# ON THE CLASSIFICATION AND ESTIMATION OF COSTS IN INFORMATION TECHNOLOGY

A dissertation by

Pétur Björn Thorsteinsson

30 ECTS credit thesis submitted to the Faculty of Engineering in partial fulfillment of the requirements for the degree of master of science in engineering.

Advisors
Kristófer Arnar Einarsson
Páll Jensson

Faculty Representative
Ebba Þóra Hvannberg

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science
School of Engineering and Natural Sciences
University of Iceland
Reykjavik, May 2010

On the classification and estimation of costs in information technology
30 ECTS credit thesis submitted in partial fulfillment of a M.Sc. degree in Financial Engineering



Faculty of Industrial Engineering, Mechanical Engineering and Computer Science
School of Engineering and Natural Sciences
University of Iceland
VR-II, Hjarðarhaga 2-6, IS-107 Reykjavik, Iceland
Phone +354 525 4700, Fax +354 525 4632
verk@hi.is
www.hi.is

Bibliographic information:
Pétur Björn Thorsteinsson, 2010, *On the classification and estimation of costs in information technology*, M.Sc. thesis, Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, 76 pages

HI-PHD: ISSN xxx-xxx

*To my newborn son, Petur Levi, and his wonderful mother, Agla Margret Egilsdottir.*

# Abstract

The complexity of IT in modern organizations is generally quite high and the costs of IT are correspondingly complex. An overview of available publications shows that there is relatively research done on the development of cost models for IT. The systematic gathering and classification of IT costs is often overlooked in organizations. The current thesis addresses many of the issues associated with the complex nature of IT costs, and proposes a cost model for IT. The necessary details of IT costs are defined in order to facilitate the collection of complete IT cost data. Categories for IT costs are proposed. The categorization is based on how modern IT is usually built up in organizations, with each category representing a distinct part of IT. Finally a method of estimating IT costs is demonstrated. The method uses simple probability densities to represent cost estimates. A simple example is shown that demonstrates that with only a few cost estimates using probability densities, the sum of the estimates will result in a distribution that closely follows the Gaussian distribution.

# Útdráttur

Flækjustig UT (upplýsingatækni) í nútima fyrirtækjum og stofnunum er almennt mjög hátt og kostnaðurinn af UT er samsvarandi flókinn. Yfirlit af útgefnu efni sýnir að það hafa fáar rannsóknir verið gerðar á þróun kostnaðarmódela fyrir UT. Kerfisleg söfnun og flokkun á UT kostnaði er almennt ekki gerð. Þessi ritgerð fer yfir mörg af þeim málum sem tengjast hinu flóknu eðli UT kostnaðar og leggur til kostnaðarmódel fyrir UT. Hin nauðsynlegu atriði varðandi UT kostnað eru skilgreind til að koma af stað söfnun á fullnægandi UT kostnaðarupplýsingum. Flokkar fyrir UT kostnað eru settir fram. Flokkunin er byggð á hvernig nútíma UT er byggt upp í samtökum, þar sem hver flokkur inniheldur kostnað úr aðskildum hliðum UT. Í lokin er kynnt aðferð til að meta UT kostnað. Aðferðin notar einföld líkindaþéttiföll til að setja fram kostnaðarmat. Einfalt dæmi er tekið sem sýnir að með aðeins fáum líkindaþéttiföllum, þá má leggja þau saman og fá útkomu sem er mjög nálægt Gauss dreifingunni.

# Contents

# List of Tables

# List of Figures

# Nomenclature

| | |
|---|---|
| capex | Capital expenditure |
| Cloud | Cloud (internet) based development and computing |
| ERP | Enterprise resource planning |
| FFT | Fast Fourier transform |
| Green IT | Movement towards more environmentally conscious IT |
| ICT | Information and communication technology |
| IT | Information technology |
| On premise | When IT systems are housed within companies |
| opex | Operating expenditure |
| OS | Operating system |
| ROI | Return on investment |
| SaaS | Software as a service |
| Server | A computer, usually in a data center. |
| SOA | Service oriented architecture |
| TCO | Total cost of ownership |
| Virtualization | Technology used to allow servers to run many operating systems simultaneously. |

# Acknowledgements

I would like to thank my spouse, Agla Margret Egilsdottir, for her invaluable support, encouragement and patience. I also owe great thanks to my parents, Birna Hreidarsdottir and Petur Gunnar Thorsteinsson, whose help and support has greatly aided me in this and all my other endeavors so far. Finally I want to give a special thank you to my family for being a constant source of motivation.

I would like to thank my instructors, Kristofer Arnar Einarsson and Pall Jensson, for their help. In addition I would like to acknowledge the help of my colleagues, Ellert Hlodversson, Thorsteinn Hallgrimsson and Arnthor B. Reynisson. Also I would like to thank Oddur Oskar Kjartanson for valuable input.

# Outline

**Introduction** An overview of the subject matter and motivation. Also a look at other literature on the subject is made.

**Chapter 1** This section will look at IT costs, make some clarifications on issues relating to cost, and define what costs will be discussed in this thesis. Uses of cost information will be presented, as well as information on acquiring costs. The chapter ends with information on the specific issues related to labor costs.

**Chapter 2** An overview and breakdown of IT infrastructure into different types of costs. Different categories for costs will be proposed, along with the rationale behind them. Data centers will be given special attention with regards to their capacity. Finally some information on doing data center comparisons.

**Chapter 3** The theoretical methods used are looked at. Probability densities that will be used to estimate costs are explained and shown graphically.

**Chapter 4** Using the categorization of costs from chapter 2, a cost estimation technique will be proposed. The cost estimates will be based on the probability distributions in chapter 3. Gathering together the cost estimates into one final estimate is explained.

**Chapter 5** An example of the usage of the cost categories and cost estimations is given

**Conclusion** Conclusion, further research, and limitations

**Appendix A** Matlab code for the example in chapter 5

# 1 Introduction

As an IT professional for the last five years, the author of this thesis has worked on a wide range of the tasks that are commonly delegated to IT departments of modern businesses. Those tasks include the sale, purchase, setup, maintenance and development of IT equipment and software for various companies. The technologies that the author has worked on range from small applications running on IP telephones to enterprise class software running on mainframes. The companies the author has worked with have been small businesses whose only computer infrastructure is the file server under the secretary's desk to businesses with large data centers where every IT system is duplicated on a second data center located in a separate geographical location.

As the economic downturn has intensified, businesses have increasingly focused on making their IT more efficient (Roberts and Sikes, 2009). In fact the assumption that investments in IT boost productivity at all has long been debated (Brynjolfsson, 1994), but IT investment has largely been been found to increase productivity (Cotteleer, 2002). The need to analyze and quantify current and planned IT setups is becoming ever greater in order to maximize the return on investment in IT.

For businesses to understand how their investments in IT are performing, many choose to find ways to assess the business value that their IT provides. This thesis will not delve into methods of how to ascertain the business value of IT. The subject of IT value is well researched and there are many publications with methods of how to determine the value of IT investments (such as (Harris et al., 2008),(Val, 2008)).

What most of this literature has in common though, is that the starting point for methods of value analysis is to discover the costs. Costs are in this respect almost an afterthought, while it is acknowledged that they should be known and accessible, methods to accomplish this are generally quite vague, but it is assumed that the costs of IT are known. There is a lack of literature that deals directly with IT costs (Golden, 2009).

The present thesis offers a methodology for aggregating and estimating costs that have or will be incurred. A starting point for a business to analyze its costs, group them, and maintain that grouping as new costs come along is proposed. The information gained can in turn be used in a wide array of analyses such as TCO, ROI, etc.[1]. The end result that will hopefully be achieved with the help of methods outlined in this thesis will be a system where an organization can readily get historical as well as current IT costs.

The thesis will begin by defining the costs that will be used and give an overview of the associated issues that can come up when doing cost calculations. Next the thesis will generalize a typical modern IT setup and apply a 'work breakdown structure' - a methodology used in project management. There an attempt will be made to divide IT into units in a logical manner. Finally a method for estimating costs in IT is proposed, where the estimations for costs are made using probability distributions. Some simple distributions for cost are proposed, and a way of calculating the total cost and it's uncertainty is shown.

## 1.1 Current research and literature on IT costs

There is a certain difficulty in assessing the amount of literature and research that exists on the subject of IT costs because the subject matter is so broad. There has been a great deal of research on costs associated with software development and pricing, books and papers on that subject are easily found. In addition there are many books on IT organization and governance that mention IT costs in passing but do not give a substantial account of IT costs.

### 1.1.1 Cost Accounting

When dealing with costs, it is natural to look to how costs are handled in modern accounting. There is in fact a field of accounting that is wholly concerned with costs. Managerial accounting is a field of accounting that provides economic and financial information to managers and other internal personnel (Weygandt et al., 2005). One of the main methods of managerial accounting is 'cost accounting'. Cost accounting is a "technique or method for determining the cost of a project, process, or thing..." (Barfield et al., 2003).

