

7-12-2004

Bayesian and Empirical Bayes Approaches to Power Law Process and Microarray Analysis

Zhao Chen
University of South Florida

Follow this and additional works at: <https://digitalcommons.usf.edu/etd>



Part of the [American Studies Commons](#)

Scholar Commons Citation

Chen, Zhao, "Bayesian and Empirical Bayes Approaches to Power Law Process and Microarray Analysis" (2004). *USF Tampa Graduate Theses and Dissertations*.
<https://digitalcommons.usf.edu/etd/992>

This Dissertation is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

Bayesian and Empirical Bayes Approaches to Power Law Process and Microarray
Analysis

by

Zhao Chen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Mathematics
College of Arts and Sciences
University of South Florida

Major Professor: A.N.V. Rao, Ph.D.
Committee member: Marcus McWaters, Ph.D.
Committee member: Kandethody Ramachandran, Ph.D.
Committee member: Lihua Li, Ph.D.
Committee member: George Yanev, Ph.D.

Date of Approval:
July 12th, 2004

Keywords: Bayesian, Empirical Bayes, Power Law Process, Nonhomogeneous Poisson
Process, Microarray, EBarrays

©Copyright 2004, Zhao Chen

DEDICATION

To my parents,
Guoren Chen and Wandu Mo,
who made all of this possible,
for their endless encouragement and patience.

•

And also to
my fiancée,
Peng Feng,
who never fails to remind me how to live as if every day were my most precious.

ACKNOWLEDGMENTS

There is one person who deserves my deepest thanks and respect for his continued support during the writing of this dissertation: my advisor, Dr. A.N.V. Rao. His technical and editorial advice was essential to the completion of this dissertation and he has taught me innumerable lessons and insights on the workings of academic research in general. I also greatly appreciate his advice in many other aspects of academic life.

My thanks also go to the members of my major committee. Dr. Chris Tsokos showed me what's most important as a statistician and how to efficiently present a work. I can't finish my graduate study without Dr. Marcus McWaters continued support. Thanks to Dr. Kandethody Ramachandran and Dr. George Yanev for reading previous drafts of this dissertation and providing many valuable comments that improved the presentation and contents of this dissertation. I'd like to give my special appreciation to Dr. Lihua Li. He offered me a great opportunity to gain some experience in applying statistics methods. He also gave me many precious advices on how to work as a professional. Thank Dr. Jiang Kun for chairing my dissertation defense.

Many thanks to Dr. Eric Haura. He is the one who motivates me to work on microarray analysis and has always been supportive on the work. I had very pleasant time to work with Daniel J. Daucer and Lanxi Song in the laboratory. A good friend of mine, Heng Li, contributed a great number of help and no exact words can describe my deep appreciation to him.

I really appreciate the mathematics department of the University of South Florida for offering me the opportunity to study here. During the past 5 years, not only did I gain

knowledge every day, but also the friendship from the people in the department. Thank you for your smiles, Ayako, Beverly, Nancy, Frances and Marianne. You made me feel like a family.

Bayesian and Empirical Bayes Approaches to Power Law Process and Microarray

Analysis

Zhao Chen

ABSTRACT

In this thesis, we apply Bayes and Empirical Bayes methods for reliability growth models based on the power law process. We also apply Bayes methods for the study of microarrays, in particular, in the selection of differentially expressed genes.

The power law process has been used extensively in reliability growth models. Chapter 1 reviews some basic concepts in reliability growth models. Chapter 2 shows classical inferences on the power law process. We also assess the goodness of fit of a power law process for a reliability growth model. In chapter 3 we develop Bayesian procedures for the power law process with failure truncated data, using non-informative priors for the scale and location parameters. In addition to obtaining the posterior density of parameters of the power law process, prediction inferences for the expected number of failures in some time interval and the probability of future failure times are also discussed. The prediction results for the software reliability model are illustrated. We compare our result with the result of Bar-Lev, S.K. *et al.* ([7]). Also, posterior densities of several parametric functions are given. Chapter 4 provides Empirical Bayes for the power law process with natural conjugate priors and nonparametric priors. For the natural conjugate priors, two-hyperparameter prior and a more generalized three-hyperparameter prior are used.

In chapter 5, we review some basic statistical procedures that are involved in microarray analysis. We will also present and compare several transformation and normal-

ization methods for probe level data. The objective of chapter 6 is to select differentially expressed genes from tens of thousands of genes. Both classical methods (fold change, T-test, Wilcoxon Rank-sum Test, SAM and local Z-score (Chen,Z.[17])) and Empirical Bayes methods (EBarrays and LIMMA) are applied to obtain the results. Outputs of a typical classical method *SAM* and a typical Empirical Bayes Method *EBarrays* are discussed in detail.

Contents

Abstract	v
List of Tables	xi
List of Figures	xii
Chapter 1 Reliability Growth and Growth Models	1
1.1 Fundamentals of Reliability	1
1.2 Counting Process	2
1.3 Homogeneous Poisson Process (HPP)	5
1.4 Nonhomogeneous Poisson Process	6
1.5 Power Law Process	8
1.5.1 Historical Review	9
1.5.2 Model Motivation	10
1.5.3 Present Study	12
1.6 Summary	14
Chapter 2 Classical Inference on the Power Law Process	15
2.1 Introduction	15
2.2 Point Estimation of Parameters β and α	16
2.2.1 Maximum Likelihood Estimates $\hat{\beta}$ and $\hat{\alpha}$	17
2.2.2 Unbiased Estimates $\bar{\beta}$	19

2.2.3	Linearly Best Efficient Estimate of $\hat{\beta}$	19
2.3	Interval Estimation and Tests of Hypothesis	20
2.4	Estimation of Intensity Function	21
2.5	Mean Time Between Failure (MTBF)	22
2.5.1	$MTBF_n$	22
2.5.2	$MTBF_Q$	23
2.6	Goodness-of-fit Tests	25
2.6.1	Duane Plot	26
2.6.2	Ratio Power Test	27
2.6.3	Log-ratio Test	28
2.7	Summary	28
 Chapter 3 Bayesian Inference on the Power Law Process		30
3.1	Likelihood Function	31
3.1.1	Likelihood of y_1	31
3.1.2	Likelihood of $\vec{y} = (y_1, y_2, y_3, \dots, y_k)$	31
3.2	Posterior Density of (μ, θ)	33
3.3	Posterior Density of θ and μ	34
3.3.1	Marginal Posterior Density of θ	34
3.3.2	Marginal Posterior Density of μ	36
3.4	Predictive Inference	38
3.4.1	Predictions Based on Posterior Density	38
3.4.2	Prediction Inference Based on Estimation of Intensity Function	40
3.4.3	Data Comparison	42
3.5	Posterior Density for Some Parametric Functions	42
3.5.1	Posterior Density for System Reliability	44

3.5.2	Posterior Density For the Expected Number of Failures in Some Time Interval	45
3.5.3	Posterior Density for Intensity Function	46
3.6	Summary	46
Chapter 4 Empirical Bayes Analysis on the Power Law Process		48
4.1	Parametric Empirical Bayes on the PLP	48
4.1.1	Two Hyperparameters (a, m)	49
4.1.2	Three Hyperparameters (a, m, y_m)	53
4.1.3	Prior Plots	54
4.2	Nonparametric Prior on the Power Law Process	55
4.3	Further Research	61
4.4	Summary	61
Chapter 5 Microarray Analysis: Normalization and Transformation of Probe-Level Data		63
5.1	Overview of Microarray Analysis	63
5.2	Data Description	64
5.3	Probe-Level Analysis of Oligonucleotide Arrays	66
5.3.1	MAS 5.0 – Microarray Analysis Suite Version 5.0	67
5.3.2	MBEI – Model-Based Expression Index (dChip 2001)	68
5.3.3	RMA – Robust Multichip Analysis	69
5.3.4	VSN – Variance Stabilization of Network	70
5.3.5	Comparison	70
5.4	Summary	71
Chapter 6 Statistical Methods of Selecting Differentially Expressed Genes		72
6.1	Select Differentially Regulated Genes Using Classical Statistics Methods . .	72

6.1.1	Fold Change Method	72
6.1.2	T-test	74
6.1.3	Wilcoxon Rank-Sum Test	76
6.1.4	SAM	77
6.1.5	Local Z-score Method	81
6.2	Select Differentially Expressed Genes by Bayesian and Empirical Bayes Methods	82
6.2.1	EBarrays	83
6.2.2	LIMMA	86
6.2.3	Cyber-T	87
6.3	Other Work in Microarray	88
6.4	Summary	89
	Bibliography	90

List of Tables

TABLE 2.1	Software Failure Times in Seconds	26
TABLE 4.1	Commonly Used Kernel Functions	60
TABLE 6.1	SAM Result Table	80
TABLE 6.2	EBarrays Result Table	84
TABLE 6.3	LIMMA Result Table	87

List of Figures

FIGURE 1.1	The Sample Path for a Counting Process	3
FIGURE 1.2	Three Different Types of Systems	11
FIGURE 2.1	Duane Plot for Data Table 2.1	27
FIGURE 2.2	Plot of Number of Failures against Cumulative Operating Time for Data Table 2.1	28
FIGURE 3.1	Marginal Density of θ	35
FIGURE 3.2	The Predictive Probability Function of $N(67, 344; 80, 000)$	43
FIGURE 3.3	The Probability of $(n + 1)$ th Failure Time	43
FIGURE 4.1	Bayesian Model for the PLP	49
FIGURE 4.2	Parametric Empirical Bayes Model for the PLP	50
FIGURE 4.3	Prior Contour Plot $a = 2$ $m = 2$ $y_m = 2$	56
FIGURE 4.4	Prior Three Dimension Plot $a = 2$ $m = 2$ $y_m = 2$	56
FIGURE 4.5	Prior Contour Plot $a = 2$ $m = 6$ $y_m = 2$	57
FIGURE 4.6	Prior Three Dimension Plot $a = 2$ $m = 6$ $y_m = 2$	57
FIGURE 4.7	Prior Contour Plot $a = 2$ $m = 2$ $y_m = 6$	58
FIGURE 4.8	Prior Three Dimension Plot $a = 2$ $m = 2$ $y_m = 6$	58
FIGURE 4.9	Prior Contour Plot $a = 6$ $m = 2$ $y_m = 2$	59
FIGURE 4.10	Prior Three Dimension Plot $a = 6$ $m = 2$ $y_m = 2$	59
FIGURE 5.1	Data Analytic Tasks in Microarray Experimentation	65
FIGURE 5.2	Signal Extraction in Probe Level Data	66
FIGURE 5.3	STAT3 v.s. GFP Mean Value $STAT3 = 0.98745 \cdot GFP$	67

FIGURE 6.1	Lowess Regression for Five Experiments	73
FIGURE 6.2	Histogram of T-statistics	75
FIGURE 6.3	Histogram of Log-expression Wilcoxon Statistics	76
FIGURE 6.4	SAM Plot	79
FIGURE 6.5	Boxplot of Gene Expression for Five Experiments and Five Controls	81
FIGURE 6.6	EBarrays Fit GG Model - From Li Wong Normalized Data	85

Chapter 1

Reliability Growth and Growth Models

Repairable systems are often modeled after a class of stochastic point processes. (Ascher(84)[2], Engelhardt(86)[26], Rigdon and Basu(1990)[49]). This class usually consists of homogeneous Poisson processes (HPP), nonhomogeneous Poisson processes (NHPP), the branching Poisson processes (BPP) and the renewal processes (RP). In this chapter we present some fundamental results on homogenous Poisson processes and nonhomogeneous Poisson processes. A particular nonhomogeneous Poisson process with Power Law intensity function can be employed as a reliability growth model and will play a major role in our research. The results that are covered in this chapter are basic for the research presented in later chapters.

