MASTER

EPT and de-convolution of probability distributions

Zijlstra, B.

Award date:
2006

EPT and De-Convolution of Probability Distributions

B. Zijlstra

SE 420471

Final Assignment

Supervisor: Prof.dr.ir. J.T. Udding
Coach: Ir. M.P.M. Hendriks

EINDHOVEN UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF MECHANICAL ENGINEERING
SYSTEMS ENGINEERING GROUP

Eindhoven, June 2006

FINAL ASSIGNMENT

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mechanical Engineering
Systems Engineering Group

August 2005

Student	B. Zijlstra
Supervisor	Prof.dr.ir. J.T. Udding
Advisor	Ir. M.P.M. Hendriks
Start	August 2005
Finish	June 2006
Title	EPT and De-Convolutions of Probability Distributions

Subject

For supply chains and manufacturing systems, it is often hard to do a performance analysis. The performance engineers should have the availability of all the throughput times with respect to all the processes in the system. For a single workstation, several sources, such as setup, machine breakdown and natural processing time, can be identified that have influence on the throughput time of a lot. Some of these sources are often hard to measure. By the introduction of EPT, that treats those sources as a black-box, the performance analysis can be done more easily. However, by looking at a workstation with an EPT approach, the details get lost. To find out, which source is responsible for a workstation to be the bottleneck, is of great importance.

Assignment

An EPT data set can be seen as a probability density function. That function is the convolution of the probability density functions of all prevalent sources of variability at the workstation. Assuming we know the form of the distributions of those sources, it is possible to derive equations, that depend on the parameters of the distributions of the sources and on the moments of the EPT probability density function. With these equations it is possible to determine the time it takes for a lot to pass through every source of variability in the workstation, when only EPTs are available.
The objective of this assignment is to propose algorithms to de-convolute probability distributions in general and to illustrate this de-convolution on EPT data sets.

Prof.dr.ir. J.T. Udding

Ir. M.P.M. Hendriks

Systems Engineering

Department of Mechanical Engineering

Abstract

In this report the de-convolution of a probability distribution is described. Such a distribution is composed of several underlying probability distributions. In this context, de-convolution implies that the underlying distributions are determined when the convoluted distribution is assumed to be known.

First of all, an application for the de-convolution of a probability distribution is given, namely the de-convolution of an EPT data set. Effective process time, EPT, can be seen as a probability distribution that is composed of several underlying distributions. In this case, the underlying distributions are the sources of variability of a workstation in a manufacturing system. Those sources are for example the processing, setup and breakdown times.

To come to the de-convolution of a probability distribution, first the relationship between a number of distributions and their convoluted distribution is determined. The moment generating function of a probability distribution is used to find those relationships. With help of the equations from these relationships, it is possible to identify the forms of the underlying distributions that can or cannot be de-convoluted.
From the distributions for which de-convolution can be applied to, only the equations from the moment generating function have to be solved to find the parameters of the underlying distributions.
From the distributions for which de-convolution cannot be applied to, observations are necessary from the underlying distributions. Those observations are processed in a classical and a Bayesian statistical way, to estimate the parameters of the corresponding underlying distributions.

Hereafter, the described method for the de-convolution of a probability distribution is illustrated on an EPT data set. The de-convolution results in the estimation of the mean and variance of the sources of variability.

Finally, conclusions are drawn with respect to the findings for the de-convolution of probability distributions.

Samenvatting

In het verslag wordt een methode beschreven voor het uiteenrafelen (deconvolutie) van een kansdichtheidsfunctie. Zo een kansdichtheidsfunctie, ofwel distributie, is opgebouwd (convolutie) uit verscheidene andere kansdichtheidsfuncties. Deconvolutie betekent in deze context, dat de onderliggende distributies worden bepaald wanneer de geconvolueerde distributie bekend is.

In het eerste hoofdstuk wordt er een toepassingsgebied beschreven waarin de deconvolutie van een kansdichtheidsfunctie van belang kan zijn, namelijk het deconvolueren van een EPT data set. Effective process time, EPT, kan worden gezien als een kansdichtheidsverdeling die is opgebouwd door middel van een optelling van een aantal onderliggende kansdichtheidsverdelingen. De onderliggende verdelingen zijn dan de de bronnen van variabiliteit in een werkstation van een productie proces. Zulke bronnen zijn bijvoorbeeld processing tijden, set-ups en breakdowns.

Om tot een deconvolutie van een kansdichtheid te komen, wordt eerst het verband tussen de onderliggende kansdichtheden en de geconvolueerde kansdichtheid beschreven. Met behulp van de moment genererende functie van een kansdichtheidsfunctie is het mogelijk om deze verbanden te vinden. De vergelijkingen uit de verbanden maken duidelijk welke families van distributies wel en welke niet bruikbaar zijn voor deconvolutie van een kansdichtheid.
Voor de bruikbare distributies hoeven slechts de verbanden tussen de vorm van de geconvolueerde kansdichtheid en de parameters van de onderliggende distributies opgelost te worden.
Voor de niet bruikbare distributies, zullen er observaties nodig zijn uit de onderliggende distributies. Deze observaties worden verwerkt, met een klassieke en Bayesiaanse statische methode, voor een schatting van de parameters van de geobserveerde distributie.

Vervolgens zal de gevonden methode, voor het deconvolueren van een kansdichtheid, geïllustreerd worden op een EPT data set, zodat de distributies van de bronnen van variabiliteit benaderd worden. Uit deze benadering valt vervolgens af te leiden, wat het gemiddelde en de variantie van elke bron is.

Tenslotte zullen er conclusies worden getrokken met betrekking tot de aangedragen methoden voor de deconvolutie van kansdichtheden.

Contents

Introduction

This paper is concerned with the de-convolution of probability distributions. De-convolution is the reverse of the convolution of probability distributions. Convolution is the sum of, for example, two independent random variables that results in a probability distribution which is the *convolution* of each of the distributions of the two variables. Then, the de-convolution of that probability distribution will result in the distributions of the two independent random variables. (Notice that the two independent events occur sequentially in time. When they occur simultaneously, we speak of a mixture of probability distributions, which is not to be discussed in this paper.)

This paper is also concerned with the effective process times, EPT. This is an application into which the de-convolution of probability distributions can be used. EPT is a powerful tool to do a performance analysis in a production system. At a workstation, multiple sources of variability can be identified that can influence the throughput time of a part, such as the natural processing time, setup time, operator availability and machine failure. These sources are hard to measure and therefor it is hard to do a performance analysis. By the introduction of EPT, that treats the underlying sources of variability at a workstation as a black box, it is more easy to do such a performance analysis as is described in [Jac03]. Even for modelling, validation and control of manufacturing systems, the EPT approach is a valuable tool as described in [Lef04]. By the introduction of EPT, the computional complexity for the evaluation of large models, as can be found in the semiconductor manufacturing industry, is reduced considerably.

In earlier studies for the performance analysis of manufacturing systems with the EPT approach, we are able to identify the workstation that is the bottleneck in the system. Which underlying source is responsible for the workstation to be the bottleneck, cannot be retrieved directly from the measured EPT data. In this paper a method is proposed to determine the parameters of the probability distributions of each source of variability at the workstation, just by looking at the EPT data set. With help of the de-convolution of the EPT data, it is possible to compute, for example, the mean and variance of each source of variability at a workstation. Then, it is possible to identify the source that causes the workstation to be the bottleneck.
Another advantage of the de-convolution of the EPT data, is to better approximate the probability distribution of the EPT. In the studies of simulating manufacturing systems with EPTs, the distribution of the EPT data set is used for the simulation of

that system. Such an EPT data set has often no closed form probability distribution. In that case, a closed form probability distribution, like the gamma distribution, is fitted on the EPT data set, which can be a poor approximation of the actual EPT distribution. With the convolution of determined the probability distributions of the sources of variability, it is possible to generate a better approximation of that EPT distribution.

The structure of this paper is as follows. In the first chapter, an introduction to EPT is given. Also the use of the de-convolution of the EPT data sets is illustrated.
In the second chapter, algorithms are proposed from which it is possible to define a set of equations to (de-)convolute probability distributions in general. These algorithms are based on a kind of Fourier transformation of the probability distributions that are convoluted. In that same chapter, it will become clear which underlying distributions are suitable (for example, the exponential distribution) and which are not suitable (for example, the normal distribution) for the de-convolution of a probability distribution.
In the third chapter three methods are proposed, to solve the set of equations from the algorithms to the de-convolute the probability distributions. First, two methods are proposed for a distribution that is not suitable (there are more unknowns than equations) to solve those equations. These two methods are based on classical statistics and on Bayesian statistics for the estimations of the unknown parameters. Hereafter, a method is proposed for a distribution that is suitable to de-convolute with the equations from the second chapter. This method is based on an optimization problem, which makes it possible to solve the non linear equations from the algorithms for the de-convolution of the probability distribution.
In the fourth chapter, the proposed method to solve the non linear equations for the de-convolution of a probability distribution, is demonstrated on EPT data sets.
In the last chapter, conclusions are drawn with respect to the findings in this paper.

Chapter 1

Effective Process Time (EPT)

Effective process time, EPT, is an application into which the de-convolution of probability distributions can be used. In this chapter, the line of thought of that EPT is explained.
Hopp and Spearman have introduced the term EPT for the first time, and is defined as [Hop01]:

> The effective process time of a job at a workstation is a random variable. The label *effective* is used as a reference to the total time "seen" by a job at a station. This is done from a logistical point of view. If machine B is idle because it is waiting for a job to finish on machine A, it does not matter whether the job is actually being processed or is being held up because machine A, for example, is being repaired or undergoing a setup. To machine B, the effects are the same. For this reason, we will combine these and other effects into one aggregate measure of variability, namely the effective process time.

Thus, EPT can be seen as the total amount of time a part could have been, or actually was, processed in a workstation. Hopp and Spearman derived the mean and variance of the EPT straightforwardly. First they distinguish the underlying sources of variability at a workstation, such as:

- Natural process time
- Random outages (eg. machine breakdown or repair)
- Setups
- Operator availability
- Recycle or rework

The queueing time of a part in front of a workstation is **not** and should **not** be incorporated in the computation of the EPT.
The mentioned sources of variability do all contribute to the mean and variance of the EPT, through the relations that are mentioned in [Hop01]. The mean and variance of every source should be known to determine the EPT of a job at a workstation. This is because [Hop01] convolutes the underlying sources of variability to determine the EPTs.