The research and literature of cost accounting contains a lot of information on industry specific methods of applying cost accounting. IT is an exception, as there is little information of how to apply cost accounting specifically to IT.

One of the few papers the author found that use cost accounting in IT is a paper on the application of activity based costing (one of the methods of cost accounting) to IT (Neumann et al., 2004). The paper deals with costs on a much more abstract level then is used in this thesis. Another paper the author found proposes a framework for using activity based costing

[1] For more information on these analysis, see chapter 2.2

to evaluate IT investments in emerging economies (Roztocki and Weistroffer, 2004). The paper describes how traditional costing techniques have increasingly become less reliable. The author did not find any research in cost accounting that could be applied to IT costs as they are presented in this thesis.

### 1.1.2 Academic literature

The economics of IT have been a subject of research ever since the adoption of the first mainframes in the 1950s. A review of the available literature reveals that it generally deals with the economics of IT on a macro or high level. Examples are various studies done relating performance indicators such as a company's stock price to it's IT investments (Bharadwaj et al., 1999), surveys of economic evaluations of IT (Mutschler et al., 2007), and relating IT investment to future firm performance (Anderson et al., 2003). While there exists a substantial amount of literature on the economics of IT, only a small portion deals directly with IT costs. This makes finding papers on IT costs difficult.

#### Authors

Notable authors on the financial aspect of IT are, Professor Dan Remenyi and Dr. Frank Bannister, both of Trinity College Dublin. They have both written a multitude of papers and books on the economic aspect of IT. A book authored by Remenyi and Bannister, along with Arthur Money, "The Effective Measurement and Management of ICT Costs and Benefits" (Remenyi et al., 2007) is considered by many to be one of the leading texts on IT costs and benefits. The book addresses a wide range of the issues related to economic evaluation of IT, and gives a good overview of the history of IT evaluation. There is also a book by Bannister, "Purchasing and Financial Management of Information Technology"(Bannister, 2004) which examines the practical aspects of IT finance.

Other notable authors in the field of IT economics are Professor Zahir Irani at Brunel University and Dr. Peter E.D. Love at Edith Cowan University. They have authored various publications on the subject of IT management and finance, and wrote a paper written on IT adoption in the construction industry in Australia that has a good survey of the various indirect or soft costs [2] of IT (Love et al., 2003). Irani and Love are also on of the editors of 'Evaluating Information Systems'(Irani and Love, 2008), where an overview of the literature on IT evaluation can be found.

#### Master's theses

There are relatively few master's theses that deal with IT costs. A thesis written in 2007 by Ken Huang named 'Towards An Information Technology Infrastructure Cost Model'(Huang, 2007). The thesis is written on how to find the cost drivers in IT infrastructure, and had

[2] See chapter 2.1.3 for more information on soft costs

also among its findings information that indicates how different sectors prioritize their IT.

### 1.1.3 Other literature

Publications by IT professionals and IT companies on the subject of IT costs are few. There are a great many books and papers on the subject of IT management and high level overviews of IT finance, there is not much dealing directly with IT costs.

A book written on IT cost accounting in 2008 gives a good overview of the research that has been done in the field, particularly in Germany. The book proposes a way method of estimating expected resource consumption in a data center in order to allocate infrastructure costs when the IT infrastructure is shared (Brandl, 2008).

### 1.1.4 Summary and relation to this thesis

The overall theme in most of the available literature on IT costs is to look at IT costs from a high level. The available material is of a very general nature, while this thesis takes a step further and delves into the actual categorizing and estimating of IT costs.

The author did not find any comparable research to what is presented in this thesis. The closest work to what is presented here is that which can be found in the books (Remenyi et al., 2007),(Bannister, 2004) but even there costs are dealt with on a higher level than they are here. There is some literature that suggests categorizing IT costs, but with few examples. The author did not find any literature that either shows how to estimate IT costs, or gives an example of how to systematically estimate IT costs.

# 2 IT and costs

Issues related to cost have always pervaded most businesses. The methods used to handle costs have evolved over time, and have to a certain extent been standardized, for example through modern accounting methods. Despite this, methods to track costs will vary greatly between businesses and industries since the needs are so various. The way that costs need to be measured in IT is often quite different from other parts of a business, which is why they deserve special consideration.

Organizations will often have a significant amount of their capital tied to IT. The purchase of IT will often constitute an expensive and long term commitment, and empirical analysis has shown that IT investment risk is higher than for many other investments (Dewan et al., 2007).

When measuring costs in IT, so many factors come into play that it is easy to get lost. In this chapter, considerations are pointed out that will be useful when recording costs. Cost is a subjective term that can refer to very different things, so the term will be narrowed down. Systems for cost gathering will be discussed, and the more common IT evaluations based on cost will be listed.

The collection of IT cost information is rarely a simple undertaking, and will often result in incomplete information. As is the case with all information - getting more detailed information needs more work - this is especially relevant with IT costs since the personnel that are capable of collecting the cost data needed are often expensive. This is why IT cost granularity will always be constrained by how much a company is willing to invest in the gathering of IT costs.

## 2.1 Costs clarified

When starting with an analysis of IT costs, one will need to begin by considering what is meant by 'cost'. Since 'cost' is a very general term, it is important to clearly define what aspects of cost need to be recorded, which type of costs are needed, and how they should be kept. The type of costs incurred in IT vary widely, and cost can mean very different things depending on the context.

### 2.1.1 Accountancy and costs

The accountancy of a company is of prime importance when doing any kind of cost analysis. This is why the subject of accountancy will be broached, albeit with broad strokes only. While the way accounting is done varies between countries and organizations, there are some basic similarities that apply to most businesses of today, keeping in mind that these are not universal truths.

Two terms from accounting will often come up in relation to IT costs. They are capex and opex. Costs (or expenses) incurred by a company will generally either be handled as capital expenditures(Weygandt et al., 2005) (capex) or operational expenditures(Weygandt et al., 2005)(opex). Capex applies to expenditures on fixed assets or upgrades on fixed assets, given that the asset last longer than a year. Capex expenditures are capitalized and then depreciated over the lifetime of the asset. Opex applies to day to day expenditures for running businesses, such as wages, software licenses, utilities etc.

From an accounting standpoint, IT costs will be a mixture of capex and opex. There will also be grey areas in between, since there will be cases where a capex is incorrectly classified as on opex. This could be for example if a cheap server is bought, it would often be taken as a opex.

With this in mind, it is apparent that information on whether equipment is classified as opex or capex is needed for the correct treatment of costs. It should also be noted that whether IT costs are classified as capex or opex varies greatly between companies and will not always make sense from an IT standpoint. IT equipment will often be given a depreciation period that is highly unrealistic, or in some cases all IT expenditures are booked as opex for taxation purposes.

### 2.1.2 Time span of costs

When doing a cost analysis, care must be taken to note the time span of the costs being measured. The various time spans used in business are too many to list, but there are a few time spans that are commonly used. First there is the 'complete' time span, that is the time span that includes the costs from the moment they were incurred to the present. Annual costs are also very useful, and usually refer to the fiscal year as a time period. Finally there is the fiscal quarter, which an important time period for most businesses.

### 2.1.3 Hard and soft costs

Costs in IT can be categorized into 'hard' costs and 'soft' costs(Bannister, 2004). Hard costs will refer to costs that pertain to the purchase, maintenance, upgrade, licensing, leasing, etc of IT. There will usually be a receipt for hard costs.

Soft costs are abstract costs that are often hard to measure. They are the cost of unplanned downtime, having a overly complex infrastructure, using obsolete hardware or software, etc. Soft costs will usually be based on hard costs - so it is important to get the hard costs first. This thesis will only be concerned with hard costs.

Labor can be a hard or soft cost, depending on the context. Doing technical work on an IT system will be a hard cost, but learning and making decisions on the future direction of a company's IT is a soft cost.

### 2.1.4 Other issues with cost

**Currency**

If a business must deal with different currencies in their evaluation of IT costs then that will add a level of complexity on all of their calculations and estimations. Companies whose operational currency is not a reserve currency (i.e. USD, EUR, etc) usually have to buy software licenses denominated either directly of indirectly in another currency. Currency fluctuations will then constantly have to be measured to know the current and past costs.

**Interest rates**

It is advisable to store the interest rates on credit that are available at any given time to a company in order to use in economic evaluations of IT. The interest rate information can then be used in retrospective financial evaluation in order to assess current IT assets, and can also be used in forecasting.

## 2.2 Some uses of cost information

To date a multitude of economically driven evaluations of IT have been developed. Although this thesis will not delve into them, it is worth noting the various methods since they can be very informative when employed correctly. The methods most often mentioned can be seen in table 2.1. What all the methods have in common is that they use IT costs as a starting point. A comparison between the methods can be found in (Mutschler et al., 2007).