1.1 Fundamentals of Reliability

The **reliability function** is the probability that a system will carry out its mission through time t and is denoted by $R(t)$. Let T denote the failure time since the initial start up of the system (assumed to be at time $t = 0$). Let $N(t)$ denote the cumulative number of failures from time 0 to time t . Then the reliability function is defined as

$$R(t) = Pr[T > t] = 1 - F(t) = \int_t^{\infty} f(s) ds,$$

where $F(t)$ is the cumulative distribution function of T and $f(t)$ is the probability density function of T . We note that $R(t) = Pr[N(t) = 0]$. The reliability function is also called the survival function of T . $R(t)$ decreases in t , from 1 at $t = 0$, to 0 at $t = \infty$.

Reliability plays a key role in developing products and in enhancing competitiveness. There is a lot of literature on the reliability of nonrepairable system. However, we will focus on repairable system.

A **repairable system** is a system which, after failing to perform one or more of its designed functions satisfactorily, can be restored to a fully satisfactory performance by any method, other than replacement of the entire system (Ascher(84)[2]). A good portion of literature on repairable systems seems to be motivated from applications to mechanical systems. However, repairable systems are not limited to such cases. An important application area is the reliability of software systems.

Consider the system is tested until it fails. Since it is repairable, we will debug the software and it runs again. Then we will continue to test the system until it fails again. Before the systems are put into the market, we need reach a desirable reliability, which will reflect the quality of the final design. This process of testing a system has been referred as **reliability growth**. In what follows, we will discuss how to model this process in a suitable way.

1.2 Counting Process

Models utilizing a counting process have played a key role in the analysis of systems composed of random occurring events. By way of motivation, suppose that we are interested in observing the occurrences of a repeatable event over a period of time. One of the simplest examples is the arrival of customers at a service station, such as a bank. Another example is the occurrences of earthquakes of a specified magnitude at a particular location over time. The example that is of interest to us here is the points in time

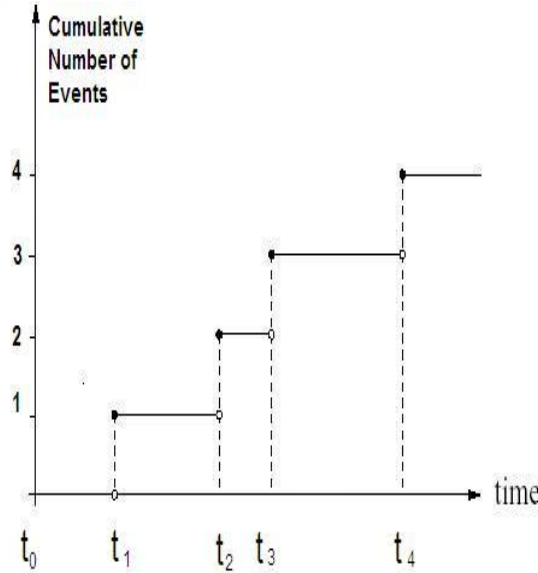


Figure 1.1: The Sample Path of a Counting Process

when a system's software fails. In all such cases, the event of interest does not occur with any regularity and is therefore not exactly predictable. We are not sure about the exact times at which the events will occur and consequently about the exact number of events that will occur in any time interval. Such a random phenomenon is called a point process.

A counting process is simply the count of the number of events that have occurred in any specified interval of time. Since $N(t)$ is unknown for any value of t , we are facing with the problem of describing our uncertainty about an infinite collection of random variables, one for each t . Any indexed collection of random variables is called stochastic process, and when the interest is focused on counts, the process is called a counting process and is denoted by $\{N(t), t \geq 0\}$.

The sample path of a counting process is given by Figure 1.1. The horizontal line is designated to represent time; the vertical line is used to represent the total number of counts over time. It is a step function starting at zero, and taking jumps of size one at each t_i , that is, the cumulative time of the i th failure.

Our purpose of this chapter is to introduce some probabilistic models for the

counting process. The most commonly used models are homogeneous and nonhomogeneous Poisson processes. We now define a Poisson process and the intensity function.

A counting process $N(t)$ is said to be a **Poisson process** if

1. $N(0) = 0$;
2. For any $a < b \leq c < d$ the random variables $N(a, b]$ and $N(c, d]$ are independent. This is called the **independent increment** property.
3. There is a function ν such that

$$\nu(t) \equiv \lim_{\Delta t \rightarrow 0} \frac{Pr\{N(t + \Delta t) - N(t) \geq 1\}}{\Delta t}.$$

The function ν is called the **intensity function** of the Poisson process.

4.

$$\lim_{\Delta t \rightarrow 0} \frac{Pr\{N(t + \Delta t) - N(t) \geq 2\}}{\Delta t} = 0$$

This precludes the possibility of simultaneous failures.

Properties (1) to (4) of the Poisson process imply that

$$P[N(t) = n] = \frac{1}{n!} \left(\int_0^t \nu(x) dx \right)^n \exp \left(- \int_0^t \nu(x) dx \right).$$

For proof, see for example Rigdon(2000)[50].

The terminology of intensity function $\nu(t)$ is often confusing with the terminology of hazard function. **Hazard function** is defined as:

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x < X \leq x + \Delta x | X > x)}{\Delta x}.$$

The hazard function is the limit of the probability that a unit fails (for the first and only time) in a small interval given that it survived to the beginning of the interval. Hazard function is a conditional probability and gives its relative rate at time t . It also can be calculated by dividing the derivative of cumulative distribution function $F(t)$ with the probability of surviving past time t , that is,

$$h(t) = \frac{f(t)}{1 - F(t)}.$$

Note that $\nu(t)\Delta t$ gives the probability of a failure in a small time interval $(t, t + \Delta t]$. In a counting process, the expected number of failures up to time t is denoted by $m(t) = E[N(t)]$. Intensity function can be obtained by taking the derivative of $m(t)$. It is an absolute rate.

1.3 Homogeneous Poisson Process (HPP)

The counting process $\{N(t), t \geq 0\}$ is said to be a **homogeneous Poisson process (HPP)** if the intensity function $\nu(t)$ is a constant, that is, $\nu(t) = \lambda$, $\lambda > 0$ and

1. $N(0) = 0$;
2. The process has independent increments and stationary increments. A point process has **stationary increments** if for all k , $P(N(t, t + s] = k)$ is independent of t .

It can be shown that the number of events in any interval of length $s = t_2 - t_1$ has a Poisson distribution with mean λs , that is

$$P[N(t_2) - N(t_1) = n] = \frac{e^{-\lambda s} (\lambda s)^n}{n!}, \quad 0 \leq t_1 \leq t_2, \quad n = 0, 1, \dots$$

The intensity function is also referred as repair rate. Homogeneous Poisson Process has the following properties, proofs are given in (Rigdon(2000)[50]):

Property 1. A process is an HPP with constant intensity function λ , if and only if the times between events are iid exponential random variables with mean $1/\lambda$.

Property 2. If $0 < T_1 < T_2 < \dots < T_n$ are the failure times from an HPP, then the joint pdf of T_1, T_2, \dots, T_n is

$$f(t_1, t_2, \dots, t_n) = \lambda^n \exp(-\lambda t_n), \quad 0 < t_1 < t_2 < \dots < t_n.$$

Property 3. The time to the n th failure from a system modeled by an HPP has a gamma distribution with parameter $\alpha = n$, $\beta = 1/\lambda$.

Property 4. For an HPP, conditional on $N(t) = n$, the failure times $0 < T_1 <$

$T_2 < \dots < T_n$ are distributed as order statistics from $UNIF(0, t)$ distribution.

Property 5. The probability of system failure after time t is

$$R(t) = Pr[T > t] = Pr[N(t) = 0] = e^{-\lambda t}.$$

The times between events mentioned in Property 1 is called the sequence of interarrival times which is denoted by $\{X_i = T_i - T_{i-1}, i = 1, 2, \dots\}$. We shall note that, in HPP, each X_i is independently identically exponentially distributed with mean $1/\lambda$. Hence, we can expect an average $1/\lambda$ events to occur within the time interval $(t_{n-1}, t_n]$.

$$\begin{aligned} Pr[X_1 > x] &= Pr[N(x) = 0] = e^{-\lambda x}, \\ Pr[X_2 > x | X_1 = x_1] &= Pr\{\text{zero event in } (x_1, x_1 + x)\} = e^{-\lambda x}, \\ &\dots = \dots \\ Pr[X_n > x | X_n = x_{n-1}] &= Pr\{\text{zero event in } (x_{n-1}, x_{n-1} + x)\} = e^{-\lambda x}. \end{aligned}$$

1.4 Nonhomogeneous Poisson Process

Nonhomogeneous Poisson process (NHPP) is a Poisson process which intensity function is not a constant. A counting process $\{N(t), t \geq 0\}$ has a nonhomogeneous Poisson process if

1. $N(0) = 0$;
2. The process has independent increments.

It can be shown that the number of failures in any interval $(t_1, t_2]$ has a Poisson distribution with mean $\int_{t_1}^{t_2} v(t) dt$. That is,

$$P(N(t_2) - N(t_1) = k) = \frac{1}{k!} \exp\left\{-\int_{t_1}^{t_2} v(t) dt\right\} \left(\int_{t_1}^{t_2} v(t) dt\right)^k.$$

For our purposes, these *occurrences in time* will be the failure times of a repairable system. Though the models discussed in the following may be applicable to other situations, we shall use the term *failures* instead of *events* from now on.

There are two different sampling protocols which provide data on repairable system: (i) failure truncated case and (ii) time truncated case.

Data are said to be **failure truncated** when testing stops after a predetermined number of failures. Suppose that a repairable system is observed till n failures occur (fixed n), so we observe the ordered failure times $t_1 < t_2 < \dots < t_n$ where t_i is the time of i th failure. In this case, the number of failures is fixed and the time when the testing stops is random.

Data are said to be **time truncated** when testing stops at a predetermined time t . We observe a set of failure time $t_1 < t_2 < \dots < t_n < t$. In this case, the time when the testing stops is fixed and the number of failures n is random. These two different cases cause slightly different likelihood functions.