1.1 Probability distributions of the sources of variability

The probability distributions of the underlying sources of the EPT are right skewed. This means that the tail of the density function, PDF or f, is on the right side. Because of the right skewness, negative times do not show up. Right skewed distributions are, for example, the exponential, gamma or Weibull distribution. The gamma distribution is a useful distribution to approximate each of the underlying sources of variability at the workstation [Law00].
In this paper we are especially concerned with a workstation that have the following underlying sources of variability:

Operator availability

An operator that operates the workstation is a source of variability. In an optimal process, the operator is always immediately available to serve the machine. In practice, however, that operator can be gone for a break or is busy doing something else. The time it takes before the operator is ready to serve the workstation, is often assumed to be exponentially distributed. The exponential distribution is a special case of the gamma distribution, namely $Exp(\theta) \equiv Gam(1, \theta)$. In figure 1.1, a corresponding probability density function, PDF, is presented.

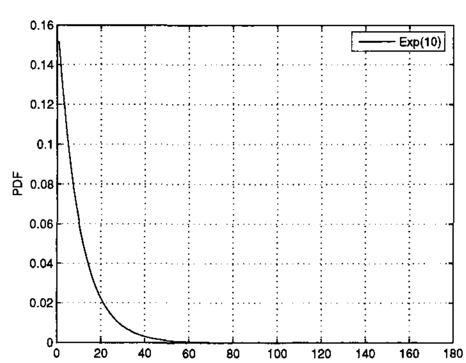

Figure 1.1: Operator availability

Setup time

When a part arrives at a workstation, a setup may be necessary before the part can be processed. That setup time is often assumed to have a gamma distribution. In figure 1.2, a corresponding density function is presented.

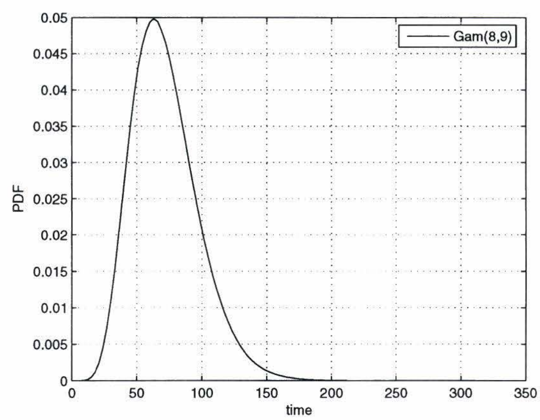

Figure 1.2: Setup

Processing time

The processing time is derived from two possible events that can occur at a workstation. When the machine does not fail, the processing time will equal the natural processing time, which is often assumed to be gamma distributed. However, when the machine breaks down the processing time equals the natural processing time plus the repair time. The sum of the natural processing time and the repair time is also assumed to be gamma distributed. With a certain probability, (1-p), that a machine breaks down we can create a mixed distribution, namely $p \cdot Gam(\alpha_1, \beta_1) + (1 - p) \cdot Gam(\alpha_2, \beta_2)$. That distribution contains both possible events, namely without breakdown and with breakdown, respectively. In figure 1.3, a corresponding density function is presented for a 20 % probability of breakdown.

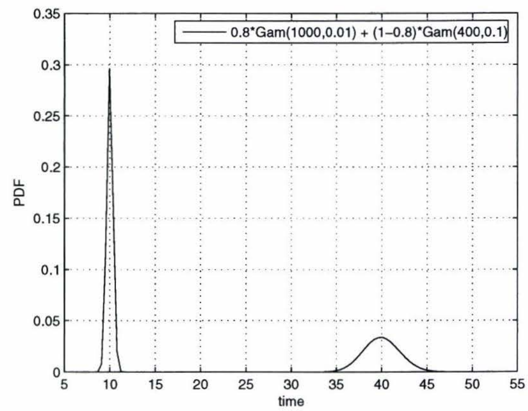

Figure 1.3: Processing time

EPT

The convolution of all the sources of variability that are mentioned, results in the EPT. The convolution of the density functions of figure 1.1 and 1.2 and 1.3, is presented in figure 1.4.

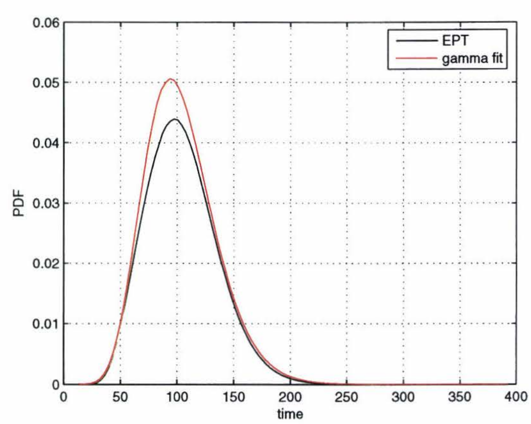

Figure 1.4: EPT and gamma fit on that EPT

In that same figure a gamma fit is plotted, that approximates the EPT density function. This is a usual method to find a distribution that corresponds with the EPT and from which samples can be taken when evaluating manufacturing models, as is described by the two-moments distribution fitting method in [Koc05]. With slightly different parameters for the distributions of the sources, the EPT density function cannot be fitted with a closed form distribution. Such an EPT density function is presented in figure 1.5.

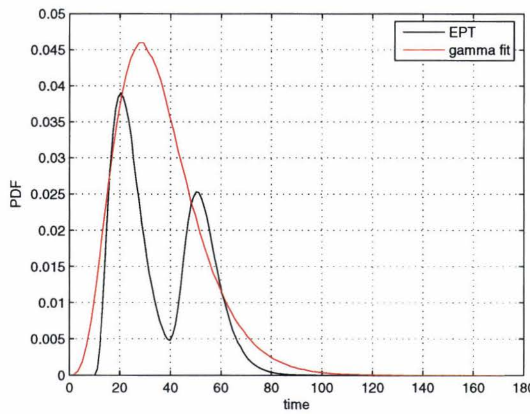

Figure 1.5: EPT and gamma fit on that EPT

The EPT distribution, that corresponds to the density function as in figure 1.5, is not suitable to fit a gamma distributions onto. That fit is a poor approximation of the density function corresponding to the EPTs. Samples that can be taken from the fitted distribution for manufacturing system models, as explained in the next section, are not optimal at all.

1.2 Applications of EPT

The two main performance criteria of a supply chain or manufacturing system, are the throughput and cycle time of the processed parts or lots. To improve the performance of such a system, the capacity losses in that system should be determined. The mean EPT and the corresponding variance are two fundamental parameters with respect to the throughput and cycle time performance. In [Jac03], a method to compute EPTs from realtime fab data is proposed for single and multiple machines with FIFO and general dispatching. Those EPT algorithms correctly compute the mean and variance of the EPT, when only track-in and track-out data is available from each workstation in the system. A case study in [Jac03] shows that the main causes of large cycle times can be identified with respect to the workstations that are present in the studied system.

In [Lef04], the EPT is used to model, validate and control manufacturing systems in the semiconductor industry. Although, the flow time of a wafer fab is in the order of two months, the natural processing time of a wafer is less than two weeks. The commonly used discrete-event and fluid models, that are used to analyze the performance of the wafer fab, can become very computionally complex when taking all the underlying sources of variability into account. By considering the EPTs as a conceptual way to cover all the variability at a workstation, the complexity of the model is considerably reduced.

Chapter 2

(De-)Convolution of Probability Distributions

In this chapter, a general method to de-convolute a probability distribution, denoted as F, is explained. An important aspect in that method is the moment generating function, which can be seen as a kind of Fourier transformation of the corresponding probability density function, PDF or f, of the distribution F. In case of convoluting and de-convoluting probability distributions, the moment generating function lends itself well to calculate with. In this chapter two distributions are examined, namely one which is not suitable to de-convolute and one that is suitable to de-convolute. These are the normal and exponential distribution, respectively.

First the de-convolution of the EPT data set, mentioned in chapter 1, is transformed into a more generalized problem:

Problem:
A data set of process Y is subject to a certain probability distribution, say F_Y. That process consists of a certain number, say n, of sequential sub-processes X_i, $0 < i \leq n$, which are independent from each other. These sub-processes are also distributed with a certain probability distribution, say F_{X_i}.
Assume that we know the data set of process Y and its distribution, F_Y. Furthermore, assume that we know how the sub-processes X_i are distributed. Is it, then, possible to determine the parameters of the F_{X_i}'s?
An illustration of this problem is in figure 2.1.

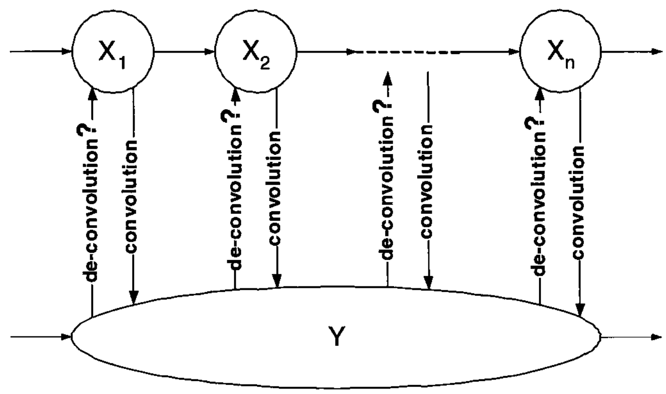

Figure 2.1: Y and the underlying processes, X_i

To come to a solution of this problem, first is investigated, how the distributions, F_{X_i}, of the sequential sub-processes contribute to the shape of the distribution, F_Y. Hereto, the concept of moment generating functions is introduced from which the moments of a probability distributions are determined. From these moments, the shape parameters of the distribution are derived. Common shape parameters are the mean, variance, skewness and kurtosis. The mean represents the value around which central clustering occurs. The variance gives an indication of the function "width" or "variability" around that value. The skewness characterizes the degree of asymmetry of the probability density function with respect to its mean. The kurtosis gives an indication of how peaked that density function is with respect to the normal distribution.

2.1 Moment generating functions

To determine the moments of a continuous probability distribution F with probability density function f, the moment generating function is defined by [Mon99]:

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx \ , \tag{2.1}$$

The moment generating function $M(t)$ will only exist if the integral from (2.1) converges. Then it is possible to determine the moments of F from this generating function.