*Table 2.1:* Economically driven evaluations of IT

| Name | Abbreviation |
|---|---|
| Total Cost of Ownership | TCO |
| Return on Investment | ROI |
| Economic Value Added | EVA |
| Real Options Valuation | ROV |
| Return on Assets | ROA |
| Return on Infrastructure Employed | ROIE |

Of the methods mentioned in table 2.1, probably the most famous (and most maligned) method to evaluate IT is TCO(Hawkins, 2001). In particular various TCO studies were conducted and presented around during the years 2000 to 2005. There were also numerous academic papers written on TCO in this period. TCO is much less touted today, as other methods have supplanted it.

Having good information on costs, either estimates or concrete data, is a motivating factor for the pursuit of economic evaluation of IT. Such evaluations will assist companies in making informed decisions.

## 2.3 Acquiring costs

### 2.3.1 The recording of IT cost information

The tool that is often used by companies to handle costs are ERP systems. Modern ERP systems have become quite powerful and it would seem obvious to delegate the task of handling IT costs to them, but they are not well suited to handle IT costs. The main reason that ERP systems are not well suited to handle IT costs is that ERP systems are in general designed to handle fixed assets, while IT is constantly changing (ITA, 2006).

The methods of categorizing and estimating costs put forth in this thesis assume that an effort is being made to both record and to maintain cost information. Cost information should be collected in a prescribed way so that the various costs are compatible with each other. Since the costs of IT will most involve most if not all parts of a company, the gathering of IT costs will involve many different parties.

Aforementioned was the assumption that IT costs are collected and maintained in a prescribed manner. The reasons for this is mainly that if there is no defined way of collecting costs, they can become very inconsistent, so much so that comparisons between costs will not make sense. Also the use of historical costs is sometimes the only way a company has to assess their current costs, since industry averages are often not available or relevant.

The major drawback of maintaining IT costs, and probably the reason why many companies do not do it, is that it can be quite expensive. That is why it is useful to use cost estimations and averages to get started with cost analysis in IT.

There is a set of business practices called IT asset management (ITAM)[1] that has among its goals to provide the means to enable the management of IT assets. Various software is available to help with ITAM. Having a thorough ITAM setup in place is a very useful way of maintaining IT costs, but many companies will be deterred by the large amount of time and money such a setup would demand.

To have a simple scalable model for assessing costs can be highly beneficial for a company.

[1] for more information see http://www.iaitam.org

There are many reasons for this. Since the costs of IT tend be spread out in a company, for example the accounting department might handle the real estate costs while the IT department handles server deployment, getting all the data needed to discover costs in IT can be tedious. Having a mechanism in place that aggregates the needed data from across the company is difficult to accomplish but is very valuable, and in fact imperative if one is to get the whole picture.

### 2.3.2 Labor cost

The cost of labor in IT deserves special consideration since there are so many different ways of measuring it. Ideally a company will have detailed records of the amount of work done on any given aspect of their IT setup. Such records would include who did the work, how long it took, and the hourly cost.

The assumption in this thesis is that work done on a company's IT is recorded onto a time sheet with a level of detail that is thorough but only to a degree that is practical. A company might not require its employees to record any of their work specifically - that would pose problems to getting any reliable costs.

There will of course be other considerations in respect to labor costs, like for example the cost of educating employees. If such costs are incurred and measured, it would make sense to attach those costs to the systems that require them.

There are many reasons why it is important to have a good account of where time is being spent on the various parts of IT, but it is particularly valuable when considering the purchase of software that is meant to be labor saving in IT. Such software can be for example automatic provisioning software, Identity management or any other similar software.

# 3 IT cost breakdown structure

The categorization of IT costs is a useful way to structure and present them in a simple and straight forward manner. Cost categories will also facilitate the making of cost comparisons, and provide a platform for discovering cost trends. This chapter will propose grouping of costs in order to to define easily understandable units which can then be used to compare to similar units inside or outside an organization.

## 3.1 Categorizing IT

The core of an organizations IT is usually its data center. The rest of an organizations IT will be somehow connected to the data center, relying on it for making available the appropriate systems , storage of important data and many other things. The data center itself might be outsourced or on premise, but will in most cases incur considerable costs. The categorization of IT costs presented here will use the data center as a starting point, and expand outwards - with the goal of creating appropriate categories for as much of an organizations IT as possible.

The costs that will be identified in this chapter are by no means exhaustive, but are meant to indicate how they can be grouped. In addition the costs mentioned will give an idea of what to look for, as well as the scope of the process, since cost gathering for IT will involve getting information from most departments of a company. Costs will often fall into grey areas in respect to how they should be classified but the hope is to minimize the occurrence of such costs.

The cost categories are built up around the costs of having the IT infrastructure 'on premise' (i.e. not outsourced or something similar). These costs are often quite unclear and deserve special attention. Also as will be apparent by looking at the groupings, a fundamental assumption on the organization of on-premise IT will be made, and this assumption will hold for the IT infrastructure of most of today's companies. In the case where some or all of the IT is outsourced, the cost structure will be simpler, and many of the categories will not be needed.

Having a predefined grouping of IT costs in place which encompasses the IT of an organization will facilitate cost estimation. The cost information available does not have to be detailed to make estimations of the total costs of each category. More information can then be gradually obtained to refine the estimations, and the process of estimation will be more easily defined. The categories are thus a prerequisite for the cost estimation techniques that are presented in a later chapter.

After the cost categories have been defined and agreed upon in an organization, they can then be used as a terminology when analyzing costs. Since the categories should be well understood, cost information will be easier to record and understand.

In project management, projects are broken down into comprehensive units with a method called 'work breakdown structure'(Maylor, 2003). This methodology will be used here, but will be adapted to costs (also used in (Garvey, 2000)). But because it is not work that is broken down, but IT costs, we will call it 'IT cost breakdown structure' instead.

## 3.2 IT cost considerations

In reference to the considerations mentioned in chapter 2.1, the costs that need to be collected are ideally annual costs, if they are not one off costs, and of a fixed currency. The costs should also be 'hard costs', as much as possible. This is easy when dealing with physical costs such as servers or housing, but more difficult when assessing labor costs.

Labor costs are often a considerable portion of IT costs, but despite this labor costs are not categorized specifically here. The reason is that since IT companies will often use a mix of in house labor and contractors, each category will include as a cost component the labor done in that specific category. In this manner, the 'hard costs' of labor are associated with the relevant category.

## 3.3 Divisions

The different parts of the IT infrastructure can be subdivided into the following parts,

- Housing and utilities
- Base infrastructure
- Support infrastructure
- Computer hardware
- Operating systems and supporting software
- Software
- Peripheral infrastructure

Note that the assumption is made that the costs that are being recorded for each category are the cost of physical assets, and additionally the cost of labor associated with them. Also the costs of maintenance and warranty contracts are assumed to be divided among the categories if needed.

The rationale behind this specific division is an attempt to divide the different parts of the infrastructure according to the personnel that service them. Dividing IT along the lines of who is doing the work, and by extension what vendor is involved, simplifies the grouping of IT costs. Even though there will be situations that the same vendor or personnel work on different parts of the IT infrastructure, they will often only work on one of the levels at any time.

When creating a division of IT costs into groups a primal concern will be to have as little overlap as possible. This will not always be feasible so it is important to denote clearly that a specific cost will be associated with more than one element of the IT of a company.

### 3.3.1 Housing and utilities

This is the foundation for all the in-house IT infrastructure of a company. Companies will usually have one or more data centers and this category of costs encapsulates the housing costs and the costs of utilities needed to run IT. This category is a useful metric to use when comparing costs between different sites.

- Housing
- Electricity
- Water
- Physical security
- Fire suppressant
- WAN connectivity
- Backup generator

The housing element will include also the basic interior such as lighting and flooring and the like. Water is mentioned here because many data centers use water for cooling their equipment. WAN connectivity comprises of the physical link of the data center to an Internet provider and any other connections that might be needed. The backup generator will often be a diesel machine that is outside the data center.

### 3.3.2 Base infrastructure

This level uses the resources supplied from the previously mentioned housing and utilities level to create an environment for IT equipment. The items mentioned here will likely comprise a large part of the total costs of the construction costs of a data center.

- Cooling
- UPS
- Server racks
- Wiring (Electrical/Networking/Fiber optic/other)

The components mentioned constitute the basic building blocks for connecting the basic work units of IT systems. It is often the same person who sets up the racks and wires them, but racks are not a prerequisite for data centers, many organizations choose to use simple shelves to store their IT hardware.

### 3.3.3 Support architecture

- Storage (SAN, NAS, etc)
- Network (Switches, routers, etc)
- Backup (Tape drives, etc)
- Blade chassis

These technologies are among the major components of modern data centers.

The context in which they are here refers to when they are systems themselves.

#### 3.3.3.1 Storage

Storage systems, such as SAN, NAS, etc, are constantly growing in capacity(Sta, 2008). and have a wide array of variables that will affect the cost of storage for a servers. This is why having detailed information of how storage is used will often make sense. There is also the growing realization in the IT sector that storage is in general highly underutilized - this may for example be seen in Symantec's storage survey (Sta, 2008).