Nonhomogeneous Poisson Process has the following properties (Rigdon(2000)[50]):

Property 1. The joint pdf of the failure times $T_1, T_2 \dots T_n$ from an NHPP with intensity function $\nu(t)$ is given by [1]:

$$f(t_1, t_2, \dots, t_n) = \left(\prod_{i=1}^n \nu(t_i) \right) \exp\left(- \int_0^w \nu(x) dx\right),$$

where w is the so-called stopping time: $w = t_n$ for the failure truncated case, $w = t$ for the time truncated case.

Property 2. If the failure times of a nonhomogeneous Poisson process are $T_1 < T_2 < \dots < T_n$ then conditioned on $T_n = t_n$, the random variables $T_1 < T_2 < \dots < T_{n-1}$ are distributed as $n - 1$ order statistics from the distribution with cumulative distribution function

$$G(y) = \begin{cases} 0, & y \leq 0, \\ m(y)/m(t_n), & 0 < y \leq t_n, \\ 1, & y > t_n. \end{cases}$$

Property 3. If a NHPP with intensity function $\nu(t)$ is observed until time t , and if the failure times are $T_1 < T_2 < \dots < T_{N(t)}$ where $N(t)$ is the random number of failures in

the interval $(0, t]$, then conditioned on $N(t) = n$, the random variables $T_1 < T_2 < \dots < T_n$ are distributed as n order statistics from the distribution with cdf

$$G(y) = \begin{cases} 0, & y \leq 0, \\ m(y)/m(t), & 0 < y \leq t, \\ 1, & y > t. \end{cases}$$

Property 4. The probability of system failure occurring after time t is known as the reliability function, $R(t)$. The nonhomogeneous Poisson process assumes that the number of failures in any interval $(t_1, t_2]$ has a Poisson distribution with mean $\int_{t_1}^{t_2} \nu(t) dt$. Hence the reliability function is

$$\begin{aligned} R(t) &= Pr[T > t] \\ &= Pr[N(t) = 0] \\ &= \frac{e^{-\int_{t_1}^{t_2} \nu(t) dt} (\int_{t_1}^{t_2} \nu(t) dt)^0}{0!} \\ &= e^{-\int_{t_1}^{t_2} \nu(t) dt} \\ &= e^{-[\lambda(t_2) - \lambda(t_1)]}. \end{aligned}$$

1.5 Power Law Process

A common function form for the intensity function in NHPP is

$$\nu(t) = \left(\frac{\beta}{\alpha}\right) \left(\frac{t}{\alpha}\right)^{\beta-1} \text{ for } \alpha > 0, \beta > 0,$$

where α and β are the scale parameter and shape parameter respectively. The intensity function is proportional to the cumulative failure time t raised to a power, therefore this special nonhomogeneous Poisson process is usually called the **Power Law Process**. The mean value function $\lambda(t)$ of the process is

$$\lambda(t) = E(N(t)) = \int_0^t \nu(s) ds = \int_0^t \frac{\beta}{\alpha^\beta} s^{\beta-1} ds = (t/\alpha)^\beta.$$

An alternative way of describing the power law process is to consider the sequence of successive failure times T_1, T_2, \dots where T_i is the time of the i th failure. Then the time of the first failure T_1 has a Weibull distribution with scale and shape parameter α and β . T_i ($i = 2, 3, \dots, n$) have left truncated Weibull distributions conditional on $T_1 = t_1, \dots, T_{i-1} = t_{i-1}$. Therefore, Power Law Process is also called Weibull Process.

1.5.1 Historical Review

The power law process has been widely used in reliability growth (Crow(1982)[22]), and software reliability models (Kyparisis and Singpurwalla(1985) [35]), and in repairable systems (Ascher and Feingold (1984)[2], Engelhardt and Bain(1986) [26], Rigdon and Basu(1989) [49]). Other names for the Power Law model are: the Duane Model (Duane(1964) [24]) and AMSAA model. AMSAA stands for the United States Army Materials System Analysis Activity.

There is a lot of literature on the power law process from a classical statistics view. Much theoretical work describing the Power Law model was performed in the 1970's (Lee, L and Lee, K.(1978)[36] and Engelhardt and Bain(1978)[26] [3]). Classical inference on the power law process, such as point estimation, confidence intervals, tests of hypothesis for parameters and estimates of the intensity function, was reviewed by Rigdon and Basu(1989)[49]. Calabria(1988)[12] examined modified maximum likelihood estimators of the expected number of failures in a given time interval and of the failure intensity and compare their mean squared errors with those MLEs. Qiao, H. and Tsokos, C.(1998) [44] obtained the best efficient estimates of intensity function.

Bayesian inference on the power law process was also studied during the past two decades. Bayesian point and interval estimates were obtained by Guida,M.(1989)[30] and Kyparisis and Singpurwalla(1985)[35]. Informative and noninformative priors were both employed on failure truncated data case. Bar-Lev *et al.*(1991)[7] discussed both time and

failure truncated data by using noninformative priors. They derived prediction distributions of future failure times and the number of failures in some future time interval. It involves complicated numerical calculation. Calabria(1990)[13] also derived prediction distribution by using noninformative and informative priors. These references are given on a single system and usually assume parameters are independent. Crow(1974)[21] and Bain(1978)[3] analyzed independent equivalent multi-system by employing power law process. Power bounds for a test of equality of trends in several independent power law processes were discussed by Calabria,R., Guida,M. and Pulcini,G.(1992)[14]. Huang and Bier (1998) [33] presented a natural conjugate prior for the PLP.

1.5.2 Model Motivation

The most commonly used models for repairable systems are the homogeneous and nonhomogeneous Poisson processes. Let us start with different data sets. Figure 1.2 displays the time dot plots and scatter plots of cumulative failure number against cumulative failure time under different situations of repairable systems.

In figure 1.2, the first situation (a) illustrates an improving system. After removing bugs, times between failures tend to get longer and system is **improving**. The intensity function decreases since the probability of failures gets smaller when system ages. This can be employed as a reliability growth model and hence is of the most interest. The second situation (b) illustrates a steady system. Times between failures tend to stay the same. The intensity function remains constant since the probability of failure does not change. The third situation (c) illustrates a deteriorating system. After removing bugs, times between failures tend to get shorter and system is **deteriorating**. The intensity function increases since the probability of failure will gets larger when system ages.

Usually the assumptions of independent and identical distribution for times be-

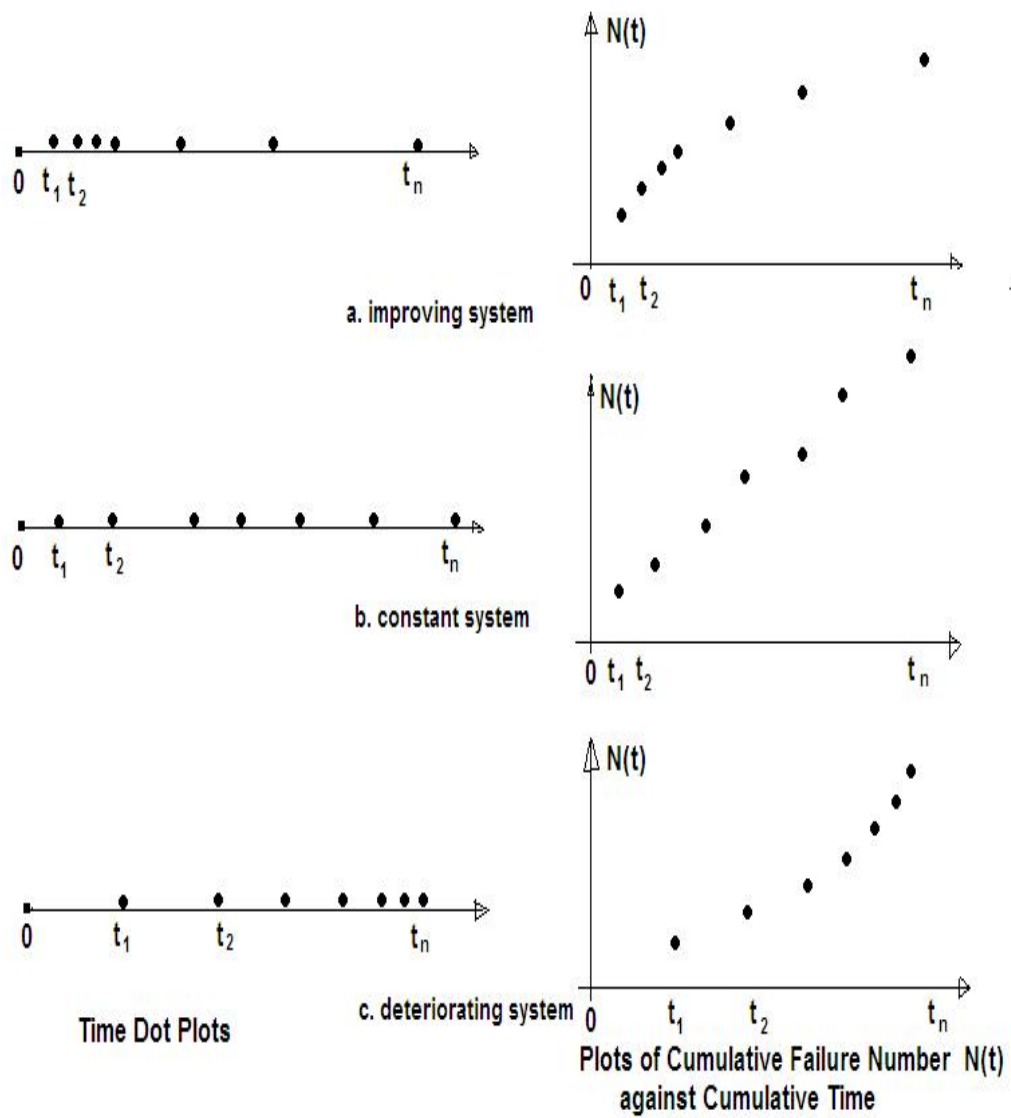


Figure 1.2: Three Different Types of Systems

tween failures in repairable system are invalid. We must consider models in which the assumptions do not hold. The intensity function (repair rate) plays an important role for selecting model because it contains the information about likelihood of a failure at or around any time t . The intensity function changes when system ages. In situation (a), the repair system is improving. In situation (c), the system is deteriorating. Under those two cases, we should employ NHPP. In situation (b), the intensity function is a constant, HPP is a more proper choice. If NHPP is selected as the model, a very commonly used process in NHPP is the power law process (PLP). Power law process is flexibly enough to set up models for those three situations by applying different values of shape parameter β . Details will be given in the next chapter. Except PLP is very flexible, the fact that mean function can be easily derived is also a plus.

1.5.3 Present Study

We make parameters transformations $\mu = \ln\alpha$, $\theta = 1/\beta$. A location parameter μ and a scale parameter θ are obtained. This makes noninformative priors more appropriate. Thus we applied non-informative priors to get the posterior densities of μ and θ and got Bayes estimators of μ and θ . A newly developed Bayes estimator of intensity function is shown. We compared the prediction result by employing posterior inferences with the result by employing a Bayesian estimator of intensity function. The current approach simplifies the calculation considerably. The posterior densities of several parameter functions are discussed in the last part. Bayesian approach requires numerical integration. We either use some approximation method or create computer program to calculate the data.