There are various ways to determine the shape parameters from (2.1). These parameters are derived from the *centered* and *uncentered moments* of the probability distribution. The definitions of the *uncentered moments* (with respect to the origin of the density function), is given by [Mon99]:

$$u_r = \left(\frac{d^r M(t)}{dt^r} \right) (0) \tag{2.2}$$

From (2.2), the mean and variance of distribution F, are determined by:

$$\begin{aligned} \text{mean}(F) &= u_1 \\ \text{variance}(F) &= u_2 - {u_1}^2 \end{aligned}$$

Considering the property of moment generating functions [Mon99]: $M_{F+a}(t) = e^{at}M(t)$, the *centered moments* (with respect to the mean of F), are given by:

$$uc_r = \left(e^{-u_1 t} \cdot \frac{d^r M(t)}{dt^r} \right)(0) \tag{2.3}$$

From (2.3), the shape parameters are determined by:

$$\begin{aligned} \text{variance}(F) &= u_2 - {u_1}^2 = uc_2 \\ \text{skewness}(F) &= \frac{uc_3}{{uc_2}^{3/2}} \\ (\textit{normal})\ \text{kurtosis}(F) &= \frac{uc_4}{{uc_2}^2} \\ \text{kurtosis}\ \textit{excess}(F) &= \frac{uc_4}{{uc_2}^2} - 3 \end{aligned}$$

In this paper the normal kurtosis is used as the shape parameter for the kurtosis. Because this kurtosis equals three for a normally distributed density function, in some literature it is scaled to zero for the normal distribution and is called the kurtosis *excess*.

Convolution

When convoluting multiple independent random variables, X_i, the moment generating function of the total probability distribution, F_y, is determined [Mon99] by:

If $X_1, X_2, \ldots, X_n$ are independent random variables with moment generating functions $M_{X_1}(t), M_{X_2}(t), \ldots, M_{X_n}(t)$, respectively, and if $Y = X_1 + X_2 + \cdots + X_n$, then the moment generating function of Y is given by (2.4) :

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) \tag{2.4}$$

This is, in this paper, the most important property of the moment generating functions. It is used to derive equations, that depends on the parameters of the underlying distributions, F_{X_i} and the moments of the convoluted distribution F_Y.

In the next section, the shape parameters of several distributions are determined. Conclusions are drawn with respect to the use of the moment generating function to (de-)convolute probability distributions.

2.2 Normal distributions

The moment generating function of a normal distribution with density function, $f = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, is [Joh94]:

$$M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \tag{2.5}$$

The shape parameters of the corresponding distributions F, are determined by (2.2) and (2.3):

$$\begin{aligned}
\text{mean}(F) &= \mu \\
\text{variance}(F) &= \sigma^2 \\
\text{skewness}(F) &= 0 \\
\text{kurtosis}(F) &= 3
\end{aligned}$$

2.2.1 Convolution

Using equation (2.4), the moment generating function of a sequence of normal distributions, is determined by:

$$M_Y(t) = \prod_{i=1}^{n} e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}} \quad = e^{(\sum_{i=1}^{n} \mu_i)t + \frac{(\sum_{i=1}^{n} \sigma_i^2)t^2}{2}} \tag{2.6}$$

The corresponding probability distribution, F_Y, is again normally distributed with density function:

$$f_Y \sim N(\sum_{i=1}^{n} \mu_i, \sqrt{\sum_{i=1}^{n} \sigma_i^2}) \tag{2.7}$$

The shape parameters of F_Y are:

$$\begin{aligned}
\text{mean}(F_Y) &= \sum_{i=1}^{n} \mu_i \\
\text{variance}(F_Y) &= \sum_{i=1}^{n} \sigma_i^2 \\
\text{skewness}(F_Y) &= 0 \\
\text{kurtosis}(F_Y) &= 3
\end{aligned}$$

A small example to illustrate the above:

$$\begin{aligned} M_Y(t) &= M_{X_1}(t) \cdot M_{X_2}(t) \\ M_Y(t) &= e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} \cdot e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}} \\ M_Y(t) &= e^{(\mu_1+\mu_2)t + \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}} \end{aligned}$$

The convoluted probability distribution, F_Y, is normally distributed. The mean of F_Y equals $\mu_1 + \mu_2$ and the variance equals $\sigma_1^2 + \sigma_2^2$.

2.2.2 De-convolution

In this paper, we are also concerned with the de-convolution of the probability distribution, F_Y. One can conclude that it is not possible to determine the distributions, F_{X_i} if only F_Y is known. This is, because, if there are n processes ($n > 1$) there are $2n$ unknowns (μ_i and σ_i^2) and only two equations (one for the mean and one for the variance of F_Y, see section 2.2.1) from the moment generating function to solve this de-convolution problem.

Equation (2.6) states that the convolution of normal distributions yields again a normal distribution. With respect to de-convolution, one can say that a normal distribution, F_Y, can be subdivided into arbitrarily many, say n, processes that are normally distributed, F_{X_i}, as long as $\mu_Y = \sum_{i=1}^n \mu_i$ and $\sigma_Y^2 = \sum_{i=1}^n \sigma_i^2$ holds.

We can conclude that for the de-convolution of a normal probability distribution, F_Y, more information is necessary than only F_Y and the number of underlying processes n. Observations should be taken from those underlying processes, X_i, to determine the parameters, μ_i and σ_i^2, of the corresponding distributions. In section 3.1, two methods are proposed for the processing of the observations that are are taken from a process X_i, that results in an estimation for the corresponding μ_i and σ_i^2. The first method is based on classical statistics and the second is based on Bayesian statistics.

2.3 Exponential distributions

The next distribution we examine is the exponential distribution. The moment generating function of an exponential distribution with density function, $f = \frac{1}{\theta} e^{\frac{-x}{\theta}}$, is [Joh94]:

$$M(t) = \frac{1}{1 - \theta t} \tag{2.8}$$

The shape parameters of the corresponding distributions F, are determined by (2.2) and (2.3):

$$\begin{aligned}\text{mean}(F) &= \theta \\ \text{variance}(F) &= \theta^2 \\ \text{skewness}(F) &= 2 \\ \text{kurtosis}(F) &= 9\end{aligned}$$

2.3.1 Convolution

Using equation (2.4), the moment generating function of a sequence of exponential distributions, is determined by:

$$M_Y(t) = \prod_{i=1}^{n} \frac{1}{1-\theta_i t} \tag{2.9}$$

The corresponding probability distribution, F_Y, is not exponentially distributed. The shape parameters of F_Y are:

$$\begin{aligned}\text{mean}(F_Y) &= \sum_{i=1}^{n} \theta_i \\ \text{variance}(F_Y) &= \sum_{i=1}^{n} \theta_i^2\end{aligned}$$

For the skewness and kurtosis of the distribution F_Y, it is hard to find simple closed forms. If $n = 2$, the skewness and kurtosis are:

$$\begin{aligned}\text{skewness}(F_Y) &= 2\sqrt{\frac{(\theta_1^3+\theta_2^3)^2}{(\theta_1^2+\theta_2^2)^3}} \\ \text{kurtosis}(F_Y) &= \frac{9\theta_1^4+6\theta_1^2\theta_2^2+9\theta_2^4}{(\theta_1^2+\theta_2^2)^2}\end{aligned}$$

In case of the convolution of exponential distributions, the skewness and kurtosis of the distribution F_Y, are depending on the θ_i's. This property is helpful for the de-convolution of F_Y.

2.3.2 De-convolution

An example illustrates the use of the moment generating function to de-convolute the distribution F_Y and to determine the input parameters, θ_i, of the underlying distributions, F_{X_i}.
Consider the distribution F_Y, that is composed of two exponential distributions, $F_{X_{1,2}}$. When it is possible (there is enough data) to find the mean and variance of F_Y, then it is possible to find θ_1 and θ_2, because:

$$\begin{aligned} \text{mean}(F_Y) &= \theta_1 + \theta_2 \\ \text{variance}(F_Y) &= \theta_1^2 + \theta_2^2 \ , \end{aligned}$$

in which there are two equations containing two unknown variables. For the de-convolution of F_Y, the two equations need to be solved. The presents of F_Y guarantees the existence of a solution for the θ_i's. With respect to the uniqueness of this solution, a non linear equation, in general, may have more than one solution.
With the information of the natural bounds on the θ_i's (for example, $0 < \theta_i < \text{mean}(F_Y)$), the equations are solved. The output, say $\hat{\theta}_1$ and $\hat{\theta}_2$, does not necessarily correspond to θ_1 and θ_2, respectively. It is possible that the output $\hat{\theta}_1$ corresponds to θ_2 and the output $\hat{\theta}_1$ corresponds to θ_2. The method of de-convolution results in a solution that indicates which exponential distributions are in F_Y. The exact sequence cannot be determined.

If F_Y is built upon three exponential distributions (with three unknowns), the skewness will provide us with the third equation. Up to four exponential distributions, it is possible to determine the θ_i's with help of the four common shape parameters. If F_Y is built upon more than four exponential distributions, it will be necessary to use more moments from the moment generating function. Then, the equations become too complex to solve them analytically. In section 3.4, a method is proposed to solve these complex problems as an optimization problem.

2.4 (De-)Convolution of a normal and exponential distributions

The normal distribution is not suitable to de-convolute into a set of unique distributions. In contrast, the convolution of a set of exponential distribution is suitable for de-convolution. In this section, we examine a convoluted probability distribution, F_Y, containing one distribution from the normal and n distributions from the exponential family.

Combining (2.5) and (2.8) with (2.4), the moment generating function, corresponding to the F_Y, is:

$$M_Y(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \cdot \prod_{i=1}^{n} \frac{1}{1 - \theta_i t} \tag{2.10}$$

From (2.10), it is possible to generate the centered and uncentered moments using (2.2) and(2.3), respectively, which is used to determine the shape parameters of F_Y:

$$\begin{aligned} \text{mean}(F_Y) &= \mu + \sum_{i=1}^{n} \theta_i \\ \text{variance}(F_Y) &= \sigma^2 + \sum_{i=1}^{n} \theta_i^2 \end{aligned}$$

For the skewness and kurtosis of F_Y, again, it is hard to find simple closed forms. For $n = 2$, that is one normal and two exponential distributions, the skewness and kurtosis are:

$$\begin{aligned} \text{skewness}(F_Y) &= 2\sqrt{\frac{(\theta_1^3 + \theta_2^3)^2}{(\sigma^2 + \theta_1^2 + \theta_2^2)^3}} \\ \text{kurtosis}(F_Y) &= \frac{3(\sigma^4 + 2(\theta_1^2 + \theta_2^2)\sigma^2 + 3\theta_1^4 + 3\theta_2^4 + 2\theta_1^2\theta_2^2)}{(\sigma^2 + \theta_1^2 + \theta_2^2)^2} \end{aligned}$$

In this case, there are four equations and four unknown parameters: μ, σ, θ_1, θ_2. For the de-convolution of F_Y, the four equations need to be solved. The presents of F_Y guarantees the existence of a solution for those parameters. With respect to the uniqueness of this solution, those non linear equations may have more than one solution. With the information of the natural bounds (for example, $0 < \mu, \theta_{1,2} < \text{mean}(F_Y)$), the equations are solved. The output for μ, say $\hat{\mu}$, as well as the output for σ, say $\hat{\sigma}$, corresponds to each other. The output of the θ_i's, say $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$, again, do not necessarily correspond to θ_1, θ_2 and θ_3, respectively. The method of de-convolution results in a solution that indicates which exponential distributions are in F_Y. The exact sequence cannot be determined.