In cases where a single server uses its own local storage, this storage should be accounted for in the cost of the server itself. And likewise when a whole storage subsystem is dedicated to just one server, than it would make sense to account for that storage with the cost of that server.

Storage virtualization systems would be included in these costs. This includes specialized servers that are used as controllers.

#### 3.3.3.2 Network

Networking infrastructure is also often difficult to quantify, because it is often not contained to just the data center, but is distributed through whole office buildings. Many modern buildings have mini-data centers on every floor to house networking equipment (often no larger than broom closets) but these locations should be included in any cost calculations for the networking infrastructure.

#### 3.3.3.3 Backup

If a company is using tape or similar technologies as a backup medium than those costs would be accounted for here. When backup is done to the Storage subsystem, the amount of space allocated to backup could be noted down and assigned to these costs.

#### 3.3.3.4 Blade chassis

One important addition to this list is the inclusion of blade chassis. The reason for them being here is that they have more to do with the the infrastructure surrounding the servers then the servers themselves. They may include important components of the network and SAN infrastructures. Although it does make sense to attribute some of the costs of the blade chassis to an individual blade server, the effort would probably not be cost effective. But as with the San and network infrastructure, the blade chassis infrastructure should be accounted for.

### 3.3.4 Computer hardware

- Rack mounted servers (1U, 2U, etc)
- Blade servers
- Other

Thorough documentation on the servers (or computers) in a data center is very beneficial to IT divisions in companies. Definitions need to be made of how servers are defined, both in size and in function. Servers will have different attributes and costs depending on their capabilities.

Cost savings associated with using virtualization to run virtual machines are much more easily quantified when good data on server costs is available. Virtual machines will use the resources of a server differently, so if a high degree of detail is needed on costs, the resources of a computer (RAM, CPU, etc) should be known both quantitatively and by the cost.

### 3.3.5 Operating systems and supporting software

- Virtualization Software
- Operating systems
- Supporting software

Often the licensing costs of operating systems can be quite significant and must be taken into account when calculating the costs associated with servers. This also holds true for virtualization software.

Supporting software is software that is always installed with the operating system, and commands a license fee. This is for example anti virus software. Often there will be no supporting software.

### 3.3.6 Software

- License fees
- Installation and maintenance work
- Upgrades
- Software development

Discovering and maintaining the costs of software can be simple as in the case of plain vanilla offerings of software vendors that can function with minimal configuration. But software is often made in-house or software from third parties can be greatly changed to adapt it to a company. Finding the costs of software can quickly become an overwhelming task, so creating basic guidelines about how software costs are defined and applying them as broadly as possible is essential.

Software costs in IT can be quite substantial in many cases and deserve special attention. Specifically the costs that relate to software development, maintenance and customization can be quite high. There is ample literature available that deals with the economic side of software so it will not be discussed further here, except to stress that a thorough understanding of current software costs is vital in order to make informed decisions in software license or labor purchases.

### 3.3.7 Peripheral infrastructure

- Monitoring
- Management software and hardware

This cost level entails costs that are commonly used by most of the IT infrastructure.

#### 3.3.7.1 Monitoring

The monitoring of hardware and software is fundamental to running a data center. The costs of monitoring depend on a large part both on the amount of personnel tasked with the monitoring, and the scope of monitoring.

It is very useful to have at least an overview for each server in a data center of the CPU and memory statistics. These are important for the discovery of trends, and also to find out if hardware is being underutilized.

#### 3.3.7.2 Management software and hardware

This should be interpreted as software used in the management of the data center. A good example of this would be backup software. Most data centers rely on a single backup solution, and software costs behind that solution would lie here.

Software that coordinates workload, or assists in provisioning servers would be accounted for here as well as SOA infrastructure, management buses, LDAP and the like. Often there is a need for specialized hardware for the operation of IT equipment, these costs will also be included here.

## 3.4 Data center capacity

An aspect of data centers that needs to be known to facilitate the recording of costs is the amount of physical space a server takes. In most cases this is a fairly straight forward process but there are variables that must be taken into account.

The space in a data center that is used for creating an infrastructure with electricity, cooling and backup power generation

The space in a data center that is available for IT hardware must be known. This is the part of the data center that at the beginning is usually just empty floorspace, and often this floor space is 'raised'. The space that is reserved for cooling, UPS, backup power and other infrastructure is excluded. How many square meters of usable floor space is in a data center it an important metric.

### 3.4.1 Capacity measurement

Most modern data centers use standard server racks to house IT equipment. Server racks are divided into units called 'U'. The standard server rack has a width of 482.6mm (19 in) and is divided into 42 'U' (Commission, 2008).

### 3.4.2 Total capacity

The total capacity of a data center is essentially how many 'U's are available to use. Over the lifetime of a data center, this amount will most likely change, whether because more server racks are put into the data center, or the racks themselves are replaced with other racks of different capacity.

Maintaining the total capacity at any given time of a data center is one of the basic metrics needed for finding costs.

### 3.4.3 Installed capacity

Knowing how much of the data center is used is useful for many reasons. For our purposes it is good to know so that the cost of unused capacity can be calculated.

### 3.4.4 Server size

IT equipment that is used in data centers nearly all 'rack' mountable. An average server may use from 1U to 4U. But this only applies to a certain class of servers, many servers are not as easily quantified. For example blade servers might just use a fraction of a U, while a mainframe might take more floorspace than a whole server rack.

#### 3.4.4.1 Blade servers

To calculate the number of U for a standard blade one would find out how many blades the blade center can have, and also how many U the blade center itself takes. So a blade center that takes 14 blades, but only uses 7U, would have blades that use 0.5U.

#### 3.4.4.2 Mainframes

Mainframes and other hardware of comparable size are normally at least as large as a full rack. If these can be found or similar hardware, than they could be approximated as taking the full 42U of a rack or more.

### 3.4.5 Data center comparisons

With capacity and cost information, making comparisons between different data centers is made easier. When comparing the costs of two data centers, thought should be put to the metric's used to compare them. Though it might be tempting to use a cost/$m^2$ metric, this would be highly misleading, since that does not take into account the computational power of each $m^2$. Studies have shown that using a cost/server metric gives better results (Karidis et al., 2009).

## 3.5 Considerations

Any categorization of something as complex as IT will give rise to various considerations and problems. Some of the considerations related to the IT breakdown structure proposed in this chapter will be addressed.

### 3.5.1 Aggregating costs

After setting up the cost categories the next step is to assign the costs themselves to them. One way of doing this is to simply put cost estimations on each category, using the cost estimation methods proposed in this thesis. Another way is to take available cost information and use that to fill up the categories. Obtaining the true costs for each category is a time consuming and labor intensive task and usually not practical, so in the end the costs will usually be a mix of real costs, estimations and averages.

The end result should then be categories with hard costs attached to them. These in turn can be used as a basis for a whole host of cost analyses, depending on how much detail is available of the cost information.

### 3.5.2 Extensions

The categorizations that have been defined in this chapter are a starting point for the grouping of IT costs for a typical organization. The IT components that were omitted from these categories can easily be added, such as an organizations telephone system or laptops, but the point is not to have a definitive list of all IT components, rather to give an example.

Extending the categories to encompass soft costs such as learning, project management, IT management etc. would also be desirable. Labor costs that can not be categorized as 'hard cost' would then be added to a new 'soft cost' category. Analysis of this 'soft cost' category would be beneficial but is out of the scope of this thesis.

## 3.6 Application to IT services

An example of how to use the cost information is to find the costs of IT services. IT services are an abstraction of the work done in IT, where tasks are organized into 'services'. The branch of IT that deals with managing IT services is called 'IT service management' (ITSM) (O'Neill and Krauss, 2006). While the IT breakdown structure does not directly provide the costs of IT services, if the costs are detailed enough, they can be used to discover the costs of IT services.

# 4 Overview of methods used

The mathematical methods used in the next chapter will be briefly described in this section. These are mostly statistical methods, and the probability densities that are used will be explained and graphs of them shown.

## 4.1 Distribution functions

The following are the probability distributions used in the next chapter.

Note that for each probability density $f(x)$ the following holds,

$$f(x) \geq 0, \quad \int_0^{\infty} f(x)dx = 1 \tag{4.1}$$

### 4.1.1 Triangular distribution

A triangular distribution $\mathbb{A}$ is analytically simple and often used in business and finance (Bojadziev and Bojadziev, 2007). Triangular distributions are defined on a supporting interval

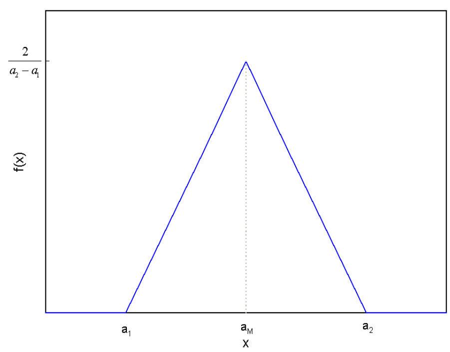


*Figure 4.1:* Triangular probability density function

$[a_1, a_2]$, and the point $a_M$ is the peak. So if one has the smallest and largest values ($a_1$ and $a_2$), along with the most likely value ($a_M$), a triangular distribution can be denoted by,

$$\mathbb{A} = (a_1, a_M, a_2) \tag{4.2}$$

The value that the triangular distribution takes is at $a_M$ is $2/(a_2 - a_1)$. It is also worth noting that the triangular distribution will often be a right triangle, that is we will have $a_1 = a_M$ or $a_2 = a_M$.