We also provide our original work by applying natural conjugate priors and non-parametric Kernel priors in Empirical Bayes analysis for the power law process. For

the natural conjugate priors, two-hyperparameter prior and a more generalized three-hyperparameter prior (Huang and Bier(1998)[33]) are used. Given the estimates of hyperparameters, we obtain closed forms for prior and posterior distributions in a special case.

Another area of current research interests focuses on Bayesian and Empirical Bayes methods on microarray. Currently we have an opportunity to work with Dr. Haura on microarray data analysis in Moffitt Lee Cancer Center—one of the largest national cancer centers. In this work, the main object is to select differentially expressed genes in around 22,000 genes. The data are nonpaired 5-control (GFP protein) and 5-experiment (Stat3) gene expression. Both classical and Bayesian methods are applied. We briefly address the statistical structure and illustrate the results of two classical methods SAM (Storey 2002), local Z-score (Chen,Z.[17]) and one parametric Empirical Bayes (Newton, 2002). Several partial lists of differentially expressed genes are shown. The results are obtained by SAS programming and research software packages. We also discuss some normalization methods and applications on probe-level and expression-level data.

Bayesian methods are developed in many other fields of microarray analysis, for instance, assessing differential expression (Newton 2002, Speed 2002, Smyth 2003), clustering (Sebastiani 2002), decomposition (Ochs 2002), principal component analysis and prediction (Mike West, 2000). The main problems on microarray are from low replicates and large amount of genes. Bayesian analysis partially contributes to solve the problem by considering the typical variability in the system. Further work can be done by employing Bayesian or Empirical Bayes analysis in microarray.

1.6 Summary

We address some fundamental concepts which are involved in reliability model and reliability growth. A class of point processes is usually used to model repairable system, such as homogeneous Poisson processes, nonhomogeneous Poisson processes, Branching Poisson processes and renewal processes. Our research will focus on a commonly used nonhomogeneous process–Power Law Process. We give a historical review on the power law process in this chapter. And we showed a brief idea how the power law process can be applied in reliability growth models. Our research interest also includes Bayesian and Empirical Bayes approaches on microarray analysis, especially on the area of selection of differentially expressed genes.

Chapter 2

Classical Inference on the Power Law Process

In this chapter, we address some classical inference results on the Power Law Process. These include point and interval estimates for the parameters, hypothesis testing, estimation of the intensity function and estimation of the mean time between failures (MTBF). We will also discuss three goodness-of-fit tests. The first two will be illustrated with real data set.

2.1 Introduction

As mentioned in the last chapter, the power law process can be described as a nonhomogeneous Poisson process $\{N(t), t \geq 0\}$ with intensity function:

$$\nu(t) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}, \text{ for } \alpha > 0, \beta > 0.$$

The mean value function $\lambda(t)$ of the process is:

$$\lambda(t) = E(N(t)) = \int_0^t \nu(s) ds = \int_0^t \left(\frac{\beta}{\alpha}\right)\left(\frac{s}{\alpha}\right)^{\beta-1} ds = (t/\alpha)^\beta.$$

The shape parameter β affects how the system deteriorates or improves over time. If $\beta > 1$, the intensity function $\nu(t)$ is increasing, then the failures tend to occur more frequently, and we call the system **deteriorating**. If $\beta < 1$, the intensity function $\nu(t)$ is decreasing, then the system is **improving**. Under this situation, the power law process can be applied as a reliability growth model. If $\beta = 1$, the power law process is reduced to

a simple homogeneous Poisson process with intensity $1/\alpha$, where α is a scale parameter.

The power law process has been widely used as models in repairable systems (Ascher And Feingold (1984)[2], Engelhardt and Bain(1986)[26], Rigdon and Basu(1989)[49]) and software reliability growth models (Kyparisis and Singpurwalla(1985) [35]). For example, Duane (1964) demonstrated that many systems developed at General Electric seemed to follow a model closely related to the power law process.

2.2 Point Estimation of Parameters β and α

There are two different sampling protocols which provide data on the power law process: (i) failure truncated case and (ii) time truncated case. These two terms are defined in the previous chapter.

The joint pdf of the failure times T_1, T_2, \dots, T_n from a NHPP with intensity function $\nu(t)$ is then given by Crow(1982)[22],

$$f(t_1, t_2, \dots, t_n) = \left(\prod_{i=1}^n \nu(t_i) \right) \exp\left(- \int_0^w \nu(x) dx\right). \quad (2.1)$$

where w is a so-called stopping time: $w = t_n$ for the failure truncated case, $w = t$ for the time truncated case. Thus for the failure truncated case with $\nu(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1}$, the joint density of $T_1 < T_2 \dots < T_n$ is obtained from equation (2.1):

$$\begin{aligned} f(t_1, t_2, \dots, t_n) &= \left(\prod_{i=1}^n \frac{\beta}{\alpha} \left(\frac{t_i}{\alpha}\right)^{\beta-1} \right) \exp\left(- \int_0^{t_n} \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} dx\right) \\ &= \left(\frac{\beta}{\alpha}\right)^n \left(\prod_{i=1}^n \frac{t_i}{\alpha}\right)^{\beta-1} \exp\left(-\frac{t_n}{\alpha}\right)^\beta. \end{aligned} \quad (2.2)$$

For the time truncated case, we observed $t_1 < t_2 < \dots < t_N < t$. The number of failures N in time truncated interval $[0, t)$ is a random variable. Given $N = n$, the distribution of $T_1 < T_2 \dots < T_n$ can be shown by using property 3 of NHPP,

$$f(t_1, t_2, \dots, t_n | n) = n! \prod_{i=1}^n \frac{\beta}{t} \left(\frac{t_i}{t}\right)^{\beta-1}$$

and the random variable N has a Poisson distribution with mean $(t/\alpha)^\beta$, so

$$f_N(n) = \frac{(t/\alpha)^{n\beta} \exp[-(t/\alpha)^\beta]}{n!} \quad n = 0, 1, \dots$$

Thus, the joint density of N and $T_1 < T_2 < \dots < T_n$ is

$$\begin{aligned} f(n, t_1, t_2, \dots, t_n) &= \frac{(t/\alpha)^{n\beta} \exp[-(t/\alpha)^\beta]}{n!} n! \prod_{i=1}^n \frac{\beta}{t} \left(\frac{t_i}{t}\right)^{\beta-1} \\ &= \left(\frac{\beta}{\alpha}\right)^n \left(\prod_{i=1}^n \frac{t_i}{\alpha}\right)^{\beta-1} \exp\left(-\frac{t}{\alpha}\right)^\beta. \end{aligned} \quad (2.3)$$

It is possible that we observe no failure before time t . In this case, $f(0) = \exp[-(t/\alpha)^\beta]$. This is of no inferential interest, and hence we won't consider this case. It is remarkable that the likelihood functions are almost identical for the failure truncated case (2.2) and the time truncated case (2.3). Those two likelihoods can be written as

$$L(\alpha, \beta | n, t_1, t_2, \dots, t_n) = \left(\frac{\beta}{\alpha}\right)^n \left(\prod_{i=1}^n \frac{t_i}{\alpha}\right)^{\beta-1} \exp\left(-\frac{w}{\alpha}\right)^\beta, \quad (2.4)$$

where $w = t_n$ for the failure truncated case, $w = t$ for the time truncated case.

2.2.1 Maximum Likelihood Estimates $\hat{\beta}$ and $\hat{\alpha}$

Given the likelihood function (2.4) as above section, the log-likelihood function is

$$L(\alpha, \beta | t_1 \dots t_n, n) = n \log \beta - n \beta \log \alpha + (\beta - 1) \sum_{i=1}^n \log t_i - \left(\frac{t_n}{\alpha}\right)^\beta.$$

Setting the first partial derivatives (with respect to β and α) equal to zero, we obtain the MLE's as following,

$$\begin{aligned} \hat{\beta} &= \frac{n}{\sum_{i=1}^n \log(w/t_i)}, \\ \hat{\alpha} &= \frac{w}{n^{1/\hat{\beta}}}, \end{aligned}$$

where $w = t_n$ for the failure truncated case, $w = t$ for the time truncated case. They are biased estimates. It is known (Guida (1989)[30] and Rigdon and Basu(1989)[49]) that

$2n\beta/\hat{\beta}$ has a chi-square distribution with $2(n - \gamma)$ degrees of freedom, where $\gamma = 1$ for failure truncated case and $\gamma = 0$ for time truncated case.

Now we will show

$$E(\hat{\beta}) = \frac{n\beta}{(n - 1 - \gamma)}, \quad Var(\hat{\beta}) = \frac{n^2\beta^2}{(n - 1 - \gamma)^2(n - 2 - \gamma)}.$$

For the proof we shall use the following Lemma.

Lemma. Let X be χ^2 distributed with n degrees of freedom, then

$$E(X^k) = 2^k \frac{\Gamma(\frac{n}{2} + k)}{\Gamma(\frac{n}{2})} \text{ where } k \text{ is an integer s.t. } \frac{n}{2} + k > 0.$$

In particular,

$$\begin{aligned} E(X) &= n, & E\left(\frac{1}{X}\right) &= \frac{1}{n - 2}, \\ E(X^2) &= n^2 + 2n, & E\left(\frac{1}{X^2}\right) &= \frac{1}{(n - 2)(n - 4)}. \end{aligned}$$

Proof. By the Lemma above, we have

$$\begin{aligned} E(\hat{\beta}) &= 2n\beta E\left(\frac{1}{\chi_{2(n-\gamma)}^2}\right) \\ &= \frac{n\beta}{n - 1 - \gamma}. \end{aligned}$$

We also have

$$\begin{aligned} E(\hat{\beta}^2) &= E\left(\frac{2n\beta}{\chi_{2(n-\gamma)}^2}\right)^2 \\ &= \frac{4n^2\beta^2}{[2(n - \gamma) - 2][2(n - \gamma) - 4]} \\ &= \frac{n^2\beta^2}{(n - 1 - \gamma)(n - 2 - \gamma)}. \end{aligned}$$

Thus, the variance of the MLE of β is

$$Var(\hat{\beta}) = E(\hat{\beta}^2) - [E(\hat{\beta})]^2 = \frac{n^2\beta^2}{(n - 1 - \gamma)^2(n - 2 - \gamma)}.$$

A modified maximum likelihood estimate of β is given by Suresh and Rao(1992).

This modified MLE is given by

$$\beta' = \frac{(n - 1)}{n} \hat{\beta}.$$

2.2.2 Unbiased Estimates $\bar{\beta}$

The MLE's are biased estimates and we can adjust them to unbiased estimates.

The unbiased estimators are (Bain and Engelhardt(1991)[4]),

$$\bar{\beta} = \frac{n-1-\gamma}{n} \hat{\beta} = \frac{n-1-\gamma}{\sum_{i=1}^n \log(t_n/t_i)},$$

$\gamma = 1$ for the failure truncated case, and $\gamma = 0$ for the time truncated case.