Chapter 3

Methods for the De-convolution of Probability Distributions

In this chapter, the theory of chapter 2 is used to propose methods for the de-convolution of a probability distribution, F_Y, into the sub-probability distributions, F_{X_i}. We assume the data set corresponding to the distribution F_Y, is known.
First the convoluted normal distribution is examined. Because this distribution is not suitable to apply the equations from the moment generating function, two methods are proposed for the estimation of the mean and variance of the distribution F_{X_i}. Those methods are based on classical statistics and on Bayesian statistics. Hereafter, a method is proposed to solve the equations from chapter 2 for the de-convolution of exponential distributions.

3.1 Normal distributions

In this section, the convoluted normal distribution, F_Y, is examined. Again, assume that the distribution is composed of the underlying normal distributions X_i, $0 < i \leq n$. The algorithm from section 2.2.2, provides us with two equations, one for the mean and one for the variance of F_Y. To determine all the distributions F_{X_i}, the parameters, μ_i and σ_i^2 have to be derived. If we know $(n-1)$ of the μ_i's and $(n-1)$ of the σ_i^2's, it is possible to solve the mentioned equations. In practice, however, we do not always know these parameters. In that case, experiments or observations from the processes X_i are necessary, to estimate the parameters μ_i and σ_i^2.
This section proposes two methods to process these observations and to determine the distributions F_{X_i}. The first method is based on the classical statistical philosophy and the second method is based on the Bayesian statistical philosophy.
The philosophy of classical mathematics, or the frequentists (notice the word frequent!), say that every proposition is held to be derived from observations or experiments. In contrast, the philosophy of Bayesian mathematics says that the mathematical theory

of probability is applicable to the degree to which a person believes a proposition. The Baysian mathematics also adapt Bayes' theorem that can be used as the basis for a rule for updating beliefs when new information is available (Bayesian inference). Those Bayesian beliefs are extensively described in [Lee97] and [Gel04].

With a small example, the difference between those philosophies, is illustrated:
Suppose, we want to know if a one euro coin is fair to toss. To find out, we do an experiment with, for example, one thousand throws. After this experiment, we count how many times heads and how many times tails has been thrown. If the outcomes are about equal, the coin is supposed to be fair. But what if we take another euro coin, and we want to find out if this coin is also fair? To find out, the frequentists say that for this new proposition, a whole new experiment is needed. The experiences from the former experiment do not apply to the new coin.
The Bayesian mathematics, however, do take the knowledge of the first experiment into account. If the first coin is expected to be fair, the new euro coin is also expected to be fair before the new experiment has started. This knowledge is then considered as the prior beliefs of the new experiment. If a multiset of observations of a new experiment is available, the information of that new experiment is modified according to the prior beliefs. This inference between the observations and prior beliefs, results in a prediction with respect to the fairness of the coin.
In this paper an application where those Bayesian beliefs can be useful, is for example, the determination of the natural processing time of, say, machine A. Suppose, we have knowledge about the natural processing time of a comparable machine, say machine B, then these prior beliefs can be used to determine the natural processing time of machine A. If the prior beliefs are accurate, then fewer observations, compared to a classical experiment, are needed to find the actual natural processing time of machine A.

In this section, we first consider one normally distributed proces, say X_1, and estimates are provided for the parameters, μ_1 and σ_1^2, based on a set of observations. Then, more underlying normally distributed processes, X_i, are examined to estimate all the parameters, μ_i and σ_i^2. If we assume that the convoluted normal distribution, F_Y, is known, experiments are done with respect to the processes, X_i. Hereto, in section 3.2, a theorem is derived to update the mean and variance of all the probability distribution, F_{X_i}, when observations from a new experiment are available. This method of updating is finally demonstrated for a probability distribution, F_Y, that is composed of two normal distributions.

3.1.1 Classical statistics

The classical approach to find the parameters, μ and σ^2, of a normal probability distribution, F, is to determine the mean and variance from an experiment that contains k observations ($\{j : 0 < j \leq k : y_j\}$).

The mean is determined by [Mon99]:

$$\bar{y} = \frac{1}{k}\sum_{j=1}^{k} y_j \ , \tag{3.1}$$

and the variance is determined by [Mon99]:

$$s^2 = \frac{1}{k-1}\sum_{j=1}^{k}(y_j - \bar{y})^2 \ . \tag{3.2}$$

The mean, $\bar{y}$, is an estimate for μ and the variance, s^2, is an estimate for σ^2. The more observations are taken (increasing k), the better the estimations are.

3.1.2 Bayesian statistics

In this section, the parameters, μ and σ^2, of a normal distribution are estimated with a Bayesian approach. This approach is based on updating beliefs when new observations are available, which is called: Bayesian inference. We assume that we have prior information about the normal probability distribution, F, before new observations are available. These priors are stochastic variables. With the Bayesian inference it is possible to create a loop for the prediction of μ and σ^2. First we define the prior beliefs on the mean and variance of a normal distribution. Then, we collect a new multiset of observations and we update those observation with the prior beliefs. The updated multiset of observations, results in the posterior predictive density for the normal density function that is examined. From this posterior predictive density, the prior beliefs are adjusted. When a new multiset of observations is available, that multiset of observations is again updated with the adjusted prior beliefs. An illustration of the Bayesian inference is in figure 3.1.

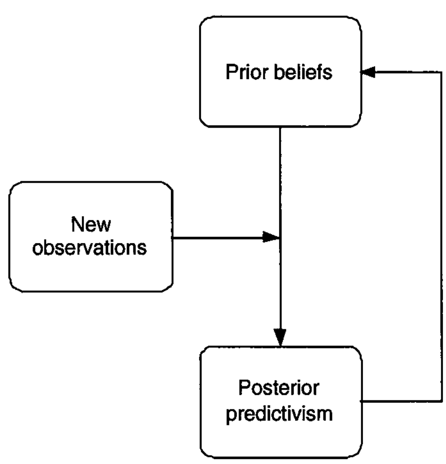

Figure 3.1: Bayesian inference

Suppose we do not know the parameters, μ and σ^2, exactly. But there is information available with respect to the mean of the normal distributions. In that case we are allowed to define a conjugate prior on the mean distribution. *Conjugate* implies that the prior distribution has the same form as the distribution from which the observations are taken, namely from the normal distribution. Then the posterior density is also of the normal form, which, again, results in a normally distributed prior when new observations are available.
In this section, only the necessary background is provided to understand how to estimate the mean and variance of the normal distribution with a simulation based on the Bayesian inference. A complete overview of this Bayesian approach can be found in [Lee97] and [Gel04].

The posterior predictive distribution, $p(\tilde{y}|y)$, of a future observation, $\tilde{y}$, given the sampled data, y, is given by [Gel04]:

$$p(\tilde{y}|y) = \int_{\text{all }\theta} p(\tilde{y}|\theta)p(\theta|y)d\theta \tag{3.3}$$

Where:

$\tilde{y}$	:	The prediction of a future observation.
y	:	The new observations.
θ	:	Prior beliefs, in this case $\theta = [\mu]$.
$p(\tilde{y}\|y)$	:	The distribution of $\tilde{y}$, given new observations, y.

For the understanding of equation 3.3, suppose, observations of the normal distribution are available (there are data points from this distribution). Then, the prior beliefs (the knowledge about that distribution before the data points were available) and the information of the observations (the data points) together, will result in a prediction of a future observation (an update of the former knowledge about the distribution), namely the posterior predictive distribution.

An important topic in the Bayesian beliefs is how to define proper prior distributions. In this paper, the assumption that there is knowledge about the mean of the normal distributions, means that we use a conjugate prior distribution for the mean [Lee97].

Variance (σ)

We have assumed, there is only prior information of the mean available for the estimation of the parameters of the normal distribution. Thus, all the information of the variance is only available through the new observations and from these observations the marginal posterior distribution, σ_k^2, is determined. The adjective *marginal* implies that the posterior on the estimation of the variance of the observations is determined by ignoring the prior knowledge. The marginal posterior variance does not necessarily coincides with the variance, σ^2, of the posterior predictive distribution. Because the estimation of the mean does also include variability. To come to a proper marginal posterior distribution of the variance, $p(\sigma_k^2|y)$, is estimated by [Gel04]:

$$\sigma_k^2 \mid y \sim \text{inv-}\chi^2(k-1, s^2) \tag{3.4}$$

Where k is again the number of observations and the quantity s^2 is the variance of the collected data, as in equation (3.2). The scaled inv-$\chi^2(k-1, s^2)$ equals the distribution of $\frac{(k-1)\cdot s^2}{\chi^2_{(k-1)}}$, i.e., it is the inverse of the more usual χ^2 distribution with $k-1$ degrees of freedom that is scaled by the quantity $(k-1)\cdot s^2$.

Mean (μ)

The conjugate prior distribution on the mean parameter of the normal distribution, is [Gel04]:

$$\mu \sim N(\mu_0, \tau_0^2),$$

where the, so called, hyperparameters μ_0 and τ_0^2 are the initial guesses for the mean and the (un-)certainty of that guess of the mean, respectively. For an informative prior distribution on the mean, the precision $\tau_0^2 \to 0$. For a non-informative prior on the mean, the precision $\tau_0^2 \to \infty$. If a multiset of new observations, y, containing k data points, is available, the posterior mean is [Gel04]:

$$\mu|\sigma_k^2, y \sim N(\mu_k, \tau_k^2) \ ,$$

in which,

$$\mu_k = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{k}{\sigma_k^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{k}{\sigma_k^2}} \quad \text{, and the posterior precision,} \quad \frac{1}{\tau_k^2} = \frac{1}{\tau_0^2} + \frac{k}{\sigma_k^2} \ . \tag{3.5}$$

Where:

μ_k : The posterior of the mean.
μ_0 : The prior on the mean.
$\bar{y}$: The mean of the new observations, as in equation (3.1) .

