000}}\cdot\frac{1}{2} = 0.5,$$

whilst

$$\mathbf{E}\{U_{rich}(a_1)\} = \sum_{x\in\{tails,heads\}} U(a_1(x)) = \frac{4.000.000}{10.000.000}\cdot\frac{1}{2} + \frac{4.000.000}{10.000.000}\cdot\frac{1}{2} = 0.4$$

and

$$\mathbf{E}\{U_{rich}(a_2)\} = \sum_{x\in\{tails,heads\}} U(a_2(x)) = \frac{10.000.000}{10.000.000}\cdot\frac{1}{2} + \frac{0}{10.000.000}\cdot\frac{1}{2} = 0.5.$$

This quantifies that action a_1 *is better than* a_2*, if not very rich, while it is the other way around, if very rich.*

Often in practice, it is quite hard to find a correct utility U. However, when that has been done, the problem of finding the best action has been reduced to an optimization problem, to find the action a that maximizes $\mathbf{E}\{U(a)\}$.

One obvious general rule for utilities, is that if an action a_1 is preferred to another action a_2, then one should have $\mathbf{E}\{U(a_1)\} > \mathbf{E}\{U(a_2)\}$.

Concave utility functions are called *risk-averse*, convex ones *risk-seeking*, while linear utilities are *risk-neutral*.

Example 2.2. (Continued) *The utility U_{poor} is risk-averse, while the utility U_{rich} is risk-neutral.*

2.1.2 More Aspects on Decision Theory

Consider a problem of parameter estimation. In general, there are many methods to choose between.

A common choice of method for parameter estimation is the *maximum likelihood* method. As indicated by the name, the parameter estimates produced by this method are those that maximize the probability that the sample has occurred.

Now, if one is estimating parameters in order to be able to *forecast* an event, should one then use maximum likelihood? Well, the answer is that one should use a method which maximizes the precision of the forecast. Albeit this is quite obvious, it could be quite difficult to find the proper method.

When estimating parameters, one should always keep in mind the purpose of the estimate, and choose a method that is a good to that end.

Example 2.3. *Consider a linear model*

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

where ϵ is zero-mean normal distributed. Given a sample $\{(x_i, y_i)\}_{i=1}^n$ of (X, Y), the parameters β_0 and β_1 can be estimated, using the maximum likelihood, by

$$\hat{\beta}_0^{ml} = \bar{y} - \hat{\beta}_1 \bar{x} \quad \textit{and} \quad \hat{\beta}_1^{ml} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

The same estimators are obtained by minimizing the sum of squared deviations

$$\sum_{i=1}^n \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2.$$

If, given a value of x, we are interested in forecasting the corresponding value of $y = \beta_0 + \beta_1 x$, then, rather than using the forecast $y = \hat{\beta}_0^{ml} + \hat{\beta}_1^{ml} x$, the best thing would be to forecast $y = \hat{\beta}_0 + \hat{\beta}_1 x$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ minimize

$$|(\beta_0 + \beta_1 x) - (\hat{\beta}_0 + \hat{\beta}_1 x)|.$$

This is done by a so called crossvalidation *to minimize the sum of deviations*

$$\sum_{i=1}^n |y_i - (\beta_0 + \beta_1 x_i)|$$

The problem can now be solved by numerical methods, without further analysis.

A numerical algorithm to solve the problem could look like this:

```
smallest scorfunction = 1000000000
```

```
for beta0 in a intervall
  for beta1 in a intervall
    calculate scorefunction
    if scorefunction<smallest scorefunction
         smallest scorefunction=scorfunction
         best beta0=beta0
         best beta1=beta1
    end if
  end
end
```

2.2 Statistics of stocks

The world of finance is very often subjected to decisions. Consider how to select a portfolio of stocks where one can choose from virtually thousands of different stocks. This can of course be done by a stock broker who is an expert in a branch but in the long run one would like to complement a portfolio with stocks that gives the highes possible yield to the least possible risk.

2.2.1 Model

A stock can be considered as a non-stationary process which of course is really hard to model. Instead one consider returns (or logreturns) of the stock, i.e. the relative movement of a stock during a time period. Historically, returns are considered to be independent and normally distributed. The independence assumption comes from the fact that no one should be able to predict the stock market tomorrow and be able to make money without risk.

In mathematics, one can consider the stock market as n stocks, $S_1(t), \ldots, S_n(t)$, with returns(or logreturns) $X_1(t), \ldots, X_n(t)$ to be assumed independent in time and normal distributed. Here we have $X_1(t) = (S_1(t) - S_1(t-1))/S_1(t-1)$ or $X_1(t) = \log(S_1(t)/S_1(t-1))$. Using this probabilistic model is it straight forward to find an optimal portfolio and that is the goal of this laboration.