The variance of $\bar{\beta}$ is

$$Var(\bar{\beta}) = Var\left(\frac{n-1-\gamma}{n} \hat{\beta}\right) = \frac{\beta^2}{n-2-\gamma}.$$

In Calabria *et al.* (1988), it was shown that the unbiased estimate of β is more efficient than the biased estimate.

2.2.3 Linearly Best Efficient Estimate of $\hat{\beta}$

In Qiao and Tsokos (1998), they showed that there exists a linearly best efficient estimate of β , denoted by $\hat{\beta}$.

Theorem. Assume $\bar{\theta}$ is an unbiased estimate of θ , and $\bar{\theta}$ has a finite variance, then there exists an unique α_0 such that

$$MSE(\alpha_0 \bar{\theta}) = \min_{\alpha} MSE(\alpha \bar{\theta}).$$

Moreover,

$$\alpha_0 = \frac{\theta^2}{\theta^2 + Var(\bar{\theta})}.$$

In the above theorem, the MSE is defined as the expected value of the square of the deviation of the estimate from the true value and it equals to the square of variance plus the square of bias.

Applying their theorem, we have

$$\alpha_0 = \frac{\beta^2}{\beta^2 + \left[\frac{\beta^2}{n-2-\gamma}\right]},$$

and the best efficient estimate of β is

$$\hat{\beta} = \frac{n-2-\gamma}{n-1-\gamma} \bar{\beta} = \frac{n-2-\gamma}{\sum_{i=1}^n \log(t_n/t_i)}.$$

We also have

$$\begin{aligned} MSE(\alpha_0 \bar{\beta}) &= \alpha_0^2 Var(\bar{\beta}) + (\alpha_0 - 1)^2 \beta^2 \\ &= \frac{\beta^4 Var(\bar{\theta})}{[\beta^2 + Var(\bar{\beta})]^2} + \frac{\beta^2 Var(\bar{\theta})^2}{[\beta^2 + Var(\bar{\beta})]^2} \\ &= \frac{\beta^2 Var(\bar{\theta})}{[\beta^2 + Var(\bar{\beta})]} \\ &= \frac{\beta^2}{n-1-\gamma}. \end{aligned}$$

The efficiency of MLE $\hat{\beta}$, unbiased estimate $\bar{\beta}$ and linearly best estimate $\hat{\beta}$ are

$$EFF(\hat{\beta}|\bar{\beta}) = \frac{MSE(\hat{\beta})}{MSE(\bar{\beta})} = \frac{n-2-\gamma}{n-1-\gamma} < 1,$$

$$EFF(\hat{\beta}|\hat{\beta}) = \frac{MSE(\hat{\beta})}{MSE(\hat{\beta})} = \frac{n-3}{n+6} < 1, \text{ (Failure truncated case)}$$

$$EFF(\hat{\beta}|\hat{\beta}) = \frac{MSE(\hat{\beta})}{MSE(\hat{\beta})} = \frac{(n-2)(n-3)}{(n-1)(n+2)} < 1, \text{ (Time truncated case)}$$

$$EFF(\bar{\beta}|\hat{\beta}) = \frac{MSE(\bar{\beta})}{MSE(\hat{\beta})} = \frac{n-2}{n+6} < 1, \text{ (Failure truncated case)}$$

$$EFF(\bar{\beta}|\hat{\beta}) = \frac{MSE(\bar{\beta})}{MSE(\hat{\beta})} = \frac{(n-2)^2}{(n-1)(n+2)} < 1. \text{ (Time truncated case)}$$

Therefore, the linearly best estimate $\hat{\beta}$ has the greatest efficiency and unbiased estimate $\bar{\beta}$ has a better efficiency than MLE $\hat{\beta}$.

2.3 Interval Estimation and Tests of Hypothesis

We again apply the fact that $2n\beta/\hat{\beta}$ has a chi-square distribution with $2(n-\gamma)$ degrees of freedom, where $\gamma = 1$ when failure truncated case and $\gamma = 0$ when time truncated case. Then we can write

$$P\left(\chi_{1-\alpha/2}^2(2(n-\gamma)) < \frac{2n\beta}{\hat{\beta}} < \chi_{\alpha/2}^2(2(n-\gamma))\right) = 1 - \alpha$$

Note: α is the significance level and not the parameter α here. So the confidence interval for β is

$$\frac{\chi_{1-\alpha/2}^2(2(n-\gamma))\hat{\beta}}{2n} < \beta < \frac{\chi_{\alpha/2}^2(2(n-\gamma))\hat{\beta}}{2n}.$$

The result that $(2n\beta)/(\hat{\beta})$ has a chi-square distribution with $2(n-\gamma)$ degrees of freedom can also be used to construct a test at significant level α for

$$H_0 : \beta = \beta_0 \text{ versus } H_1 : \beta \neq \beta_0.$$

The rule to reject H_0 is

$$\frac{2n\beta_0}{\hat{\beta}} < \chi_{1-\alpha/2}^2 2(n-\gamma) \text{ or } \frac{2n\beta_0}{\hat{\beta}} > \chi_{\alpha/2}^2 2(n-\gamma),$$

that is,

$$\hat{\beta} < \frac{2n\beta_0}{\chi_{\alpha/2}^2 2(n-\gamma)} \text{ or } \hat{\beta} > \frac{2n\beta_0}{\chi_{1-\alpha/2}^2 2(n-\gamma)}.$$

It is often useful to test $H_0 : \beta = 1$ versus $H_1 : \beta \neq 1$. The power law process reduces to the homogeneous Poisson process when $\beta = 1$, and it tests whether the system is remaining stable or not. Alternative test can also be $H_1 : \beta > 1$ which means the system is deteriorating or $H_1 : \beta < 1$ which means the system is improving.

2.4 Estimation of Intensity Function

Recall Power Law process's intensity function is

$$\nu(t) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}, \text{ for } \alpha > 0, \beta > 0.$$

The simplest way is using maximum likelihood estimates of α and β to evaluate $\nu(t)$. We have

$$\hat{\nu}(t) = \left(\frac{\hat{\beta}}{\hat{\alpha}}\right)\left(\frac{w}{\hat{\alpha}}\right)^{\hat{\beta}-1} = \frac{n\hat{\beta}}{w}$$

where $w = t_n$ for the failure truncated case, $w = t$ for the time truncated case. Other estimates are obtained by Tsokos and Rao (1995)[59] and Qiao and Tsokos (1998)[44].

Rigdon and Basu(2000)[50] combined the failure truncated case and time truncated case together, in summary, the estimates of intensity function are unbiased Estimate

$$\hat{v}_{UB} = \frac{(n-1)(n-2)\hat{\beta}}{nw}$$

and best efficient estimate with minimum mean squared error

$$\hat{v}_{MMSE} = \frac{(n-2)(n-3)\hat{\beta}}{nw}.$$

2.5 Mean Time Between Failure (MTBF)

Mean time between failures is defined to be the average time that a component works without failure. It is an important metric that assesses the reliability of repairable system. The reciprocal of the intensity function is widely accepted as an approximate estimate of the MTBF (Cox & Lewis 1966)[20], denoted by $MTBF_A$. However, such a relationship is only true for HPP (Thompson(1981)[57]). The expected time between the n th and the $(n+1)$ th failure, denoted by $MTBF_n$, is the mean time between failure in time interval (t_n, t_{n+1}) . In this section, we first derive $MTBF_n$ as a function of α and β , then we show an estimate of $MTBF_n$ derived by Qiao and Tsokos[44]. Their estimate of $MTBF_n$ will be referred as $MTBF_Q$.

2.5.1 $MTBF_n$

The probability $F(t_{n+1}|t_n)$ of system failing after time t_{n+1} , given that the system last failed at time t_n is equivalent to the probability of the system experiences zero failures between (t_n, t_{n+1}) . This can be used to derive the distribution of $MTBF_n$ for the power law process. From previous section,

$$Pr[N(t_{n+1}) - N(t_n) = 0] = exp\{-[\lambda(t_{n+1}) - \lambda(t_n)]\}.$$

Hence

$$f(t_{n+1}|t_n) = \frac{d}{dt_{n+1}}F(t_{n+1}|t_n) = \frac{d}{dt_{n+1}}\{1 - Pr[N(t_{n+1}) - N(t_n) = 0]\}$$

$$= \nu(t_{n+1})\exp\{-[\lambda(t_{n+1}) - \lambda(t_n)]\}.$$

$$\begin{aligned} MTBF_n &= \left(\int_{t_n}^{\infty} t_{n+1} f_{n+1}(t_{n+1}/t_n) dt_{n+1} \right) - t_n \\ &= \left(\int_{t_n}^{\infty} t_{n+1} \frac{\beta}{\alpha^\beta} (t_{n+1})^{\beta-1} \exp\left(-\frac{1}{\alpha^\beta} [t_{n+1}^\beta - t_n^\beta]\right) dt_{n+1} \right) - t_n \\ &= \left(\frac{1}{\alpha^\beta}\right)^{-1/\beta} \exp\left(\left(\frac{t_n}{\alpha}\right)^\beta\right) \Gamma\left(\frac{1}{\beta} + 1\right) \bar{F}_x. \end{aligned}$$

where $\bar{F}_x = 1 - Pr[X > (\frac{t_n}{\alpha})^\beta]$, $X \sim \Gamma(\frac{1}{\beta} + 1, 1)$.