τ_k^2 : The variance of the posterior mean.
τ_0^2 : The variance of the prior mean.
σ_k^2 : The marginal posterior of the variance of the new observations, as in equation (3.4).

k : The number of new observations.

From equation (3.5), the relationship between μ_k and the number of observations is visible. If $k \to \infty$ or as $\tau_0 \to \infty$, then $\mu_k \to \bar{y}$ and $\tau_k^2 \to \frac{\sigma_k^2}{k}$, the Bayesian approach coincides with the central limit theorem from the classical statistics.

This central limit theorem, from the classical statistics, basically states that as the sample size, k, becomes large, the following occur [Mon99]:

1. The distribution of the mean of the observations becomes approximately normal in this case $\mu \sim N(\mu_k, \tau_k)$. (In general, this is regardless of the form of the distribution from which the observations are collected.)

2. The mean of the observations, μ_k, and the variance, τ_k^2, of that mean approaches:

$$\mu_k = \mu \tag{3.6}$$

$$\tau_k^2 = \frac{\sigma_k^2}{k.} \tag{3.7}$$

Simulation

Equations (3.4) and (3.5) lend themselves well to simulate. The posterior predictive distribution of a future observation, $\tilde{y}$, $p(\tilde{y}|y)$, can be approximated by simulation, with simulation steps:

begin loop
  Draw (μ, σ_k^2) from $p(\mu, \sigma_k^2|y, \theta)$:
    1) draw σ_k^2 from $\sigma_k^2|y \sim \text{inv-}\chi^2(k-1, s^2)$
    2) draw μ from $\mu|\mu_k, \tau_k^2 \sim N(\mu_k, \tau_k^2)$
  Draw $\tilde{y}$ from $\tilde{y} \sim N(\mu, \sigma_k^2)$
end loop

The expectation is of the posterior predictive distribution, $\tilde{y}$, is $\mu = \mu_k$ and the variance is $\sigma^2 = \sigma_k^2 + \tau_k^2$.

Suppose there is prior information about the mean of a normal distribution, then the use of the Bayesian beliefs can be helpful to estimate the parameters, μ and σ^2, of that distribution. Especially, when only a few observations are available. In case of the de-convolution of a known normal probability distribution, F_Y, into underlying normal distributions, F_{X_i}, the simulation based on this Bayesian method can also be of interest. In case of two underlying normal distributions, F_{X_1} and F_{X_2}, only the parameters of one of the two distributions should be known. Then it is possible to determine the parameters of the other distribution with the equations from section 2.2.1. In case of this two distributions, F_{X_1} and F_{X_2}, we can, also, take observations from both processes, X_1 and X_2. When the first experiment is done at process X_1, μ_1 and σ_1^2 can be estimated by the simulation steps of this section. With the equations from section 2.2.1, the corresponding estimates for μ_2 and σ_2^2 are derived. If the next available observations are from the other proces, X_2, new estimates are determined for μ_2 and σ_2^2. In order to take also the estimated μ_2 and σ_2^2 of the first experiment into account, the estimated mean and variance of the first experiment are updated with the estimated mean and variance of the second experiment. Because, the corresponding observations are not taken from the same process, it is not possible to update them directly. In the next section, theorems are derived which makes it possible to update the mean and variance of a process, when new observations are available. This update need not be expressed in terms of the values of the formerly collected observations. Only the mean, variance and the number of the former observations are necessary to update the mean and variance of the corresponding process.

3.2 Updating the mean and variance

If new observations of a process are available, the estimates of the mean and variance of the formerly collected observation are updated with the estimated mean and variance of the new observations. In this section two theorems are derived. The first theorem updates the mean and the second theorem updates the variance of the former observations when new observations are available. These updates are not expressed in terms of the individual values of the formerly collected observations. Only the mean, variance and the number of the former observations are of interest. The interpretations of the collected observations, are:

- Y_l : The multiset of the formerly collected observations, $y_j,\ 0 < j \leq l$, containing $|Y_l| = l$ observations.
- Y_k : The multiset of the newly available observations, $y_j,\ l < j \leq l + k$, containing $|Y_k| = k$ observations.
- Y_m : The update of the multiset of all available observations, $y_j,\ 0 < j \leq m$, containing $|Y_m| = m$ observations, where $m = l + k$.

Theorems

Theorem 1: Update the mean

$$\bar{y}_m = \bar{y}_l + \frac{k(\bar{y}_k - \bar{y}_l)}{l+k} \tag{3.8}$$

Where:

$\bar{y}_l$	:	The mean of the multiset of formerly collected observations, Y_l.
$\bar{y}_k$	:	The mean of the multiset of the newly available observations, Y_k.
$\bar{y}_m$	:	The mean of the multiset of variance of all available observations, Y_m.

The proof of **Theorem 1** can be found in Appendix A.

Theorem 2: Update the variance

$$s_m^2 = \frac{1}{l+k-1}\left\{(l-1)s_l^2 + (k-1)s_k^2 + \left(\frac{lk}{l+k}\right)(\bar{y}_k - \bar{y}_l)^2\right\}$$

Where:

s_l^2	:	The variance of the multiset of the formerly collected observations, Y_l.
s_k^2	:	The variance of the multiset of the newly available observations, Y_k.
s_m^2	:	The update of the multiset of variance of all available observations, Y_m.

The proof of **Theorem 2** can be found in Appendix A.

Iteration steps to update the mean and variance

With help of the iteration steps of **Theorem 1** and **Theorem 2**, it is possible to collect new observations (new data points) and update the mean and variance of the formerly collected data with that new data set.

The iteration steps to update the mean and variance, are:

```
begin loop
  collect new observations, Y_k:
    determine ȳ_m with the iteration step of Theorem 1
    determine s_m^2 with the iteration step of Theorem 2
  update: Y_l = Y_m and l = m
end loop
```

The update from these iteration steps are exemplified in the next section. In that section, the means and variances of two normally distributed processes are estimated.

3.3 Two normal distributions

In this section, the means and variances of two processes, that have normally distributed throughput times, are estimated with the iteration steps of the previous section. The positioning of the processes are:

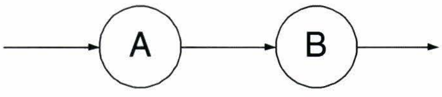

Figure 3.2: Two normally distributed processes

We consider that we have a large amount of data points of the total throughput times in that chain. Because it is not possible to retrieve the means and variances of the two processes from the equations of the moment generating function, we need to take observations from the processes. In this section, an illustration is presented, how to determine the means and variances of both processes, with just a small amount of observations from the processes A and B. We assume that there is prior information available for the means of those processes. The mean and variance are estimated with the Bayesian simulation, from section 3.1.2. To update the mean and variance, the iteration steps from section 3.2 are used.

The steps to de-convolute the two normal distributions are given by:

Step 1: Collect the prior beliefs with respect to the means of the processes.
Step 2: Collect an observation of one of the two processes.
Estimate the mean and variance in a **Bayesian** way.
Determine the mean and variance of the remaining process (**moments**).
Step 3: Collect a new observation of one of the two processes.
Update the observed mean and variance with help of the **Theorems**.
Estimate the mean and variance in a **Bayesian** way.
Determine the mean and variance of the remaining process (**moments**).

Step 3 represents the final loop, thus, whenever new observations are available, step 3 should be carried out.

In the next example, an illustration is given of how fast the shape parameters of both processes are determined with respect to the collected observations, if prior information of the mean is available. Suppose we have collected data of the total throughput times, which is normally distributed like $data \sim N(360, 410)$. The processes A and B are distributed like $A \sim N(160, 40^2)$ and $B \sim N(200, 50^2)$. If we have information of the means of two identical processes, say A' and B', we can use their means as an informative prior for the means of both processes. Every experiment contains 25 observations and these experiments are available for both processes.

The estimation of the mean and variance of process A, with help of the three steps, is presented in in figure 3.3. The black line represents outcome of the three steps with the Bayesian estimates and the updates on the mean and variance. The red dots and lines represent the estimation in a classical way and its 95% confidence interval of that estimation, respectively.

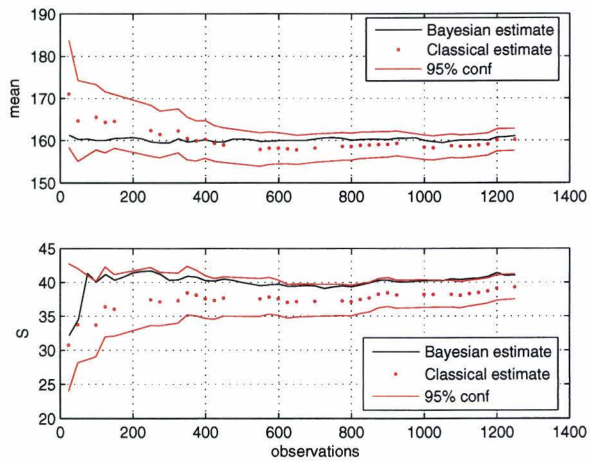

Figure 3.3: Process A: Estimates for the μ and σ

The estimation of the mean and variance of process B with help of the three steps, is presented in in figure 3.4. The black line represents outcome of the three steps with the Bayesian estimates and the updates on the mean and variance. The red dots and lines represent the estimation in a classical way and its 95% confidence interval of that estimation, respectively.

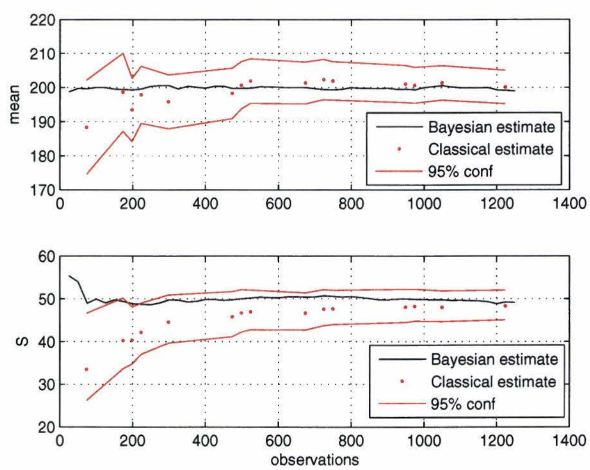

Figure 3.4: Process B: Estimates for the μ and σ

The Bayesian estimates for the mean and variance of the two processes are, especially for a small number of observation, more accurate. The more observations are available, the more the classical and Bayesian estimates resemble each other, as described in section 3.1.2.