The formula for the triangular probability density function is given by,

$$f(x|a_1, a_M, a_2) = \begin{cases} \frac{2(x-a_1)}{(a_2-a_1)(a_M-a_1)} & \text{for } a_1 \leq x \leq a_M \\ \frac{2(a_2-x)}{(a_2-a_1)(a_2-a_M)} & \text{for } a_M \leq x \leq a_2 \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

The mean($\mu$) and variance ($\sigma^2$) of the triangular distribution are,

$$\mu = \frac{a_1 + a_M + a_3}{3} \tag{4.4}$$

$$\sigma^2 = \frac{a_1^2 + a_2^2 + a_M^2 - a_1 a_M - a_1 a_2 - a_M a_2}{18} \tag{4.5}$$

### 4.1.2 Trapezoidal distribution

A trapezoidal distribution **A** is defined on the interval $[a_1, a_2]$. It has a flat segment on the interval $[b_1, b_2]$. Trapezoidal distributions can be denoted by,

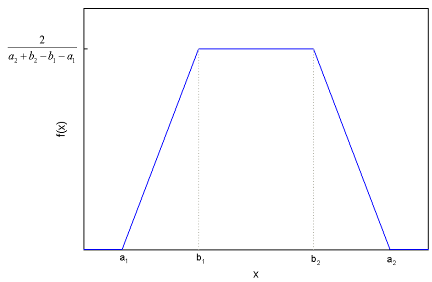


*Figure 4.2:* Trapezoidal probability density function

$$\mathbb{A} = (a_1, b_1, b_2, a_2) \tag{4.6}$$

The trapezoidal probability density function takes the value $2/(a_2 - a_1 + b_2 - b_1)$ on the interval $[b_1, b_2]$, which is the flat segment.

The formula for the trapezoidal probability density function is,

$$f(x|a_1,b_1,b_2,a_2) = \begin{cases} \frac{2}{b_2+a_2-a_1-b_1}\frac{1}{b_1-a_1}(x-a_1) & \text{for } a_1 \leq x < b_1 \\ \\ \frac{2}{b_2+a_2-a_1-b_1} & \text{for } b_1 \leq x < b_2 \\ \\ \frac{2}{b_2+a_2-a_1-b_1}\frac{1}{a_2-b_2}(a_2-x) & \text{for } b_2 \leq x \leq a_2 \\ \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

The mean($\mu$) and variance ($\sigma^2$) of the trapezoidal distribution are,

$$\mu = \frac{((b_2+a_2)^2 - b_2a_2) - ((a_1+b_1)^2 - a_1b_1)}{3(b_2+a_2-a_1-b_1)} \tag{4.8}$$

$$\sigma^2 = \frac{(b_2^2+a_2^2)(b_2+a_2) - (a_1^2+b_1^2)(a_1+b_1)}{6(b_2+a_2-a_1-b_1)} - \mu^2 \tag{4.9}$$

### 4.1.3 Uniform distribution

The well known uniform distribution can be described as a special case of the trapezoidal distribution where $a_1 = b_1$ and $a_2 = b_2$.

Uniform distributions can be denoted by, Trapezoidal distributions can be denoted by,

$$\mathbb{A} = (a_1, a_2) \tag{4.10}$$

The formula for the uniform probability density function is,

$$f(x|a_1,a_2) = \begin{cases} \frac{1}{a_2-a_1} & \text{for } a_1 \leq x \leq a_2 \\ \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

The mean($\mu$) and variance ($\sigma^2$) of the uniform distribution are,

$$\mu = \frac{1}{2}(a_1+a_2) \tag{4.12}$$

$$\sigma^2 = \frac{1}{12}(a_2-a_1)^2 \tag{4.13}$$

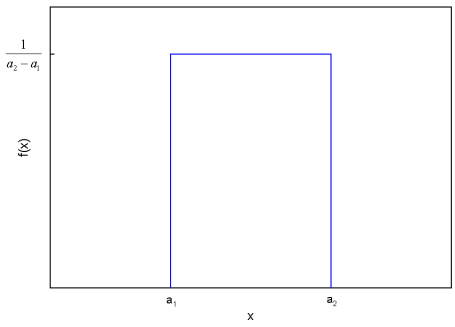


*Figure 4.3:* Uniform probability density function

### 4.1.4 Beta distribution

The beta distribution is defined on the interval (0,1) and is parameterized by two positive shape parameters called $\alpha$ and $\beta$. The beta distribution is often used in place of the triangular distribution, since it can have a similar shape and is defined on a finite interval.

The formula for the beta probability density function is,

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} \tag{4.14}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1} \tag{4.15}$$

where $\Gamma$ is the gamma function.

The mean($\mu$) and variance ($\sigma^2$) of the beta distribution are,

$$\mu = \frac{\alpha}{\alpha+\beta} \tag{4.16}$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{4.17}$$

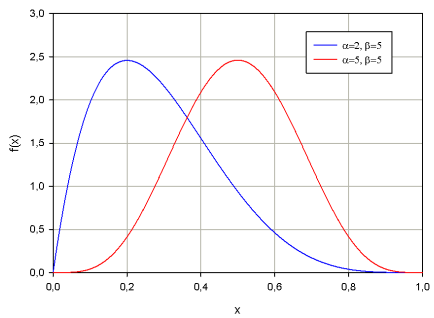


*Figure 4.4:* Beta probability density function, two examples

### 4.1.5 Gaussian distribution

The Gaussian (normal) distribution is defined by its mean $\mu$ and variance $\sigma^2$. It is an important distribution that is often encountered in practice. Gaussian distributions are denoted as,

$$\mathcal{N}(\mu, \sigma^2) \tag{4.18}$$

The formula for the Gaussian probability density function is,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4.19}$$

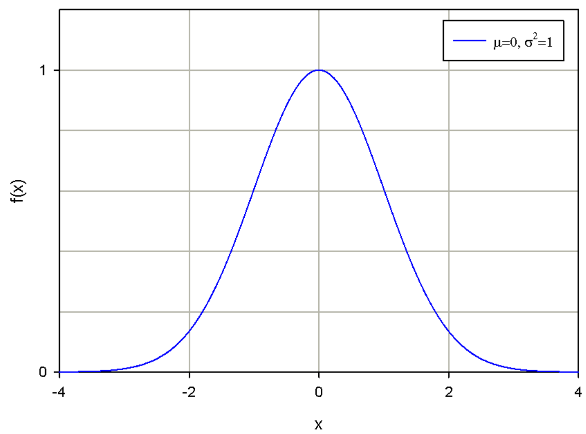


*Figure 4.5:* Gaussian probability density

## 4.2 The Central Limit Theorem

One of probability theory's most useful theorems is the central limit theorem. The central limit theorem states that when there are a sufficient amount of random variables, and they meet certain conditions, they will be approximately normally distributed(Montgomery and Runger, 2007).

The conditions that need to be met are that the random variables should be mutually independent, each with a constant mean and variance. The random variables should also contribute equally, that is, none should dominate in standard deviation.(Garvey, 2000)

## 4.3 Sums of random variables

The probability density function of the sum of two random variables, where both variables have separate probability densities, is the convolution of their density functions.

### 4.3.1 Convolution

The following is the continuous convolution equation,

$$f_{A+B}(x) = \int_{-\infty}^{\infty} f_A(x-y) f_B(y) dy \quad \text{for} \quad -\infty < x < \infty \tag{4.20}$$

and the discrete time version of convolution is given by,

$$f_{A+B}[n] = \sum_{m=-\infty}^{\infty} f_A[n-m] f_B[m] \tag{4.21}$$

$A$ and $B$ represent two independent random variables, each with a probability density function.

### 4.3.2 Frequency domain transformation

The convolution of PDFs (probability density functions) can be a computationally intensive operation, depending on the PDFs involved. For this reason it is desirable to find a faster way of performing convolutions. One method that is used extensively in signal processing is frequency domain transformation(Oppenheim et al., 1996).

Frequency domain transformation involves taking a function in the time domain, such as a PDF, and transforming it into the frequency domain. In the frequency domain, convolution is a simple multiplicative operation, which is computationally efficient. The result is then transformed back into the time domain with an inverse frequency domain transformation.