2.2.2 Model Check

In Project 1 we learnt how to check distribution assumptions with QQ-plots and e.g. the Kolmogorov-smirnov test. However, in mathematical finance it is also of importance to investigate the independence assumption. Remember that the definition of independence for two random variables X and Y is

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

but how do one test for independence in time if one has observations $X_1(0), X_1(1), \ldots, X_1(t_k)$? The answer is partly that one do not. Instead of testing independence one test linear dependance (correlation) which in the normal case is equivalent. Correlation between to random variables X and Y is defined as

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{E\{(X - E(Y))(Y - E(Y))\}}{\sqrt{var(X)var(Y)}}.$$

Remember from basic course: The correlation is always 0 for two independent random variables. The opposite is not true. Two random variables with 0 correlation can indeed be dependent. To see this, regard the following two random variables

$$X = \begin{cases} 1 & , \text{ with prob. } 1/3, \\ 0 & , \text{ with prob. } 1/3, \\ -1 & , \text{ with prob. } 1/3, \end{cases}$$

and $Y = X^2$. Then, X, Y are dependent by the definition of Y. Now, since $E(X) = 0$, the covariance is

$$\begin{aligned} Cov(X,Y) = E\{XY\} &= \sum_{i\in\{-1,0,1\}} \sum_{j\in\{0,1\}} ijP(X=i, Y=j) = \sum_{i\in\{-1,1\}} iP(X=i, Y=1) \\ &= \sum_{i\in\{-1,1\}} iP(X=i) = -\frac{1}{3} + \frac{1}{3} = 0, \end{aligned}$$

which means that the correlation is also 0. Dependent random variables can have 0 correlation. If we instead would have defined $Y = aX$, then $Cov(X,Y) = a^2 \cdot var(X)$. As stated above, covariance is a measure of linear dependence!

Now, if we continue with the returns from stock 1, $X_1(0), \ldots, X_1(t)$, we must consider the time aspect. In this case we will regard $X_1(0), \ldots, X_1(t)$ as observations from a realization of a stochastic process. We will then regard the *autocorrelation function* (acf), which is defined as

$$r_{X_1}(t,h) = \frac{Cov(X_1(t), X_1(t+h))}{\sqrt{var(X_1(t))\, var(X_1(t+h))}}.$$

Note that $r_{X_1}(t,t) = var(X_1(t))$. Under some assumptions of *stationarity*, the acf only depends on the time lag, h. This implies

$$r_{X_1}(h) = \frac{Cov(X_1(0), X_1(h))}{var(X_1(0))}.$$

We can estimate the acf by

$$\hat{r}_{X_1}(h) = \frac{(t+1)\sum_{i=0}^{t-h}(X_1(i) - \overline{X}_1)(X_1(i+h) - \overline{X}_1)}{(t-h+1)\sum_{i=0}^{t}(X_1(i) - \overline{X}_1)^2},$$

where

$$\overline{X}_1 = \frac{1}{t+1}\sum_{i=0}^{t} X_1(i).$$

The function $\hat{r}_{X_1}(h)$ is known as the sample auto correlation function.

Now, as one is interested to investigate linear independence one has to ask oneself how small the sample acf must be to be interpret as linear independence. It turns out that, under some assumptions, for independent data,

$$\lim_{n\to\infty} \sqrt{n}\hat{r}_{X_1}(h) \sim N(0,1),$$

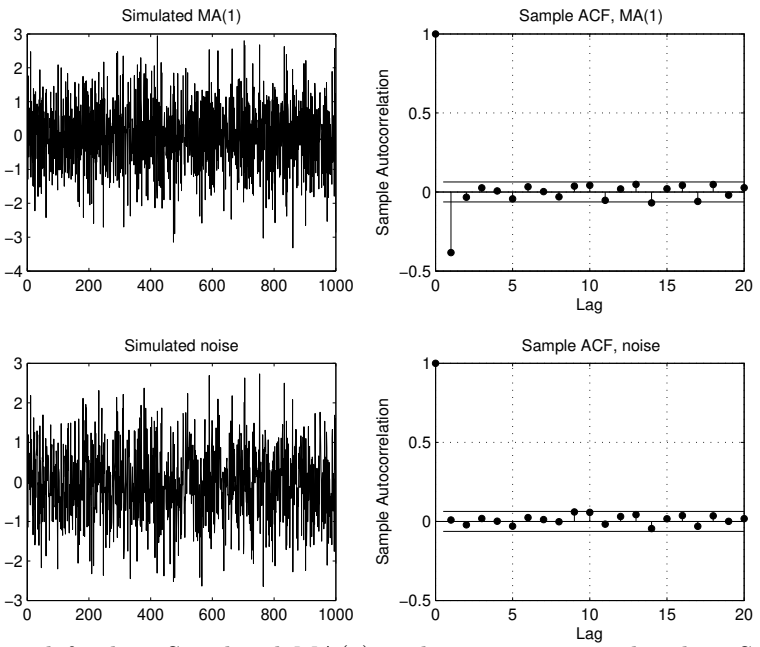

Figure 2.1: Top left plot: Simulated MA(1) with a=0.5. Top right plot: Sample ACF for simulated MA(1). Straight line is the 95% confidence interval for independence. Bottom left plot: Simulated white noise. Bottom right plot: Sample ACF for simulated noise. Straight line is the 95% confidence interval for independence.