The example is done for only two processes. If there are more processes, the approach for the de-convolution is slightly different. Suppose that there are N processes, obviously we need observations of $N-1$ processes to estimate the means and variances of all the processes. Then, when new observations are available from a process, the mean and variance of that process are updated. The mean and variance of the process with the biggest ratio $\frac{\sigma^2}{n}$, are adjusted with the equation from section 2.2.1. The ratio $\frac{\sigma^2}{n}$ comes again from the central limit theorem, as in equation (3.7), which implies the uncertainty of the sampled mean.
Another approach is to wait until we have new observations of $N-1$ processes. The means and variances of each processes are updated with the new observations. Hereafter, the mean and variance of the resulting process are determined with the equations from section 2.2.1. A disadvantage is, however, that for a great amount of processes, we have to wait until we have the availability of the observations of those $N-1$ processes.

3.4 Exponential distributions

In this section the de-convolution of exponential distributions is examined. The involved θ_i's are determined by solving the equations from the moment generating function. For n processes, that poses the number n of unknown parameters, we should retrieve n equations from the moment generating functions. First two processes, that are exponentially distributed, are examined.

3.4.1 Two processes

Consider two processes, with exponentially distributed throughput times, such as presented in figure 3.5.

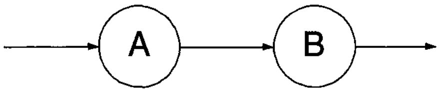

Figure 3.5: Two exponentially distributed processes

We consider that we have a great amount of data points of the total throughput times in that chain. From section 2.3.2, the equations to determine the mean and variance of process A and B, are:

$$\begin{aligned}\mu_{data} &= \theta_A + \theta_B \\ \sigma^2_{data} &= \theta_A^2 + \theta_B^2\end{aligned}$$

For two processes it is not difficult to determine both θ_A and θ_B. For more processes, more equations from the moment generating function are needed to determine the θ_i's. Because of the complexity of these equations and the inaccuracies of the estimation of the moments with respect to the data, it is not always possible to retrieve the θ_i's analytically. The approach in this paper to cope with that problem is to optimize the so called error function. A method to define such an optimization problem, is proposed in the next section.

3.4.2 Multiple processes

For a probability distribution, F_Y, that is composed of n sub-processes X_i, $0 < i \leq n$, n equations, that are derived from the moment generating functions, have to be solved. Those equations, as derived in section 2.3.1, can become very complex as the number of sub-processes increases. In that case, it is not always possible to compute the corresponding θ_i's analytically. This is because, the moments of the distribution F_Y, cannot be determined accurately enough (there are not enough data points). In that case, an optimization problem is defined, that determines the best set θ_i's. From the first and second moment of F_Y, we can determine the relationship of the means and variances with respect to the θ_i's. These two constraints are of great importance and should hold for all estimated θ_i. The equations from the other, $n-2$, moments are transcribed into an optimization objective to minimize the error when solving the equations. Whenever the error is zero, the estimated point is the same as the numerical outcome when solving those equations. When the error does not equal zero, the outcome of the optimization is the best outcome for this problem.

The optimization problem for the exponential de-convolution is:

$$\begin{aligned} \text{Objective:} & \quad \min_{\underline{q}} \underline{\epsilon}^T \cdot \underline{\epsilon} \\ \text{subject to:} & \\ \text{mean}(F_Y) - \sum_{i=1}^{n} \theta_i &= 0 \\ \text{variance}(F_Y) - \sum_{i=1}^{n} \theta_i^2 &= 0 \\ \underline{q} &= \underline{\epsilon} \end{aligned}$$

and

$$q_i = m_i - eq_i \quad 3 \leq i \leq n$$

where m_i stands for the i^{th} moment of F_Y and eq_i stands for the corresponding equation from the moment generating function. The output of this optimization problem is an estimation for θ_i, namely $\hat{\theta}_i$.

Example

To illustrate this optimization problem, we look at a probability distributions, F_Y, that is composed of five, $n = 5$, exponential distributions.
Those five exponential distributions have input parameters:

$$\theta = [4\ 3\ 2\ 1\ 0.5]$$

To solve the optimization problem, the n moments, m_i, are determined from equations (2.4), (2.3) and (2.2). To determine the n estimators, $\hat{\theta}_i$'s, we use the corresponding n equations, eq_i, from the moment generating function of F_Y to solve the optimization problem.
The output of the optimization problem is:

$$\hat{\theta} = [0.5000\ 1.0000\ 2.0000\ 3.0000\ 4.0000]$$

The parameters of $\hat{\theta}$, correspond to parameters of the exponential distributions that are convoluted. As explained in section 2.3.2, only the parameters of those exponential distributions can be determined, not the sequence in which the distributions are convoluted.

The optimization problem as described in this section holds also for the de-convolution of other probability distributions than the exponential. As long as there are enough equations from the moment generating function to solve the unknown parameters of the underlying sub-processes, that optimization problem will determine those unknown parameters. The computation of the unknown parameters of the distributions of the sub-processes, is exemplified in the next chapter. In that chapter, the convoluted distribution, F_Y, is an EPT data set and the sub-processes are the sources of variability of that EPT.

Chapter 4

De-convolution of EPT Data Sets

In this chapter, the use of de-convoluting a probability distribution is demonstrated on EPT data sets. We consider the EPT data set to have a probability distribution F_{EPT} and the n sources of variability to have a distribution F_{X_i}, $0 < i \leq n$. Then, it possible to de-convolute the EPT distribution and to estimate the mean and variance of each of the sources of variability.
In the sections, EPT data sets with changing sources of variability, are examined. In the end of this chapter, the de-convolution is illustrated on an EPT data set, that is composed of all the sources of variability that are mentioned in chapter 1.1.
The probability distributions of the sources of variability, do all descend from the gamma distribution, as explained in chapter 1.1. The moment generating function of the gamma distribution, $Gam(\alpha, \beta)$, is [Joh94]:

$$M(t) = \frac{1}{(1 - \beta t)^{\alpha}} \tag{4.1}$$

When convoluting the moment generating functions of the distributions corresponding to the sources of variability, it is possible to determine dependencies of these sources on the shape of the EPT distribution.

In every section, the approach for the de-convolution of an EPT data set is structured the same. First, we define the underlying sources of variability that are present in the workstation. Those sources have a probability distribution as assumed in chapter 1.1. Hereafter, values are ascribed to the parameters of the distributions of the sources of variability. With these parameters, we can determine the exact mean and variance of each source.
To create an EPT data set, the distributions of the sources of variability are convoluted. Then, we will determine the dependencies of the parameters of the underlying distribution on the shape of the EPT distribution. From the moment generating function of F_{EPT}, we are able to express those dependencies by means of a set of equations. If we

solve those equations as an optimization problem, like in section 3.4.2, we retrieve estimators for the input parameters of the distributions from which the EPT is composed. Those estimators are used to determine the mean and variance of every source in the workstation. Finally, the estimated mean and variance are compared with the mean and variance that have been ascribed to the parameters of the distribution from which the EPT is convoluted.

4.1 Operator availability and setup time

In this section we consider an imaginary workstation that has two sources of variability, namely the operator availability and the setup time (the natural processing time is zero and there is no breakdown). The convolution of those sources results in the EPT data set with probability distribution F_{EPT}. The operator availability is exponentially distributed with parameter θ. The mean of that source of variability is θ and the corresponding variance is θ^2. The setup time has a gamma distribution with parameters α and β. The mean setup time is $\alpha \cdot \beta$ and the corresponding variance is $\alpha \cdot \beta^2$.
From the convolution of the moment generating functions of the exponential and gamma distribution, we determine the dependencies of θ, α and β on the shape of the convoluted EPT distribution, F_{EPT}. Those dependencies are used to estimate the parameters of the sources of variability, namely $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$. With those estimators, the mean and variance of each source can be determined.

4.1.1 EPT

Suppose, for example, that the parameters of the distributions of the sources of variability are $\theta = 2$, $\alpha = 5$ and $\beta = 0.2$. The convolution of those distributions results in an EPT distribution. The corresponding probability density function, PDF, is presented in figure 4.1 (labeled as Exact).

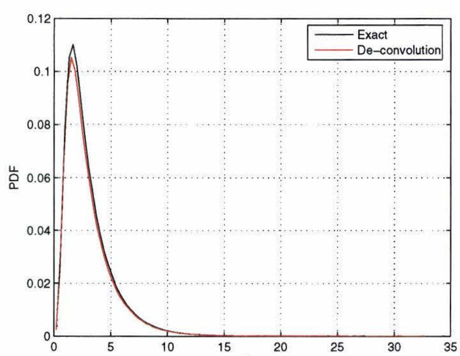

Figure 4.1: EPT

From this probability density function, is is possible to determine the moments of the corresponding EPT distribution.

4.1.2 Moment generating function

The moment generating function corresponding to the EPT distribution, F_{EPT}, is:

$$M_{F_{EPT}}(t) = \frac{1}{(1-\theta t)} \cdot \frac{1}{(1-\beta t)^{\alpha}} \tag{4.2}$$

From this moment generating function, equations for the dependencies of θ, α and β on the shape of F_{EPT}, can be determined.

4.1.3 Estimators for the mean and variance of the sources

When we solve the equations from the moment generating function, estimators are derived for θ, α and β. From these estimators, $\hat{\theta}$, $\hat{\alpha}$ and $\hat{\beta}$, the mean and variance of every source can be determined.

For several number of data points, the estimated mean and variance of the sources of variability, are:

Operator availability:

Data points:	∞	10e6	10e5	10e4	10e3	10e2	10e1
θ	2	2	2	2	2	2	2
$\hat{\theta}$	2.0000	1.9993	1.9932	2.0049	2.0163	1.9309	2.0505
θ^2	4	4	4	4	4	4	4
$\hat{\theta}^2$	4.0000	3.9974	3.9729	4.0197	4.0655	3.7283	4.2044

Setup time:

Data points:	∞	10e6	10e5	10e4	10e3	10e2	10e1
$\alpha \cdot \beta$	1	1	1	1	1	1	1
$\hat{\alpha} \cdot \hat{\beta}$	1.0000	1.0004	1.0066	0.9944	0.9934	1.0269	0.9975
$\alpha \cdot \beta^2$	0.2	0.2	0.2	0.2	0.2	0.2	0.2
$\hat{\alpha} \cdot \hat{\beta}^2$	0.2000	0.2014	0.2137	0.1958	0.2488	0.3525	0.6205

Where, ∞ implies that the exact moments are used to solve the optimization problem. The estimators can be used to recreate the EPT density function. That EPT approximation is also presented in figure 4.1 (labeled as De-convolution).