There are various transforms available that transform a time domain function into its frequency domain representation. The transform that will be used in this thesis is the discrete Fourier transform.

### 4.3.3 Fast Fourier Transform

The FFT (fast Fourier Transform) is an efficient algorithm to calculate the discrete Fourier transform, and its inverse. Note that in the case of discrete time PDFs, the choice between using convolution and FFT is usually dependent on how many data points are in the PDF. In practice, if the data points are more than 100, it is generally faster to use the FFT.

# 5 Cost model

Implementing a cost model that adequately describes an organizations IT costs will provide the organization with a simplified view of its often complex IT costs. The cost model will, among other things, assist in making informed decisions on IT investment, help locate cost drivers and identify sunk costs. The end result will be more cost effective IT.

The cost model proposed here will be a framework in which costs will be categorized with the use of the cost categories described in chapter 3. In addition, the methods shown in chapter 4 will be used to to estimate costs that are unknown or partly known. The cost model is designed to be extensible and to incorporate the uncertainty of the available cost information.

Each cost category defined will have a cost associated with it. The cost amount on each category can be a pure estimate, an aggregation of the actual costs, or a mix of the two. In the case where estimations are involved, probability density functions will be used.

Gathering together the costs associated with a company's IT infrastructure is complex and time consuming. An assumption that all IT costs are available at the outset is not made, and in fact the cost model can be applied with no cost information at all, as long as plausible estimates can be made. The accuracy and usability of the cost model depends on how much effort is put into collecting and estimating costs. A starting point to find costs for data centers, if absolutely no information is available is by using industry standard costs (Koomey, 2007).

## 5.1 IT cost levels

In chapter 3 various IT costs have been defined and categorized. The results will be used as a basis for the cost model.

The following is a summary from chapter 3 on the cost categories defined in the IT cost breakdown structure. Each category will be called a 'level' and assigned a variable.

Having absolute values for each of the costs listed above, the total cost of a company's IT

*Table 5.1:* Cost levels

| Cost level | Variable | Details |
|---|---|---|
| Level 1 | $L_1$ | Housing and utilities |
| Level 2 | $L_2$ | Base infrastructure |
| Level 3 | $L_3$ | Support infrastructure |
| Level 4 | $L_4$ | Computer hardware |
| Level 5 | $L_5$ | Operating systems and supporting software |
| Level 6 | $L_6$ | Software |
| Peripheral | $L_P$ | Peripheral infrastructure |

can be calculated to be

$$\text{Cost}_{IT} = L_1 + L_2 + \cdots + L_6 + L_P \tag{5.1}$$

Going deeper into the cost levels, each can be assigned subcategories or sub levels. An example of sub levels for cost level $L_1$ is given in table 5.2.

*Table 5.2:* Example of cost sub levels

| Level 1 | $L_1$ | Basic infrastructure |
|---|---|---|
| Sub level 1 | $L_{11}$ | Housing |
| Sub level 2 | $L_{12}$ | Cooling |
| Sub level 3 | $L_{13}$ | Electricity |
| Sub level 4 | $L_{14}$ | Security |

The cost level $L_1$ can then be denoted as,

$$L_1 = L_{11} + L_{12} + L_{13} + L_{14} \tag{5.2}$$

If we had information on a particular server's electricity usage, let's call the server 'server X' we would call that variable $L_{13serverX}$.

### 5.1.1 Example:Cost of a server

In chapter 3.4.1 the various capacities are defined and we will use the following variables for capacities,

| | |
|---|---|
| $Cap_{total}$ | total capacity |
| $Cap_{installed}$ | installed capacity |

To find the cost of a particular server, it's size measured in 'U' is needed. The first three levels of data center costs can then be aggregated and divided by the capacity,

$$Cost_{server}(size) = size \cdot \frac{L_1 + L_2 + L_3 + L_P}{Cap_{total}} + L_{5server} + L_{6server} \tag{5.3}$$

## 5.2 Assigning costs to cost levels

Each cost level needs to be assigned costs. The costs that can be assigned depend on the information that is available, but there three main cases:

### No cost information available

While it is highly unlikely that an organization would have no cost information available to it, the information might be unavailable. People with an overview of IT costs can usually be found and asked to give estimations.

### Some cost information available

This is probably the most common case. The cost information is categorized and any missing costs are estimated. Sometimes cost averages are used when deemed necessary. Cost averages are obtained by getting the average costs of a standard unit of IT, and using that cost on other similar IT units. The end result will be that each cost category will be a mix of estimations and real costs.

### All cost information available

When all cost information is available, it can be categorized and no estimates will be needed. This will usually involve a great amount of effort, but the result will be very good cost information.

## 5.3 Cost estimation

Using approximations for costs is usually the fastest way to get any cost model up and running. But the problem is that often the uncertainty rises very quickly, and if many people are giving estimations for various costs, the uncertainty of an estimate will be that much harder to grasp.

This is where a framework that takes into account cost uncertainty can be useful. If the associated probability distribution can be given with each cost, not only will the cost metric be available to an organization, but the uncertainty metric also. The long term goal will then be to keep costs at a minimum, and to keep the uncertainty of the costs also at a minimum.

The probability distributions that will be proposed will be kept as simple as possible. The cost equation (5.2) can be used as an example, with the adjustment that each of the variables is an independent random variable with a specified probability distribution.

Since the variables are independent, we have

$$E(\text{Cost}_{IT}) = E(L_1) + E(L_2) + \cdots + E(L_6) + E(L_P) \tag{5.4}$$

$$Var(\text{Cost}_{IT}) = Var(L_1) + Var(L_2) + \cdots + Var(L_6) + Var(L_P) \tag{5.5}$$

The cost levels are thus represented by probability densities instead of plain cost figures.

### 5.3.1 Probability densities for costs

It is preferable to have easily understood probability densities in order to make cost estimations. Four probability densities are suggested. An overview of the mathematical properties as well as graphs of the probability densities can be found in chapter 4.1.

**Triangular**

The triangular probability density is useful when the most likely cost is known, but there is some uncertainty that should be accounted for. The triangular distribution is denoted by $\mathbb{A} = (a_1, a_M, a_2)$, where $a_M$ is the likely cost, while $a_1$ is the lower bound and $a_2$ is the upper bound.

**Trapezoidal**

The trapezoidal probability density is used when the most likely cost less well known than in the case of the triangular distribution. The trapezoidal probability density is denoted by $\mathbb{A} = (a_1, b_1, b_2, a_2)$. The likely cost is then a range of costs that are equally likely (the flat segment, $b_1$ to $b_2$), as well as upper and lower bounds that fall steeply and are considerably less likely ($a_1$ and $a_2$).

**Uniform**

The uniform probability density is used when the costs are especially unclear. It is denoted by $\mathbb{A} = (a_1, a_2)$. It is most useful when estimating a cost that is equally likely to be a range of different costs, that is the cost could be any cost on the range from $a_1$ to $a_2$.

**Beta**

The beta distribution is flexible and is useful when a analytical solution is needed. It can be parametrized to have a very similar shape to the triangular distribution.

### 5.3.2 Cost distribution

Once the cost estimations have been made and the distributions are available, the estimated cost and its variance can be determined. According to the central limit theorem mentioned in chapter 4.2, when there are sufficiently many random variables, the normal distribution can be assumed when dealing with the resulting probability density.

Whether or not the normal distribution is plausible can be determined in various ways. Monte Carlo simulation can be used, along with visual inspection of the resulting distribution. A more rigorous explanation can be given by the Kolmogorov-Smirnov test (DeGroot, 1986), which will reveal whether the normal distribution will suit for a given random distribution.

When dealing with relatively few costs, the assumption that the addition of the probability densities will result in a normal distribution will usually not hold. In those cases it would be possible to use continuous time convolution to discover the resulting probability density. Direct calculation of the probability density will then be needed.

The probability density function of a sum of independent variables with probability densities $A$ and $B$ is given by the convolution of their density functions.

While convolution can be used for finding the probability density, it is computationally intensive. A more efficient way of achieving the convolution of two sequences is by use of the fast Fourier transform (FFT)(Oppenheim et al., 1996)(Sakamoto et al., 1997).

#### 5.3.2.1 Assumption of independent costs

In chapter 5.3.2 it is noted that the costs are assumed to be independent from each other. This will not always be the case, costs can and will be in many cases correlated with each other. Correlation will have an effect on the variance of the resulting probability density. When estimating costs it is wise to keep in mind an correlations within them.

# 6 Application example

Using the methods introduced in this thesis, cost for a fictional IT department is estimated. The accounting practices are purely fictional, as are all the prices. All prices are measured in ISK. The costs will not be differentiated by whether they are capital or operational expenses, but that could easily be incorporated. All costs are quoted on a yearly basis.

## 6.1 Costs and their distributions

The IT department that needs to make cost estimations does not have time to look up all the relevant costs, so it assigns the task of estimating the costs to selected employees that have a good understanding of the IT costs.