2.5.2 $MTBF_Q$

Qiao and Tsokos(1998)[45] investigated the relation between MTBF and the reciprocal of the intensity function. They provided upper and lower bounds for the estimate of $MTBF_n$. What it follows is a brief proof for the reliability growth model ($\beta < 1$). Consider the intensity function

$$\nu(t) = \left(\frac{\beta}{\alpha}\right) \left(\frac{t_n}{\alpha}\right)^{\beta-1},$$

Let $\theta = 1/\beta$, $h = (t_n/\alpha)$, then

$$\frac{1}{\nu(t)} = \theta \alpha h^{\theta-1}.$$

We can also rewrite $MTBF_Q$ as

$$MTBF_Q = \alpha e^{h\theta} \int_h^{\infty} e^{-t\theta-1} dt.$$

Case 1: $\beta \leq \frac{1}{2}$

This case is equivalent to $\theta \geq 2$. Therefore we may assume $\theta = m + 1 + \delta$, where $m \geq 1$ is an integer and $\delta \in [0, 1)$. We shall take the special case $\delta = 0$. In this case, the $MTBF_Q$ can be expanded as

$$MTBF_Q = \alpha e^h e^{-h} \sum_{i=0}^m \prod_{k=0}^i (\theta - k) h^{m-i}$$

$$\begin{aligned}
&= \theta h^m \sum_{i=0}^m \frac{1}{h^i} \prod_{k=0}^i (\theta - k) \\
&= \frac{1}{\nu(t)} \left[1 + (\theta - 1) \frac{1}{h} + \dots + \frac{(\theta - 1)(\theta - 2) \dots (\theta - m)}{h^m} \right].
\end{aligned}$$

The expression above shows that the difference between $\frac{1}{\nu(t)}$ is given by

$$\frac{1}{\nu(t)} \left[(\theta - 1) \frac{1}{h} + \dots + \frac{(\theta - 1)(\theta - 2) \dots (\theta - m)}{h^m} \right].$$

For general δ , we have

$$MTBF_Q = \alpha h^{\delta+m} \left[\sum_{i=0}^m \prod_{k=0}^i (\theta - k) h^{m-i} + \theta e^h h^{-\delta} \int_h^{\infty} t^{\theta-m-1} e^{-t} dt \right].$$

From the expression of $MTBF_Q$, it can be easily seen that

$$\begin{aligned}
MTBF_Q &\geq \frac{1}{\nu(t)} \left[(\theta - 1) \frac{1}{h} + \dots + \frac{(\theta - 1)(\theta - 2) \dots (\theta - m)}{h^m} \right] \\
&\geq \frac{1}{\nu(t)} \left[1 + \frac{(\theta - 1)}{h} \right],
\end{aligned}$$

and

$$\begin{aligned}
MTBF_Q &\leq \frac{1}{\nu(t)} \sum_{i=0}^m \frac{\theta - 1^i}{h} \\
&\leq \frac{1}{\nu(t)} \frac{1}{1 - \frac{\theta-1}{h}}.
\end{aligned}$$

Hence,

$$\frac{1}{\nu(t)} \left[1 + \frac{(\theta - 1)}{h} \right] \leq MTBF_Q \leq \frac{1}{\nu(t)} \frac{1}{1 - \frac{\theta-1}{h}},$$

where $\theta = \frac{1}{\beta}$ with $\beta \leq 1/2$, $h = (\frac{t_n}{\alpha})^\beta$ and $\nu(t) = \frac{1}{\frac{\alpha}{\beta} (\frac{t_n}{\alpha})^{\beta-1}}$. Thus the point estimate for $MTBF_Q$ is

$$MTBF_Q = \frac{1}{2\nu(t)} \left[\left(1 + \frac{\alpha - 1}{h} \right) + \left(1 - \frac{\alpha - 1}{h} \right)^{-1} \right]. \quad (2.5)$$

Case 2: $\frac{1}{2} < \beta \leq 1$

In this case, we may write $\theta = 1 + \delta$ with $0 \leq \delta < 1$. Thus

$$MTBF_Q = \alpha e^h \theta \int_h^{+\infty} e^{-t} t^\delta dt \geq \alpha \theta h^\delta e^h \int_h^{+\infty} e^{-t} dt = \alpha \theta h^\delta = \frac{1}{\nu(t)}.$$

We may also obtain

$$MTBF_Q = \alpha e^{h\theta} \left[h^\delta e^{-h} + \delta \int_h^{+\infty} h^{\delta-1} e^{-t} dt \right] \leq \frac{1}{\nu(t)} \left[1 + \frac{\delta}{h} \right].$$

Therefore, in this case, we have

$$\frac{1}{\nu(t)} \leq MTBF_Q \leq \frac{1}{\nu(t)} \left[1 + \frac{\delta}{h} \right].$$

Thus the point estimate for $MTBF_Q$ for this case is

$$MTBF_Q = \frac{1}{2\nu(t)} \left[2 + \frac{\delta}{h} \right].$$

For the intensity function $\nu(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta-1}$, we conclude

$$MTBF_Q = \begin{cases} \frac{MTBF_A}{2} \left[\left(1 + \frac{\theta-1}{h} \right) + \left(\frac{1}{1-\frac{\theta-1}{h}} \right) \right], & 0 < \beta \leq \frac{1}{2} \\ \frac{MTBF_A}{2} \left(2 + \frac{\delta}{h} \right), & \frac{1}{2} < \beta \leq 1 \\ MTBF_n. & \beta > 1 \end{cases}$$

2.6 Goodness-of-fit Tests

There are several ways to assess the fit of power law process. The Duane plot is an informal graphical method. Exact goodness-of-fit tests can be constructed by making an appropriate transformation of the failure times. Such transformations include ratio-power transformation and log-ratio transformation. We use the following software failure time table to illustrate how the Duane plot and ratio-power transformation work in testing the goodness-of-fit. This data set consists of 38 software failure times taken from Musa(1979)[41].

Software Failure Times in Seconds

i	t_i	i	t_i	i	t_i	i	t_i
1	115	11	1955	21	6162	31	36800
2	115	12	2026	22	6552	32	37363
3	198	13	2632	23	8415	33	40133
4	376	14	3821	24	9752	34	40785
5	570	15	3861	25	14260	35	46378
6	706	16	4649	26	15094	36	58074
7	1780	17	4871	27	18494	37	64798
8	1798	18	4943	28	18500	38	67344
9	1813	19	5558	29	23061		
10	1905	20	6147	30	26229		

i : Failure number; t_i : Cumulative Failure time. (Musa(1979)[41])

Table 2.1

2.6.1 Duane Plot

If the power law process is the correct model, Duane plots should be roughly linear.

This is derived from the following:

$$E[N(t)] = \lambda(t) = \left(\frac{t}{\alpha}\right)^\beta.$$

Thus,

$$E\left[\frac{N(t)}{t}\right] = \frac{1}{t}\lambda(t) = \left(\frac{t^{\beta-1}}{\alpha^\beta}\right).$$

After taking natural logarithm of both sides, we have

$$\log E\left[\frac{N(t)}{t}\right] = (\beta - 1)\log t - \beta \log \alpha,$$

this shows $\log E[N(t)/t]$ is a linear function of $\log t$ assuming t is fixed, in which $N(t)$ is random. From Figure 2.1, it suggests the power law process is indeed proper since the Duane plot shows a linear relation.

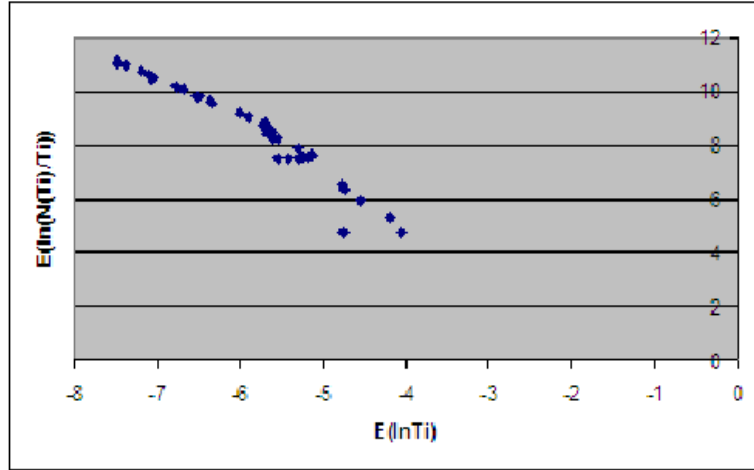


Figure 2.1: Duane Plot for Data Table 2.1

2.6.2 Ratio Power Test

A common goodness-of-fit is the ratio-power transformation, which is defined by $\hat{R}_i = (t_i/t_n)^{\bar{\beta}}$, $i = 1, 2, \dots, n - 1$, where $\bar{\beta}$ is an unbiased estimator which is obtained as in subsection 2.1.2.

H_0 : The power law process is correct model.

H_1 : The power law process is not correct model.

The test statistic for the Cramer-von Mises test is

$$C_R^2 = \frac{1}{12(n-1)} + \sum_{i=1}^{n-1} \left(\hat{R}_i - \frac{2i-1}{2(n-1)} \right)^2.$$

From the data table 2.1, the statistic $C_R^2 = 0.25$. We accept the null hypothesis at 5% level. Furthermore, Figure 2.2 is a scatter plot of number of failures. It suggests an improving system since it is concave down.

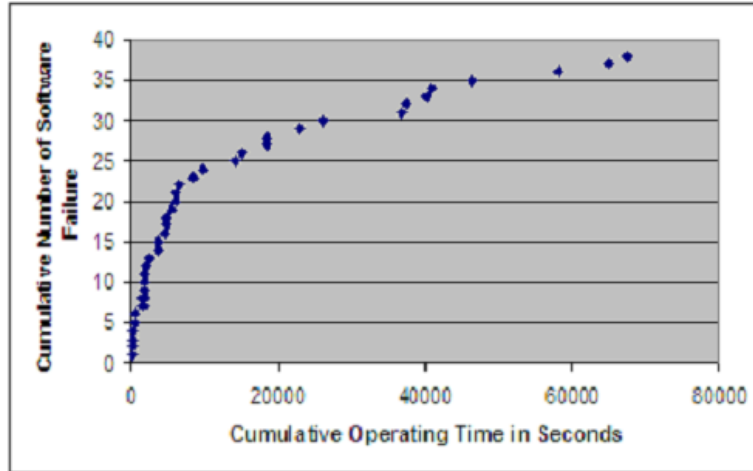


Figure 2.2: Plot of Number of Failures against Cumulative Operating Time for Data Table 2.1

2.6.3 Log-ratio Test

In Log-ratio goodness-of-fit test for the power law process, we make a log-ratio transformation

$$U_i = \log(t_n/t_{n-i}).$$

If the power law process with parameters β and α is the proper model, it can be shown that $U_1 < U_2 < \dots < U_{n-1}$ are distributed as $n - 1$ order statistics from an exponential distribution with mean $1/\beta$. Thus, any goodness-of-fit test for the exponential distribution with unknown mean (β is usually unknown) can be used to test the adequacy of the power law process. For example, Lilliefors' test (1969), Shapiro-Wilk W test(1972) and Stephens' test (1974), etc.

2.7 Summary

Classical inferences on the power law process have been done during the past decades[36, 26, 3, 49, 12, 44]. Basic inferences such as point and interval estimates,

hypothesis test of model parameters were given. We derived a unified form of linearly best efficient estimates of the scale parameter for the failure truncated and time truncated data. For completeness, we also included the estimation of intensity function and mean time between failures (MTBF). In the last part, we applied real data set to show how three existing goodness of fit tests for the PLP work in model check.

Chapter 3

Bayesian Inference on the Power Law Process

Recall that the power law process can be described as a nonhomogeneous Poisson process $\{N(t), t \geq 0\}$ with intensity function

$$\nu(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1}, \text{ for } \alpha > 0, \beta > 0.$$

The mean value function $\lambda(t)$ of the process is:

$$\lambda(t) = E(N(t)) = \int_0^t \nu(s) ds = \int_0^t \frac{\beta}{\alpha} \left(\frac{s}{\alpha}\right)^{\beta-1} ds = (t/\alpha)^\beta.$$

In this chapter, we use the transformation $\mu = \ln \alpha$, $\theta = 1/\beta$ and obtain the location parameter μ and the scale parameter θ . We develop Bayesian procedures for the power law process with failure truncation data, using non-informative priors for the scale parameter θ and the location parameter μ . Bayesian inference is different from the classical methods since we take the parameters θ and μ as random variables instead of fixed numbers. In addition to obtaining the posterior density of parameters of the power law process, Bayesian prediction inferences for the expected number of failures and the future failure times are discussed. Predictive inference based on Bayesian estimation of the intensity function greatly simplifies the calculations. We compare our results with the paper of Bar-Lev, S.K., Lavi, I. and Reiser, B. (1992) [7] by using the data set from Musa (1979). We also derive posterior densities of system reliability, mean value function and intensity function.