4.2 Natural processing time and machine breakdown

In this section we consider a workstation that has two sources of variability, namely the natural processing time and breakdown time. The convolution of those sources results in the EPT data set with probability distribution F_{EPT}. The natural processing time is gamma distributed with parameters α_1 and β_1. The mean of that source of variability is $\alpha_1 \cdot \beta_1$ and the corresponding variance is $\alpha_1 \cdot \beta_1^2$. The breakdown time has also a gamma distribution with parameters α_2 and β_2. The mean breakdown time is $\alpha_2 \cdot \beta_2$ and the corresponding variance is $\alpha_2 \cdot \beta_2^2$.
From the convolution of the moment generating functions of the two gamma distributions, we determine the dependencies of α_1, β_1, α_2 and β_2 on the shape of the convoluted EPT distribution, F_{EPT}. Those dependencies are used to estimate the parameters of the sources of variability, namely $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\alpha}_2$ and $\hat{\beta}_2$. With those estimators, the mean and variance of each source can be determined.

4.2.1 EPT

If a breakdown does not occur, the processing time equals the natural processing time. If a breakdown does occur, the processing time equals the breakdown time. That breakdown time is the convolution of the repair time and natural processing time.

With a probability of $(1 - p)$ that a breakdown occurs, the mixed distribution corresponding to the EPT data set is:

$$p \cdot Gam(\alpha_1, \beta_1) + (1 - p) \cdot Gam(\alpha_2, \beta_2) \ ,$$

To create an EPT data set, suppose, for example, that the parameters of the distributions of the sources of variability are $\alpha_1 = 100$, $\beta_1 = 0.05$, $\alpha_2 = 100$ and $\beta_2 = 0.2$. The probability that a breakdown occurs is $(1 - p) = 0.1$ (10%). The convolution of those distributions results in an EPT distribution. The corresponding probability density function, PDF, is presented in figure 4.2 (labeled as Exact).

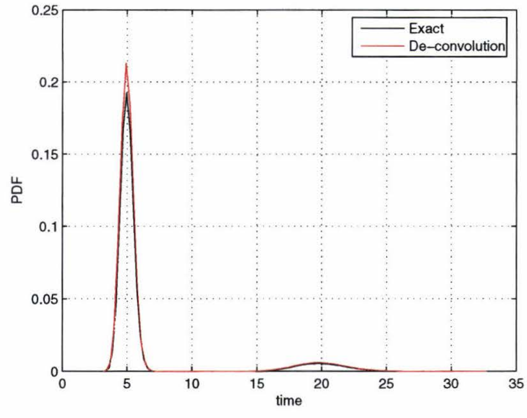

Figure 4.2: EPT

From this probability density function, is is possible to determine the moments of the corresponding EPT distribution.

4.2.2 Moment generating function

The moment generating function corresponding to the EPT distribution, F_{EPT}, is:

$$M_{F_{EPT}}(t) = p \cdot \frac{1}{(1-\beta_1 t)^{\alpha_1}} + (1-p) \cdot \frac{1}{(1-\beta_2 t)^{\alpha_2}} \tag{4.3}$$

From this moment generating function, equations for the dependencies of α_1, β_1, α_2, β_2 and p on the shape of F_{EPT}, can be determined.

4.2.3 Estimators for the mean and variance of the sources

When we solve the equations from the moment generating function, estimators are derived for α_1, β_1, α_2, β_2 and p. From these estimators, $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\alpha}_2$, $\hat{\beta}_2$ and $\hat{p}$, the mean and variance of every source can be determined.

For several number of data points, the estimated mean and variance of the sources of variability, are:

Natural processing time:

Data points:	∞	10e6	10e5	10e4	10e3	10e2	10e1
$\alpha_1 \cdot \beta_1$	5	5	5	5	5	5	5
$\hat{\alpha}_1 \cdot \hat{\beta}_1$	5.0000	4.9996	4.9999	4.9992	5.0033	5.0136	4.9454
$\alpha_1 \cdot \beta_1^2$	0.25	0.25	0.25	0.25	0.25	0.25	0.25
$\hat{\alpha}_1 \cdot \hat{\beta}_1^2$	0.2500	0.2499	0.2503	0.2512	0.2572	0.2572	0.1112

Breakdown time:

Data points:	∞	10e6	10e5	10e4	10e3	10e2	10e1
$\alpha_2 \cdot \beta_2$	20	20	20	20	20	20	20
$\hat{\alpha}_2 \cdot \hat{\beta}_2$	20.000	19.998	19.990	19.984	20.021	20.067	20.628
$\alpha_2 \cdot \beta_2^2$	4	4	4	4	4	4	4
$\hat{\alpha}_2 \cdot \hat{\beta}_2^2$	4.0000	4.0047	4.0181	3.8932	3.5964	3.4085	2.9036

The estimated probability of breakdown is:

Probability of breakdown $(1 - p)$:

Data points:	∞	10e6	10e5	10e4	10e3	10e2	10e1
$(1 - p)$	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$(1 - \hat{p})$	0.1000	0.1000	0.1000	0.0999	0.0999	0.0997	0.1026

The estimators can be used to recreate the EPT density function. That EPT approximation is also presented in figure 4.2 (labeled as De-convolution).

4.3 Four sources of variability

In this section we look at the total number of sources of variability, that were considered to be responsible for the EPT, as is assumed in chapter 1.1. These sources are the operator availability, the setup time, the natural processing time and breakdown. The distributions of those sources are the same as in the previous sections, namely:

- Operator availability: Exponential distribution with parameter θ
- Setup time: Gamma distribution with parameters α_1 and β_1
- Natural processing time: Gamma distribution with parameters α_2 and β_2
- Breakdown time: Gamma distribution with parameters α_3 and β_3

4.3.1 EPT

To create an EPT data set, suppose, for example, that the parameters of the distributions of the sources of variability are:

- Operator availability: $\theta = 2$
- Setup time: $\alpha_1 = 5$ and $\beta_1 = 0.2$
- Natural processing time: $\alpha_2 = 100$ and $\beta_2 = 0.05$
- Breakdown time: $\alpha_3 = 100$ and $\beta_3 = 0.2$

The probability that a breakdown occurs is $(1 - p) = 0.1$ (10%). The convolution of those distributions results in an EPT distribution. The corresponding probability density function, PDF, is presented in figure 4.3 (labeled as Exact).

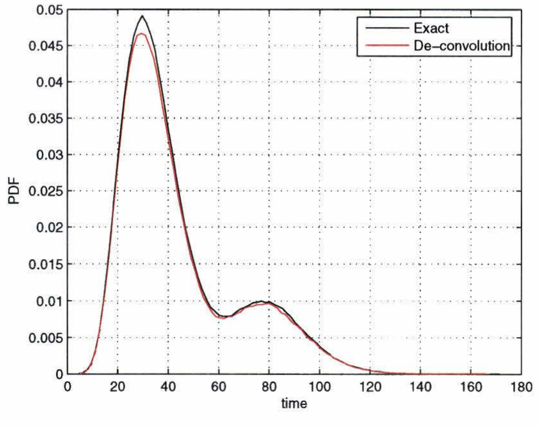

Figure 4.3: EPT

From this probability density function, is is possible to determine the moments of the corresponding EPT distribution.

4.3.2 Moment generating function

The moment generating function corresponding to the EPT distribution, F_{EPT}, is:

$$M_{F_{EPT}}(t) = \frac{1}{(1-\theta t)} \cdot \frac{1}{(1-\beta_1 t)^{\alpha_1}} \cdot \left(p \cdot \frac{1}{(1-\beta_2 t)^{\alpha_2}} + (1-p) \cdot \frac{1}{(1-\beta_3 t)^{\alpha_3}} \right) \quad (4.4)$$

From this moment generating function, equations for the dependencies of θ, α_1, β_1, α_2, β_2, α_3, β_3 and p on the shape of F_{EPT}, can be determined.

4.3.3 Estimators for the mean and variance of the sources

When we solve the equations from the moment generating function, estimators are derived for θ, α_1, β_1, α_2, β_2, α_3, β_3 and p. From these estimators, $\hat{\theta}$, $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\alpha}_2$, $\hat{\beta}_2$, $\hat{\alpha}_3$, $\hat{\beta}_3$ and $\hat{p}$, the mean and variance of every source can be determined.

For several number of data points, the estimated mean and variance of the sources of variability, are:

Operator availability:

Data points:	∞	10e6	10e5	10e4	10e3	10e2	10e1
θ	2	2	2	2	2	2	2
$\hat{\theta}$	2.0000	2.0053	1.9694	2.0158	1.8973	2.0042	1.8456
θ^2	4	4	4	4	4	4	4
$\hat{\theta}^2$	4.0000	4.0213	3.8786	4.0634	3.5996	4.0166	3.4062

Setup time:

Data points:	∞	10e6	10e5	10e4	10e3	10e2	10e1
$\alpha_1 \cdot \beta_1$	1	1	1	1	1	1	1
$\hat{\alpha}_1 \cdot \hat{\beta}_1$	1.0000	0.9746	1.1759	0.8960	1.5745	1.3424	2.2242
$\alpha_1 \cdot \beta_1^2$	0.2	0.2	0.2	0.2	0.2	0.2	0.2
$\hat{\alpha}_1 \cdot \hat{\beta}_1^2$	0.2000	0.1900	0.2765	0.1606	0.4954	0.3560	0.9725

Natural processing time:

Data points:	∞	10e6	10e5	10e4	10e3	10e2	10e1
$\alpha_2 \cdot \beta_2$	5	5	5	5	5	5	5
$\hat{\alpha}_2 \cdot \hat{\beta}_2$	5.0000	5.0216	4.8514	5.0968	4.5263	4.6492	3.9836
$\alpha_2 \cdot \beta_2^2$	0.25	0.25	0.25	0.25	0.25	0.25	0.25
$\hat{\alpha}_2 \cdot \hat{\beta}_2^2$	0.2500	0.2522	0.2354	0.2598	0.2049	0.2161	0.1580

Breakdown time:

Data points:	∞	10e6	10e5	10e4	10e3	10e2	10e1
$\alpha_3 \cdot \beta_3$ 20	20	20	20	20	20	20	20
$\hat{\alpha}_3 \cdot \hat{\beta}_3$	20.000	20.027	19.865	20.125	19.482	19.991	18.216
$\alpha_3 \cdot \beta_3^2$	4	4	4	4	4	4	4
$\hat{\alpha}_3 \cdot \hat{\beta}_3^2$	4.0000	4.0109	3.9464	4.0503	3.7960	3.9970	3.3370

The estimated probability of breakdown is:

Probability of breakdown $(1 - p)$:

Data points:	∞	10e6	10e5	10e4	10e3	10e2	10e1
$(1-p)$	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$(1-\hat{p})$	0.1000	0.0999	0.1002	0.0994	0.0991	0.0950	0.0927

The estimators can be used to recreate the EPT density function. That EPT approximation is also presented in figure 4.3 (labeled as De-convolution).