Since this is the first time they are analyzing their costs in this manner, only part of their IT will be analyzed. The IT costs that will be analyzed correspond to levels 1-3 and level P, they are explained in chapter 3, but are repeated here.

| Level | Variable | Description |
|---|---|---|
| Level 1 | $L_1$ | Housing and utilities |
| Level 2 | $L_2$ | Base infrastructure |
| Level 3 | $L_3$ | Support infrastructure |
| Peripheral | $L_P$ | Peripheral infrastructure |

The costs were then estimated by the employees and the estimates were explained.

$L_1$ These costs are known reasonably well so they are estimated using a triangular distribution

$L_2$ These costs are also known well so they are estimated using a triangular distribution

$L_3$ These costs are not well known, but there is some knowledge of them so they are estimated using a trapezoidal distribution.

$L_P$ These costs have rarely been considered so they are given a wide uniform distribution.

The following cost figures were given to the probability densities, using the notation introduced in chapter 4.1.

| Name | variable | Distribution | Cost function |
|---|---|---|---|
| Level 1 | $L_1$ | Triangular | (3.000.000, 4.000.000, 6.000.000) |
| Level 2 | $L_2$ | Triangular | (250.000, 300.000, 350.000) |
| Level 3 | $L_3$ | Trapezoidal | (2.000.000, 2.250.000, 2.750.000, 3.000.000) |
| Peripheral | $L_P$ | Uniform | (1.000.000, 2.000.000) |

The plots for the cost distributions are the following,

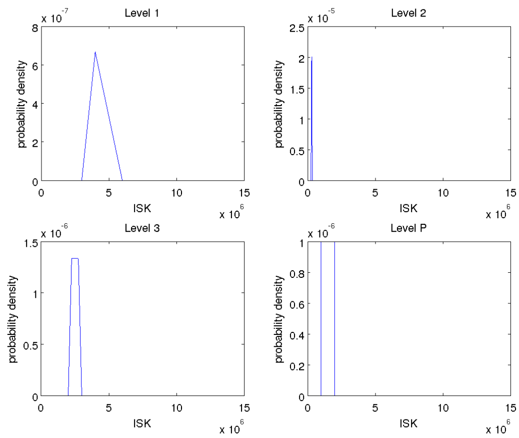


*Figure 6.1:* Plots of cost distribution functions

## 6.2 Data center costs

With the information we have we can find the yearly cost. To accomplish this we use the fast Fourier transform(FFT) on the probability densities. The probability densities are thus transformed into the frequency domain, where they are multiplied together. Next an inverse FFT is performed on the result to obtain the combined distribution.

A plot of the result of the combined distribution in figure 6.2 shows that although we have only four cost distributions, the combined distribution is very close to being normal. The results are summarized in table 6.1

*Table 6.1:* Summary of results

| | |
|---|---|
| mean | 8.472.000 |
| $\sigma^2$ | $5.5 \cdot 10^{11}$ |
| 95% confidence interval | (-12.9%)7.380.000 to 10.175.000(+20%) |

Note that the confidence interval is not symmetric, it has a negative skew. This is because the underlying distributions are not all symmetric. With more cost estimations, the result

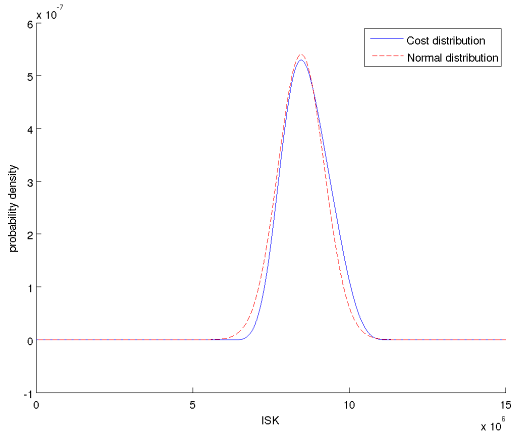


*Figure 6.2:* Plot of combined estimated cost distribution and normal distribution

will quickly lose the skew and become symmetric.

## 6.3 Server cost

With the data from the previous section, we can estimate the yearly cost of a server and its software. The costs of the servers are known, and they are in the following table. The data

*Table 6.2:* Yearly server costs

| | |
|---|---|
| Server | 200.000 |
| Operating system license (yearly) | 150.000 |
| Software license | 300.000 |
| Maintenance | 200.000 |
| total | 850.000 |

center in question has 16 racks, each with 42U. Of the available rack space, 75% is used. The server in question is a 4U server. The data center has a total capacity of $42 \cdot 16 = 672U$. Using the estimated cost, the costs are summarized in table 6.3.

*Table 6.3:* Data center costs related to capacity

| | |
|---|---|
| Cost of used space | $8.472.000 \cdot 75\% = 6.350.000$ |
| Cost of unused space | $8.472.000 \cdot 25\% = 2.150.000$ |
| Cost per single 'U' | 12.600 |
| Cost for 4U server | 51.000 |
| 95% confidence interval | 44.400 to 61.200 |

So the end result is, that the yearly costs for our server is,

$$\text{Yearly cost for server} = 901.000 \tag{6.1}$$

$$95\% \text{ confidence interval} = 894.400 \text{ to } 911.200 \tag{6.2}$$

# 7 Conclusions and future work

In this chapter an attempt will be made to identify some of the limitations of the methods used. Further development of the cost determination model will be identified, and concluding remarks will be made.

## 7.1 Limitations

When attempting to determine costs, methods used must often by very general and not to specific. The reason for this is that costs invariably come in unforeseen circumstances or highly specialized ways that might be hard to measure. The attempt made here is rather specific in its grouping of costs. Although the assumption made is that the way the costs are grouped should apply to most companies, it is not a one size fits all solution.

While the costs that are explained and estimated in this thesis are mostly so called 'hard' costs, i.e. costs that there is some kind of receipt for, there are various other costs, so called 'soft costs' that are unaccounted for. These are for example the so called 'costs of complexity'(Drury, 2001). These costs include for example the costs incurred by companies that have to maintain many different technologies or standards. Other costs that are difficult to analyze are costs of overhead, or the costs of downtime, i.e. what does it cost the company in lost productivity when the main ERP system goes down?

It must be noted that costs are only a part of the bigger picture. Having the costs of IT is useful, but it is the value that IT brings that is most important. Establishing a system that makes IT costs readily available will possibly bring unwelcome pressure to reduce costs, at the expense of IT performance and end user satisfaction. For this reason the determination of IT costs should be only a part of the overall strategy to assess the value IT brings to a company.

## 7.2 Future work

This thesis proposes various cost distributions. Research on how cost distribution estimates correlate with actual costs would be quite useful. Such work will be made harder by the fact that few organizations readily give out their cost figures in such a specific manner.

A useful addition to incorporate into IT cost structures is real options analysis. That is, incorporating put and call option valuation techniques. With the methods mentioned here, comparing two data centers by their costs will not take into account the additional features of a data center such as increased power redundancy, except by higher costs per square meter.

With a sufficiently developed and specific cost model, a great variety of opportunities for IT optimization opens up. With virtualization technologies virtual machines can be moved between servers in an instant. This, along with having cost estimates, gives companies a chance to computationally minimize costs and maximize computing power at the same time.

Development of IT specific accounting practices would be a valuable asset in cost measurement of IT. This is especially relevant as IT becomes an ever larger part of companies' expenditures. Information on accounting and IT can be found in the following paper (Dehning and Richardson, 2001) .

## 7.3 Conclusion

This thesis proposes a method of estimating costs in IT and of categorizing them. Having and maintaining costs of IT has never been as important, and will give companies a competitive edge. This field has great potential as companies have a desire of cutting costs in an effective way, while maximizing the effectiveness of their IT infrastructure. With companies contemplating the adoption of Cloud computing and SaaS offerings, discovering their actual costs is vital if they want to make informed decisions.

The proposed model offers a quick way of getting cost estimates, without hiding the inaccuracy of the estimates, as long as the cost distributions are reasonable. Knowing the accuracy of the estimate motivates businesses to improve the cost estimates until they have an acceptable granularity. With this information, cost drivers will be more easily discovered, and this will in turn assist managers in making business decisions.

Organizing costs into different 'levels' will also facilitate the understanding of the complexity of a company's IT infrastructure. Making comparisons between IT technologies and data centers in particular is made easier, as this will be an apples to apples comparison. The end result will be a breakthrough in measuring IT ROI and other similar metrics.