3.1 Likelihood Function

3.1.1 Likelihood of y_1

As pointed out in section 1.5, the time of the first failure T_1 has a Weibull distribution with scale parameter α and shape parameter β , that is, $T_1 \sim \text{Weibull}(\alpha, \beta)$ with pdf

$$f(t_1) = \frac{\beta}{\alpha} \left(\frac{t_1}{\alpha}\right)^{\beta-1} e^{-(t_1/\alpha)^\beta}$$

and cdf

$$\begin{aligned} F(t_1) &= \int_0^{t_1} f(s) ds = \int_0^{t_1} \frac{\beta}{\alpha} \left(\frac{s}{\alpha}\right)^{\beta-1} e^{-(s/\alpha)^\beta} ds \\ &= -e^{-(s/\alpha)^\beta} \Big|_0^{t_1} = 1 - e^{-(t_1/\alpha)^\beta}. \end{aligned}$$

Let $Y_1 = \ln T_1$, then

$$F(y_1) = Pr(Y_1 < y_1) = Pr(\ln T_1 < y_1) = Pr(T_1 < e^{y_1}) = F(e^{y_1}) = 1 - \exp\left\{-\left(\frac{e^{y_1}}{\alpha}\right)^\beta\right\}.$$

Take derivative of $F(y_1)$, pdf of Y_1 becomes

$$f(y_1) = \frac{d}{dy_1} F(y_1) = \frac{d}{dy_1} Pr(Y_1 < y_1) = \frac{\beta}{\alpha^\beta} e^{y_1 \beta} e^{-\left(\frac{e^{y_1}}{\alpha}\right)^\beta}.$$

Let $\mu = \ln \alpha$, $\theta = \frac{1}{\beta}$, we have

$$f(y_1) = \frac{1}{e^{\frac{\mu}{\theta}}} \cdot e^{\frac{y_1}{\theta}} \cdot e^{-e^{\frac{y_1 - \mu}{\theta}}} = \frac{1}{\theta} \cdot \exp\left\{\frac{y_1 - \mu}{\theta} - \exp\left(\frac{y_1 - \mu}{\theta}\right)\right\}, \quad \theta > 0.$$

3.1.2 Likelihood of $\vec{y} = (y_1, y_2, y_3, \dots, y_k)$

Let $Y_k = \ln T_k$, for $k = 2, \dots, n$, where T_k is the k th ordered failure time, then

$$\begin{aligned} Pr\{Y_k > y_k | t_{k-1}\} &= Pr\{\ln T_k > y_k | t_{k-1}\} = Pr\{T_k > e^{y_k} | e^{y_{k-1}}\} \\ &= Pr\{\text{zero failure in } (e^{y_{k-1}}, e^{y_k}]\} = \exp\left(-\int_{e^{y_{k-1}}}^{e^{y_k}} \nu(s) ds\right) \\ &= \exp(-\lambda(s)|_{e^{y_{k-1}}}^{e^{y_k}}). \end{aligned}$$

Since the mean value function is $\lambda(t) = (t/\alpha)^\beta$, we have

$$\lambda(s) \Big|_{e^{y_{k-1}}}^{e^{y_k}} = \frac{1}{\alpha^\beta} \cdot e^{y_k \beta} - \frac{1}{\alpha^\beta} \cdot e^{y_{k-1} \beta} \text{ for } k = 2, \dots, n.$$

Therefore, cdf of Y_k conditional on observation y_{k-1} is

$$\begin{aligned} F(y_k|y_{k-1}) &= Pr\{Y_k > y_k|y_{k-1}\} \\ &= \exp\left\{-\lambda(y) \Big|_{e^{y_{k-1}}}^{e^{y_k}}\right\} \\ &= \exp\left\{-\left(\frac{e^{y_k}}{\alpha}\right)^\beta + \left(\frac{e^{y_{k-1}}}{\alpha}\right)^\beta\right\} \end{aligned}$$

and the pdf of Y_k conditional on observation y_{k-1} is

$$f(y_k|y_{k-1}) = \left(\frac{\beta}{\alpha^\beta}\right) \cdot \exp\left\{y_k \beta - \left(\frac{e^{y_k}}{\alpha}\right)^\beta + \left(\frac{e^{y_{k-1}}}{\alpha}\right)^\beta\right\}.$$

Hence, the joint likelihood function of \vec{y} for the failure truncated case is

$$\begin{aligned} L(\vec{y}) &= L(y_n, y_{n-1}, y_{n-2}, \dots, y_1) \\ &= f(y_n|y_{n-1}) \cdot f(y_{n-1}|y_{n-2}) \cdots f(y_2|y_1) \cdot f(y_1) \\ &= \left(\frac{\beta}{\alpha^\beta}\right)^n \exp\left\{\beta \sum_1^n y_i - \frac{1}{\alpha^\beta} e^{y_n \beta}\right\}. \end{aligned} \tag{3.1}$$

Similarly, it can be shown that for time truncated data, likelihood function is

$$L(\vec{y}) = \left(\frac{\beta}{\alpha^\beta}\right)^n \exp\left\{\beta \sum_1^n y_i - \frac{1}{\alpha^\beta} e^{y^\beta}\right\}$$

where $y = \ln t$ and t is the stopping time. To simplify our work, we only consider the failure truncated case. However, inferences for the time truncated case is similar since we have a similar likelihood function.

Now we make a transformation on the parameters α and β . Let $\mu = \ln \alpha$, $\theta = \frac{1}{\beta}$, then the joint likelihood function (3.1) of \vec{y} becomes

$$L(\vec{y}) = L(y_n, y_{n-1}, y_{n-2}, \dots, y_1) = \left(\frac{1}{\theta}\right)^n \exp\left\{\frac{\sum_{i=1}^n (y_i - \mu)}{\theta} - \exp\left(\frac{y_n - \mu}{\theta}\right)\right\}. \tag{3.2}$$

Hence, from the likelihood function (3.2), the classical MLE's of θ and μ are obtained as

$$\hat{\theta} = y_n - \sum_1^n y_i/n$$

and

$$\hat{\mu} = y_n - \hat{\theta} \cdot \ln(n).$$

3.2 Posterior Density of (μ, θ)

We note that μ is a location parameter and θ is a scale parameter. From Box and Tiao(1973)[10], the noninformative priors for μ and θ are

$$\pi_o(\mu) = \text{constant}, \quad \pi_o(\theta) \propto \frac{1}{\theta}$$

We also note that these priors are both improper since integral of priors are not finite. If μ and θ are independent, the joint prior is $\pi_o(\mu, \theta) \propto 1/\theta$. If μ and θ are dependent, $\pi_o(\mu, \theta) \propto 1/\theta^2$ (Bar-Lev,S.K. *et al.*(1992)[7]). Here we assume that μ and θ are independent. Hence, by the Bayes' rule the posterior density $\pi(\mu, \theta|\vec{y})$ is:

$$\pi(\mu, \theta|\vec{y}) = \frac{L(\vec{y}|\mu, \theta) \times \frac{1}{\theta}}{\int_{-\infty}^{\infty} \int_0^{\infty} L(\vec{y}|\mu, \theta) \times \frac{1}{\theta} d\theta d\mu}. \quad (3.3)$$

We now compute the marginal density $m(\vec{y})$ of \vec{y} . It is given by

$$\begin{aligned} m(\vec{y}) &= \int_{-\infty}^{\infty} \int_0^{\infty} L(\vec{y}|\mu, \theta) \times \frac{1}{\theta} d\theta d\mu \\ &= \int_{-\infty}^{\infty} \int_0^{\infty} \left(\frac{1}{\theta}\right)^{n+1} \exp\left\{\frac{\sum_1^n (y_i - \mu)}{\theta} - \exp\left(\frac{y_n - \mu}{\theta}\right)\right\} d\theta d\mu \\ &= \int_0^{\infty} \left(\frac{1}{\theta}\right)^{n+1} \int_{-\infty}^{\infty} \exp\left\{\frac{\sum_1^n y_i}{\theta}\right\} \exp\left\{\frac{-n\mu}{\theta}\right\} \exp\left\{-e^{\frac{y_n - \mu}{\theta}}\right\} d\mu d\theta \\ &= \int_0^{\infty} \left(\frac{1}{\theta}\right)^{n+1} \exp\left\{\frac{\sum_1^n y_i}{\theta}\right\} \exp\left\{\frac{-ny_n}{\theta}\right\} \int_{-\infty}^{\infty} \exp\left\{\frac{n(y_n - \mu)}{\theta}\right\} \exp\left\{-e^{\frac{y_n - \mu}{\theta}}\right\} d\mu d\theta \end{aligned}$$

Let $h = \exp\left\{\frac{y_n - \mu}{\theta}\right\}$, then

$$\frac{dh}{d\mu} = e^{\frac{y_n - \mu}{\theta}} \cdot \left(-\frac{1}{\theta}\right).$$

We also have

$$\int_{-\infty}^{\infty} \exp\left\{\frac{n(y_n - \mu)}{\theta}\right\} \exp\left\{-e^{\frac{y_n - \mu}{\theta}}\right\} d\mu = \theta \Gamma(n).$$

Thus,

$$\begin{aligned}
m(\vec{y}) &= \Gamma(n) \int_0^\infty \left(\frac{1}{\theta}\right)^n \exp\left\{\frac{\sum_1^n y_i - ny_n}{\theta}\right\} d\theta \\
&= \frac{\Gamma(n)}{(ny_n - \sum_1^n y_i)^n} \int_0^\infty \left(\frac{ny_n - \sum_1^n y_i}{\theta}\right)^n \exp\left\{-\left(\frac{ny_n - \sum_1^n y_i}{\theta}\right)\right\} d\theta \\
&= \frac{\Gamma(n)\Gamma(n-1)}{(ny_n - \sum_1^n y_i)^{n-1}}.
\end{aligned}$$

Using this in (3.3), we obtain the joint posterior density of (μ, θ) as

$$\pi(\mu, \theta | \vec{y}) = c(\vec{y}) \left(\frac{1}{\theta}\right)^{n+1} \exp\left\{\frac{\sum(y_i - \mu)}{\theta} - \exp\left(\frac{y_n - \mu}{\theta}\right)\right\} \quad (3.4)$$

where

$$c(\vec{y}) = \frac{1}{m(\vec{y})} = \frac{(ny_n - \sum_1^n y_i)^{n-1}}{\Gamma(n)\Gamma(n-1)}.$$

3.3 Posterior Density of θ and μ

3.3.1 Marginal Posterior Density of θ

The marginal posterior density of θ is obtained by taking integral of joint posterior density (3.4) with respect to μ . That is

$$\pi(\theta | \vec{y}) = \int_{-\infty}^{\infty} \pi(\mu, \theta | \vec{y}) d\mu = c(\vec{y}) \Gamma(n) \times \left(\frac{1}{\theta}\right)^n \exp\left\{\frac{\sum_1^n y_i - ny_n}{\theta}\right\}.$$

Figure 3.1 gives the posterior density of θ for the data in table 2.1.