This can be used to construct a 95% confidence interval for independent data. Hence, for n sufficient large we assume linear independence if

$$|\hat{r}_{X_1}(h)| \leq 1.96/\sqrt{n}.$$

Now, to investigate independence it is common to continue the analysis with estimation of $f(X_1(0)), \ldots, f(X_1(t))$, for instance with $f(x) = |x|$. If one still has no linear dependence for the transformed data one assumes independence.

Example 2.4. *Consider a process $X(t) = \epsilon(t) - a\epsilon(t-1)$ where $\epsilon(t)$, $t = 0, \ldots$ are iid. This process is called a Moving Average process of degree 1 (MA(1)). Clearly, we will have a dependence between $X(t)$ and $X(t+h)$ for $h = 1$, otherwise we have independence. In Figure 2.1 we have simulated the process with noise modelled as $\epsilon(t) \sim N(0,1)$ and a=0.5 and estimated the sample acf. We have also plotted the noise and its sample acf. Although the processes looks similar they have very different dependence structure.*

Example 2.5. *Consider a process $X(t) = \epsilon(t) + aX(t-1)$ where $\epsilon(t)$, $t = 0, \ldots$ are iid. This process is called an Auto Regressive process of degree 1 (AR(1)). Clearly, we will have decreasing dependence between $X(t)$ and $X(t+h)$ for $h > 1$. In Figure 2.2 we have simulated the process with noise modelled as $\epsilon(t) \sim N(0,1)$, a=0.5 and X(0)=0 and estimated the sample acf. We have also plotted the noise and its sample acf. Although the processes looks simular they have very different dependence structure.*

2.3 Mathematica

Mathematica is a very versatile mathematical program package. As in virtually any program, there is a helpbrowser, which is activated with punching *help*: One should get

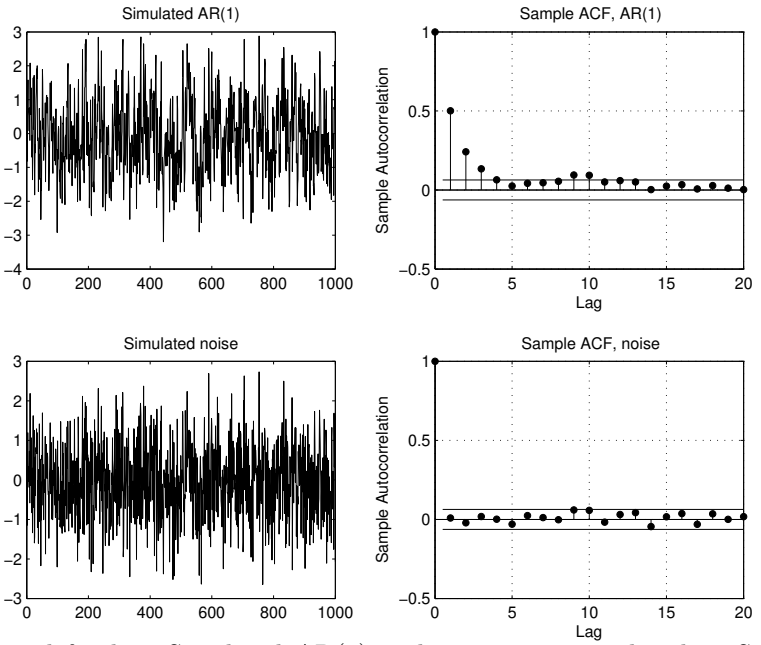

Figure 2.2: Top left plot: Simulated AR(1) with a=0.5. Top right plot: Sample ACF for simulated AR(1). Straight line is the 95% confidence interval for independence. Bottom left plot: Simulated white noise. Bottom right plot: Sample ACF for simulated noise. Straight line is the 95% confidence interval for independence.

used to make use of this helpbrowser!

All Mathematica commands begin with a capital letter. Arguments of Mathematica functions are given within brackets, for example, the sample mean of `data` is calculated with `Mean[data]`. A command is executed with *shift+return*.

Mathematica have many libraries, that have to be loaded to be accesable. For example, the command `<<Statistics‘DescriptiveStatistics‘` loads the library for desciptive statistics, that include the `Mean`-funtion.