4.4 Conclusions

With the de-convolution of the EPT data sets, we are able to estimate the mean and variance of every source of variability at a workstation. Especially, when there are just a few sources of variability and we have the availability of a lot of data points, the estimates for the means and variances of those sources are very accurate ($< 5\%$).

For less data points or more complex equations, the estimation of the means and variances are less accurate. To improve the accuracy, research can be done to find out which equations from the moment generating function of the distribution F_{EPT}, are less complex to solve. These equations can be used to solve the optimization problem. Also, research can be done to find out which shape parameters, or moments, of the EPT distribution are less sensitive to estimate, when a smaller amount of data points is available. The corresponding equations from the moment generating function can also be used to solve the optimization problem.

Chapter 5

Conclusions and Recommendations

In this report the de-convolution of a probability distribution is described. Such a distribution is composed of several underlying probability distributions. In this context, de-convolution implies that the underlying distributions are determined when the convoluted distribution is assumed to be known. Also the forms and the number of the underlying distributions are supposed to be known.

From the moment generating functions from chapter 2, we are able to derive equations to determine the dependencies of the parameters of the distributions of the sub-processes, F_{X_i}, on the moments of the convoluted distribution, F_Y. With these equations, we can identify the forms of distributions for F_X, that can be de-convoluted, like the exponential distribution, or that cannot be de-convoluted, like the normal distribution.

If the sub-processes are exponentially distributed, the parameters of those distributions can be determined with an optimization problem as described in chapter 3. This optimization problem determines estimators for the unknown parameters from the equations from the moment generating function, of F_Y.
In that same chapter, two methods are proposed for the de-convolution with respect to normally distributed sub-processes. These methods, respectively, are based on classical statistical beliefs and on Bayesian statistical beliefs for an estimation of the unknown parameters of the normal distributions. The Bayesian method is favorable when prior information is available of the distributions of the sub-processes.
With the theorems for the update of the mean and variance, all observations from all the sub-processes can be taken into account, for an accurate estimate of the unknown parameters of the normal distributions.

In the last chapter, the de-convolution of EPT data sets is demonstrated. Effective process time, EPT, is an aggregate measure into which all sources of variability at a

workstation are combined.
With the de-convolution of the EPT data set, estimators are determined for the parameters of the distributions of the sources of variability. From these estimators, we are able to determine the mean and variance of each source of variability.

If we have the availability of lot of data points (the moments of the EPT distribution do correspond more to the theoretical moments of that EPT distribution), the estimates for the mean and variance of the sources of variability are accurate ($< 5\%$). Also, for a small number of sources of variability (the equations from the moment generating function are less complex), the estimation of both parameters is also accurate ($< 5\%$). For less data points or more complex equations, the estimation of the means and variances are less accurate. To improve the accuracy, research can be done to find out which equations from the moment generating function are less complex to solve. Also, research can be done to find out which shape parameters, or moments, of the EPT distribution are less sensitive to estimate, when a smaller amount of data points is available.

Also, research into other applications for the de-convolution of a probability distributions can be done. One might think of, for example, destructive testing, which is often necessary in the bulb or car industry to test the products.

Bibliography

[Gel04] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis.* Chapmann & Hall/CRC, second edition, 2004.

[Hop01] W.J. Hopp and M.L. Spearman. *Factory Physics: Foundations of Manufacturing Management.* Irwin McGraw-Hill, second edition, 2001.

[Jac03] J.H. Jacobs, L.F.P. Etman, E.J.J. van Campen, and J.E. Rooda. Characterization of operational time variability using effective process times. *IEEE Trans. Semiconduct. Manufact.*, 16(3):511–520, 2003.

[Joh94] N.L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Volume 1.* John Wiley & Sons, Inc., second edition, 1994.

[Koc05] A.A.A. Kock, F.J.J. F.J.J.Wullems, L.F.P. Etman, I.J.B.F. Adan, and J.E. Rooda. Performance evaluation and lumped parameter modelling of single server flowline subject to blocking: An effective process time approach. *Fifth International Conference*, pages 137–144, 2005.

[Law00] A.M. Law and W.D. Kelton. *Simulation Modeling and Analysis.* Irwin McGraw-Hill, third edition, 2000.

[Lee97] P.M. Lee. *Bayesian Statistics.* ARNOLD, second edition, 1997.

[Lef04] A.A.J. Lefeber, R.A. van den Berg, and J.E. Rooda. Modeling, validation and control of manufacturing systems. *Proceedings of the 2004 American Control Conference; Boston, MA, United States*, 5:4583–4588, 2004.

[Mon99] D.C. Montgomery and G.C. Runger. *Applied Statistics and Probability for Engineers.* John Wiley & Sons, Inc., 1999.

Appendix A

Theorems

In this appendix, the theorems from sections 3.2 are derived.

Theorem 1: Update the mean

$$\bar{y}_m = \bar{y}_l + \frac{k(\bar{y}_k - \bar{y}_l)}{l + k} \tag{A.1}$$

Proof:
Considering equation (3.1),

$$\begin{aligned}
\bar{y}_m &= \frac{l \cdot \bar{y}_l + k \cdot \bar{y}_k}{l + k} \\
&= \frac{(l + k)\bar{y}_l - k(\bar{y}_k - \bar{y}_l)}{(l + k)} \\
&= \bar{y}_l + \frac{k(\bar{y}_k - \bar{y}_l)}{(l + k)}
\end{aligned}$$

Theorem 2: Update the variance

$$s_m^2 = \frac{1}{l+k-1}\left\{(l-1)s_l^2 + (k-1)s_k^2 + \left(\frac{lk}{l+k}\right)(\bar{y}_k - \bar{y}_l)^2\right\}$$

Proof:
Considering equation (3.2) and assume $s_1^2 = 0$,

$$\begin{aligned}
s_m^2 &= \frac{1}{l+k-1}\sum_{j=1}^{l+k}[y_j - \bar{y}_m]^2 \\
(l+k-1)s_m^2 &= \sum_{j=1}^{l+k}[(y_j - \bar{y}_l) + (\bar{y}_l - \bar{y}_m)]^2 \\
&= \sum_{j=1}^{l+k}(y_j - \bar{y}_l)^2 + \\
&\quad 2\sum_{j=1}^{l+k}[(y_j - \bar{y}_l)(\bar{y}_l - \bar{y}_m)] + \\
&\quad \sum_{j=1}^{l+k}(\bar{y}_l - \bar{y}_m)^2
\end{aligned}$$

The three terms are derived separately. Considering the first term,

$$\begin{aligned}
\sum_{j=1}^{l+k}(y_j - \bar{y}_l)^2 &= \sum_{j=1}^{l}(y_j - \bar{y}_l)^2 + \sum_{j=l+1}^{l+k}(y_j - \bar{y}_l)^2 \\
&= (l-1)s_l^2 + \sum_{j=l+1}^{l+k}((y_j - \bar{y}_k) + (\bar{y}_k - \bar{y}_l))^2 \\
&= (l-1)s_l^2 + \sum_{j=l+1}^{l+k}(y_j - \bar{y}_k)^2 + \\
&\quad 2(\bar{y}_k - \bar{y}_l)\sum_{j=l+1}^{l+k}(y_j - \bar{y}_k) + \sum_{j=l+1}^{l+k}(\bar{y}_k - \bar{y}_l)^2 \\
&= (l-1)s_l^2 + (k-1)s_k^2 + 0 + k(\bar{y}_k - \bar{y}_l)^2
\end{aligned}$$

Considering the second term,

$$\begin{aligned}
2\sum_{j=1}^{l+k}[(y_j-\bar{y}_l)(\bar{y}_l-\bar{y}_m)] &= 2\cdot(\bar{y}_l-\bar{y}_m)\sum_{j=1}^{l+k}(y_j-\bar{y}_l) \\
&= 2\cdot(\bar{y}_l-\bar{y}_m)\left(\sum_{j=1}^{l+k}y_j-(l+k)\bar{y}_l\right) \\
&= 2\cdot(\bar{y}_l-\bar{y}_m)\big((l+k)\bar{y}_m-(l+k)\bar{y}_l\big) \\
&= -2\cdot(l+k)(\bar{y}_l-\bar{y}_m)^2 \\
&= -2\cdot\frac{k^2}{l+k}(\bar{y}_k-\bar{y}_l)^2 \qquad \text{(with } (A.1))
\end{aligned}$$

Considering the third term,

$$\begin{aligned}
\sum_{j=1}^{l+k}(\bar{y}_l-\bar{y}_m)^2 &= (l+k)(\bar{y}_l-\bar{y}_m)^2 \\
&= \frac{k^2}{l+k}(\bar{y}_k-\bar{y}_l)^2 \qquad \text{(with } (A.1))
\end{aligned}$$

(with (A.1)) Combining these three terms,

$$\begin{aligned}
(l+k-1)s_m^2 &= [(l-1)s_l^2+(k-1)s_k^2+k(\bar{y}_k-\bar{y}_l)^2]+ \\
&\quad [-2\cdot\frac{k^2}{l+k}(\bar{y}_k-\bar{y}_l)^2]+[\frac{k^2}{l+k}(\bar{y}_k-\bar{y}_l)^2] \\
&= (l-1)s_l^2+(k-1)s_k^2+\left(k-\frac{k^2}{l+k}\right)(\bar{y}_k-\bar{y}_l)^2 \\
&= (l-1)s_l^2+(k-1)s_k^2+\left(\frac{lk}{l+k}\right)(\bar{y}_k-\bar{y}_l)^2
\end{aligned}$$

Which yields,

$$s_m^2 = \frac{1}{l+k-1}\left\{(l-1)s_l^2+(k-1)s_k^2+\left(\frac{lk}{l+k}\right)(\bar{y}_k-\bar{y}_l)^2\right\}$$