# A Application Example Matlab code

```
% Example of usage of the probability method
% Numbers are in thousands, but should be in millions
m = 10^6;
x = 0:15*m;
% Level 1 costs
level1func = m*[3 4 6];
level1 = trimf(x, level1func);
%trimf gives a density where the peak has a value of 1, need to scale it down
%so that the area of the triangle is 1. area of triangle is given by A = 0.5 *b*h
L1height = 2 / (level1func(3)-level1func(1));
level1 = level1.* L1height;
% trapz(level1) should give a result of 100
% Level 2 costs
level2func = m*[0.25 0.3 0.35];
level2 = trimf(x, level2func);
L2height = 2 / (level2func(3)-level2func(1));
level2 = level2.*L2height;
% Level 3 costs
level3func = m*[2 2.25 2.75 3];
level3 = trapmf(x, level3func);
%Area of a trapezoid is given by A = h*(b1 + b2)/2
% where b1 is length of bottom and b2 is length of top
L3height = 2/ (level3func(4) - level3func(1) + level3func(3) - level3func(2));
level3 = L3height.*level3;
% Peripheral costs
levelPfunc = m*[1 1 2 2];
```

```
levelP = trapmf(x, levelPfunc);
LPheight = 1 / (levelPfunc(4) - levelPfunc(1));
levelP = LPheight.*levelP;

% Plot of costs

figure(1)
subplot(2,2,1)
plot(x, level1);
title('Level 1')
xlabel('ISK')
ylabel('probability density')
subplot(2,2,2)
plot(x, level2);
title('Level 2')
xlabel('ISK')
ylabel('probability density')
subplot(2,2,3)
plot(x, level3);
title('Level 3')
xlabel('ISK')
ylabel('probability density')
subplot(2,2,4)
plot(x, levelP);
title('Level P')
xlabel('ISK')
ylabel('probability density')
print -dpng exampleLevels.png

result = real(ifft(fft(level1).*fft(level2).*fft(level3).*fft(levelP)));




%Isolate distribution

ratio= 1000;
```

```
[maxD maxIndex]= max(result);
fprintf('Mean is %d ISK \n', maxIndex);
ratio2max = maxD/ratio;
startPoint=0;
endPoint=0;
for i=1:length(result)
        if i < maxIndex && startPoint == 0
                if result(i) > ratio2max
                        startPoint = i;
                end
        end
        if i > maxIndex && endPoint == 0
                if result(i) < ratio2max
                        endPoint = i;
                        break;
                end
        end
end

distro = result(startPoint:endPoint);
interval = 1:length(distro);
m = mean(interval);
% Variance
ss = sum(((interval-m).^2).*distro);
fprintf('Variance is : %f \n', ss);
%Compare to Normal distribution
figure(3)
hold on
plot(x, result)
plot(x,normpdf(x, maxIndex, sqrt(ss)), 'r--');
legend('Cost distribution', 'Normal distribution')
xlabel('ISK');
ylabel('probability density');
hold off
print -dpng exampleResult.png
%Numeric method to find 95% confidence interval
%Naive approach, begins at the extreme ends and works its way in until it hits 95%
```

```
x=fminbnd(@(x)(trapz(distro(1+ceil(x):length(distro)-ceil(x)))-0.95)^2,0,length(di

conf = [startPoint + x, endPoint - x];
fprintf('95%% confidence interval is from is %d ISK to %d ISK\n', conf(1), conf(2)

% Datacenter stuff
dUsed = 0.75;
availableU = 16*42;

fprintf('Cost of used space %d ISK \n', maxIndex*dUsed);
fprintf('Cost of unused space %d ISK \n', maxIndex*(1-dUsed));
serverCost = maxIndex / availableU * 4;
fprintf('4U server cost %d ISK \n', serverCost);
```

# Bibliography

(2006). The value of integrating it asset management with erp applications - white paper. http://www.rubiksolutions.com/Admin/Public/DWSDownload.aspx?File=%2FFiles%2FFiler%2FPDF+Service+Management+Center%2FAsset+Center%2FWhite+paper%2F4aa0-6567enw.asset_erp_swp.pdf.

(2008). State of the data center. Technical report, Symantec. http://www.symantec.com/about/news/resources/press_kits/detail.jsp?pkid=sdcreport.

(2008). Val it framework 2.0. http://isaca.org/Template.cfm?Section=COBIT6&Template=/ContentManagement/ContentDisplay.cfm&ContentFileID=18930.

Anderson, M., Banker, R. D., and Hu, N. (2003). Returns on investment in information technology. Technical report, School of Management, The University of Texas at Dallas.

Bannister, F. (2004). *Purchasing and Financial Management of Information Technology*. Elsevier Butterworth-Heinemann.

Barfield, J. T., Raiborn, C. A., and Kinney, M. R. (2003). *Cost Accounting:Traditions and Innovations*. South-Western College Pub, 5 edition.

Bharadwaj, A. S., Bharadwaj, S. G., and Konsynski, B. R. (1999). *Information Technology Effects on Firm Performance as Measured by Tobin's q*, volume 45 of *Management Science*, pages 1008–1024. INFORMS.

Bojadziev, G. and Bojadziev, M. (2007). *Fuzzy Logic for Business, Finance, and Management*. Advances in Fuzzy Systems:Applications and Theory - Vol. 23. World Scientific, 2 edition.

Brandl, R. (2008). *Cost Accounting for Shared IT Infrastructures*. Gabler verlag.

Brynjolfsson, E. (1994). The productivity paradox of information technology: Review and assessment. *Communications of the ACM*.

Commission, I. E. (2008). Iec 60297-3-100.

Cotteleer, M. J. (2002). An empirical study of operational performance convergence following enterprise-it implementation. Technical report, Harvard Business School.

DeGroot, M. H. (1986). *Probability and Statistics*. Addison-Wesley publishing company, 2 edition.

Dehning, B. and Richardson, V. J. (2001). Returns on investments in information technology: A research synthesis. University of New Hampshire, University of Kansas.

Dewan, S., Shi, C., and Gurbaxani, V. (2007). Investigating the risk–return relationship of information technology investment: Firm-level empirical analysis. *Management Science*, 53(12):1829–1842.

Drury, D. H. (2001). Determining it tco: Lessons and extensions. In *ECIS*.

Garvey, P. R. (2000). *Probability methods for cost uncertainty analysis: a systems engineering perspective*. Marcel Dekker Inc.

Golden, B. (2009). The case against cloud computing, part four. *CIO*.

Harris, M. D., Herron, D. E., and Iwanicki, S. (2008). *The business value of IT*. Auerbach publications.

Hawkins, M. W. (2001). Total cost of ownership: The driver for it infrastructure management. Technical report, Prentice Hall Professional. http://www.informit.com/articles/article.aspx?p=21841.

Huang, K. (2007). Towards an information technology infrastructure cost model. Master's thesis, Massachusetts institute of technology.

Irani, Z. and Love, P., editors (2008). *Evaluating Information Systems, Public and Private Sector*. Butterworth-Heinemann.

Karidis, J., Moreira, J. E., and Moreno, J. (2009). True value: assessing and optimizing the cost of computing at the data center level. In *CF '09: Proceedings of the 6th ACM conference on Computing frontiers*, pages 185–192, New York, NY, USA. ACM.

Koomey, J. (2007). A simple model for determining true total cost of ownership for data centers. Technical report, Uptime institute. Version 2.

Love, P. E., Irani, Z., and Fulford, R. (2003). Understanding it costs: An exploratory study using the structured case method. In *7th Pacific Asia Conference on Information Systems*.

Maylor, H. (2003). *Project management*. Prentice Hall, 3 edition.

Montgomery, D. C. and Runger, G. C. (2007). *Applied Statistics and Probability for Engineers*. John Wiley & Sons, Inc., 4 edition.

Mutschler, B., Zarvi?, N., and Reichert, M. (2007). A survey on economic-driven evaluations of information technology.

Neumann, B. R., Gerlach, J. H., Moldauer, E., Finch, M., and Olson, C. (2004). Cost management using abc for it activities and services. *Management Accounting Quarterly*, 6(1):29–40.

O'Neill, P. and Krauss, D. (2006). Topic overview: It service management. http://www.forrester.com/go?docid=40558.

Oppenheim, A. V., Willsky, A. S., and Hamid, S. (1996). *Signals and Systems*. Prentice Hall, 2 edition.

Remenyi, D., Bannister, F., and Money, A. (2007). *The Effective Measurement and Management of ICT Costs and Benefits*. CIMA Publishing, 3 edition.

Roberts, R. and Sikes, J. (2009). It in the new normal: Mckinsey global survey results. https://www.mckinseyquarterly.com/Business_Technology/BT_Strategy/IT_in_the_new_normal_McKinsey_Global_Survey_results_2473.

Roztocki, N. and Weistroffer, H. R. (2004). Using activity-based costing for evaluating information technology related investments in emerging economies: A framework. In *Proceedings of the Tenth Americas Conference on Information Systems*, New York, NY, USA.

Sakamoto, J., Mori, Y., and Sekioka, T. (1997). Probability analysis method using fast fourier transform and its application. *Structural Safety*, 19(1):21–36. Asian-Pacific Symposium on Structural Reliability and Its Applications.

Weygandt, J. J., Kieso, D. E., and Kimmel, P. D. (2005). *Principles of financial accounting*. John Wiley & Sons, Inc., 7 edition.