Notice that if, for example, one use the function `Covariance` before loading the corresponding library with `<<Statistics‘MultiDescriptiveStatistics‘`, then the command `Covariance` is already “occupied” when one tries loads the library, which makes the command breakdown. (At least there is a risk for that.)

A very important tip is to always use the command `Clear[Symbol]`, before defining `Symbol`, in order to avoid that old conflicting definitions “disturb” new definition.

Mathematica can fabricate plots in *.eps/.ps format* in several ways: One way is to mark the image by left clicking on it, and right click in the frame that occurs around the image. Select *File* in the menu, and then *Print Selection*. If you do not want headers or footers on the image, they can be removed by clicking *File* followed by *Headers and Footers*, and then select *No header on first page* or *No footers on first page*, respectively. Another way is to name the plot, *name=Plot[...]*, and then just type *Export[“image.eps”,name]*.

Here is an example of a Mathematica program:

```
In[1]:= Clear[w]; w:={w1,w2,w3}
In[2]:= Clear[f]; f[w_]:=Sum[Sqrt[w[[i]]],{i,1,3}]
In[3]:= NMaximize[Flatten[{f[w],Table[w[[i]]>=0,{i,1,3}],Sum[w[[i]],
```

```
            {i,1,3}]==1}],w]
Out[3]= {1.73205,{w1->0.333333,w2->0.333333,w3->0.333333}}
```

which could also have been written

```
In[4]:= Clear[f]; f[w1,w2,w3]:=Sqrt[w1]+Sqrt[w2]+Sqrt[w3]
In[5]:= NMaximize[{f[w1,w2,w3],w1>=0,w2>=0,w3>=0,w1+w2+w3==1},
            {w1,w2,w3}]
Out[5]= {1.73205,{w1->0.333333,w2->0.333333,w3->0.333333}}
```

Also, if one would like to manipulate matrices, for instant to create logreturns from stock data the function *Table[...]* is very useful.

2.4 Laboration

2.4.1 Portfolio Optimization

A person has 7 different stocks available to create a portfolio. Of course, the person would like to earn as much money as possible with as low risk as possible. There exist time series from 2004-06-03 to 2006-06-01 for stock prices in the file *stockdata.tsv*. The stocks are AstraZenica, Electrolux, Ericsson, Gambio, Nokia, Swedish Match and Svenska Handelsbanken which are in column 2 to 8. The first column is a date vector.

This laboration has two parts. The first one is a statistical investigation and in the second part a portfolio optimization is performed. It is preferable to use matlab in the first part.

2.4.2 Model investigation

1. Play around with the stock data to check for the distribution of logreturns. Give standard statistics and interpret the dependence between the stocks in terms of economics.

2. Estimate the acf for logreturns of a few different stocks, together with confidence intervals for independent data. Do it also for squared logreturns. Conclusions? Try to interpret the findings in terms of economics.

2.4.3 Portfolio optimization

The following utility function is used

$$U(x) = 1 - \mathrm{e}^{-kx},$$

where $k > 0$ is a constant. The utility function can describe the persons relation to risk versus money.

1. Play around with the utility and describe its features, for different values of k.

2. Calculate, using Mathematica, the expected utility when the returns are jointly normal distributed, with different vectors of expected values and covariance matrices.

Notice that, under the above assumption, the portfolio becomes normal distributed, with expected value $z = \mu^T w$ and variance $\sigma^2 = w^t Q w$, where w is a vector describing the proportion of money is spent in different stocks, μ is the expected value of the joint normal distribution and Q its covariance matrix. This gives

$$\mathbf{E}\{U(w)\} = \int_{-\infty}^{\infty} (1 - \mathrm{e}^{-kx}) \, \frac{1}{\sqrt{2\pi}\,\sigma} \, \mathrm{e}^{-((x-z)^2/(2\sigma^2)} \, dx.$$

3. Estimate the meanvector and the covariance matrix, for the above mentioned joint normal distribution, using the data file *stockdata.tsv*.

4. We are faced with problem to maximize $\mathbf{E}\{U(w)\}$ under the constraints

$$w_1, \ldots, w_n \geq 0 \quad \text{and} \quad \sum_{i=1}^{n} w_i = 1.$$

Show that the problem is equivalent with the quadratic problem to maximize

$$\mu^T w - \frac{k}{2} w^t Q w.$$

5. The task now is to investigate how the selection of k alters the proportion of money in each stock. Use `NMaximize` in Mathematica to solve this problem. Why do the the weights alter as they do?

   It could be a good idea to try to solve the problem first for only two stocks, and then add more stocks when that simpler problem is solved.

6. Compare with a naive way of investing: What is the expected utility if dividing money equally in the stocks?

7. Remember the first task of the laboration. Discuss how the result from that part influences the optimal portfolio. Could one do it in another way?