Therefore, Bayesian point estimate for θ under squared error loss is

$$\tilde{\theta}_B = E(\theta | \vec{y}) = \int_0^\infty \theta f(\theta | \vec{y}) d\theta = \frac{ny_n - \sum_1^n y_i}{n-2}.$$

The Bayesian maximum likelihood estimate of θ (obtained as the maxima of the posterior p.d.f of θ) is

$$\hat{\theta}_B = \frac{ny_n - \sum_1^n y_i}{n}.$$

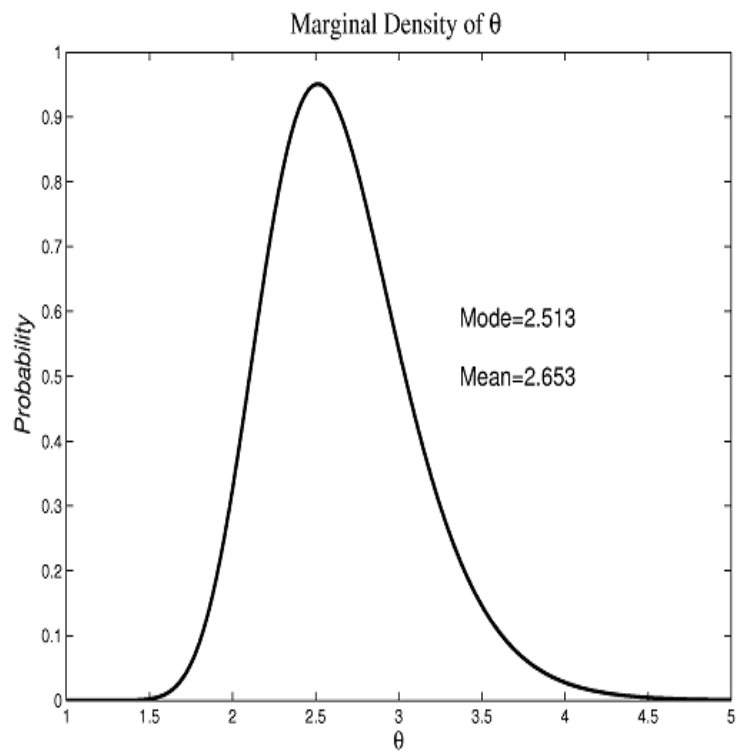


Figure 3.1: Marginal Density of θ

Evaluated from data table 2.1, we have $\tilde{\theta}_B = 2.653$ and $\hat{\theta}_B = 2.513$.

The marginal posterior variance of θ , denoted by $V_{\tilde{\theta}_B}^{\pi(\theta|\vec{y})}$, is the estimation error for $\tilde{\theta}_B$.

$$\begin{aligned} V_{\tilde{\theta}_B}^{\pi(\theta|\vec{y})} &= E^{\pi(\theta|\vec{y})}(\theta - \tilde{\theta}_B)^2 \\ &= E^{\pi(\theta|\vec{y})}(\theta^2) - 2\tilde{\theta}_B E^{\pi(\theta|\vec{y})}(\theta) + \tilde{\theta}_B^2 \\ &= E(\theta^2) - \tilde{\theta}_B^2. \end{aligned}$$

Since

$$\begin{aligned} E(\theta^2) &= \int_0^\infty c(\vec{y})\Gamma(n) \times \left(\frac{1}{\theta}\right)^{n-2} \exp\left\{\frac{\sum_1^n y_i - ny_n}{\theta}\right\} d\theta \\ &= \frac{(ny_n - \sum_1^n y_i)^2}{(n-2)(n-3)}, \end{aligned}$$

it follows that

$$V_{\tilde{\theta}_B}^{\pi(\theta|\vec{y})} = \frac{(ny_n - \sum_1^n y_i)^2}{(n-2)^2(n-3)}.$$

Moreover, the m th moment is given by

$$E(\theta^m) = \frac{(ny_n - \sum_1^n y_i)^m \Gamma(n-m-1)}{\Gamma(n-1)} \quad m = 1, 2, \dots, n.$$

3.3.2 Marginal Posterior Density of μ

The marginal posterior density of μ is obtained by taking integral of joint posterior density (3.4) with respect to θ , hence,

$$\begin{aligned} \pi(\mu|\vec{y}) &= \int_0^\infty \pi(\mu, \theta|\vec{y}) d\theta \\ &= c(\vec{y}) \int_{-\infty}^\infty \left(\frac{1}{\theta}\right)^{n+1} \exp\left\{\frac{\sum(y_i - \mu)}{\theta} - \exp\left(\frac{y_n - \mu}{\theta}\right)\right\} d\theta. \end{aligned}$$

The evaluation of this integral requires numerical procedures. Here we derive an approximate estimator of μ by using Lindley's approximation.

Lindley's Approximation: Lindley (1980) [54] developed an asymptotic approximation to the ratio

$$I = \frac{\int_{\Omega} w(\psi) e^{L(\psi)} d\psi}{\int_{\Omega} \nu(\psi) e^{L(\psi)} d\psi}$$

where $\psi = (\psi_1, \dots, \psi_m)$, $L(\psi)$ is the logarithmic of the likelihood function, $w(\psi)$ and $\nu(\psi)$ are arbitrary functions of ψ and Ω represents the range space of ψ . Clearly, if $w(\psi) = u(\psi)\nu(\psi)$ and $\nu(\psi)$ is the prior distribution of ψ , then Posterior expectation of $u(\psi)$ given the data $x = (x_1, \dots, x_n)$ is

$$I = E(u(\psi)|x)$$

which is the Bayes estimator of $u(\psi)$ under the squared-error-loss function.

To obtain Bayes estimate of μ , we need to approximate I for $m = 2$ and assume ψ_1 and ψ_2 are independent. Lindley gave the following expansion:

$$\begin{aligned} I = & u + \frac{1}{2}(u_{11}\sigma_{11} + u_{22}\sigma_{22}) + \rho_1 u_1 \sigma_{11} + \rho_2 u_2 \sigma_{22} + \\ & + \frac{1}{2}(L_{30}u_1\sigma_{11}^2 + L_{03}u_2\sigma_{22}^2 + L_{21}u_2\sigma_{11}\sigma_{22} + L_{12}u_1\sigma_{22}\sigma_{11}) \end{aligned} \quad (3.5)$$

evaluated at $(\hat{\psi}_1, \hat{\psi}_2)$ and

$$\begin{aligned} u_{11} &= \frac{\partial^2 u}{\partial \psi_1^2}, \quad u_{22} = \frac{\partial^2 u}{\partial \psi_2^2}, \\ L_{30} &= \frac{\partial^3 L}{\partial \psi_1^3}, \quad L_{03} = \frac{\partial^3 L}{\partial \psi_2^3}, \\ L_{21} &= \frac{\partial^3 L}{\partial \psi_1^2 \psi_2}, \quad L_{12} = \frac{\partial^3 L}{\partial \psi_2^2 \psi_1}, \\ \sigma_{11} &= (-L_{20})^{-1} = \left(-\frac{\partial^2 L}{\partial \psi_1^2}\right)^{-1}, \\ \sigma_{22} &= (-L_{02})^{-1} = \left(-\frac{\partial^2 L}{\partial \psi_2^2}\right)^{-1}, \\ \rho_1 &= \frac{\partial}{\partial \psi_1}(-2 \log v(\psi)), \quad \rho_2 = \frac{\partial}{\partial \psi_2}(-2 \log v(\psi)). \end{aligned}$$

In our case, $\psi = (\mu, \theta)$,

$$\begin{aligned} U(\mu, \theta) &= \mu, \quad V(\mu, \theta) = (1/\theta)^{n+1}, \quad L(\mu, \theta) = \frac{\sum(y_i - \mu)}{\theta} - \exp\left(\frac{y_n - \mu}{\theta}\right), \\ U_1 &= 1, \quad U_2 = 0, \quad U_{11} = 0, \quad U_{22} = 0, \\ \rho_1 &= 0, \quad \rho_2 = -1/\theta^2, \\ L_{10} &= -\frac{n}{\theta} + \frac{1}{\theta} \exp\left(\frac{y_n - \mu}{\theta}\right), \quad L_{20} = -\frac{1}{\theta^2} \exp\left(\frac{y_n - \mu}{\theta}\right), \end{aligned}$$

$$\begin{aligned}
L_{21} &= \frac{2}{\theta^3} \exp\left(\frac{y_n - \mu}{\theta}\right) + \frac{1}{\theta^4} \exp\left(\frac{y_n - \mu}{\theta}\right), \quad L_{30} = \frac{1}{\theta^3} \exp\left(\frac{y_n - \mu}{\theta}\right) \\
L_{01} &= -\frac{\sum(y_i - \mu)}{\theta} + \frac{y_n - \mu}{\theta^2} \exp\left(\frac{y_n - \mu}{\theta}\right), \\
L_{02} &= \frac{2\sum(y_i - \mu)}{\theta^3} - \frac{2(y_n - \mu)}{\theta^3} \exp\left(\frac{y_n - \mu}{\theta}\right) - \frac{y_n - \mu}{\theta^4} \exp\left(\frac{y_n - \mu}{\theta}\right), \\
L_{03} &= -\frac{6\sum(y_i - \mu)}{\theta^4} + \frac{6(y_n - \mu)}{\theta^4} \exp\left(\frac{y_n - \mu}{\theta}\right) + \frac{6(y_n - \mu)}{\theta^5} \exp\left(\frac{y_n - \mu}{\theta}\right) \\
&\quad + \frac{y_n - \mu}{\theta^6} \exp\left(\frac{y_n - \mu}{\theta}\right), \\
L_{12} &= -\frac{2n}{\theta^3} + \frac{2}{\theta^3} \exp\left(\frac{y_n - \mu}{\theta}\right) + \frac{2(y_n - \mu)}{\theta^4} \exp\left(\frac{y_n - \mu}{\theta}\right) + \frac{1}{\theta^4} \exp\left(\frac{y_n - \mu}{\theta}\right) \\
&\quad + \frac{y_n - \mu}{\theta^5} \exp\left(\frac{y_n - \mu}{\theta}\right).
\end{aligned}$$

Using the expression (3.5) and some of terms are equal to zero, we obtain

$$\begin{aligned}
E(\mu|\vec{x}) &= \mu + \frac{1}{2}L_{30}\sigma_{11}^2 + \frac{1}{2}L_{12}\sigma_{22}\sigma_{11} \\
&= \mu + \frac{1}{2}L_{30}\left(\frac{1}{L_{20}}\right)^2 + \frac{1}{2}L_{12}\frac{1}{L_{20}L_{02}} \\
&= \mu + \frac{\theta}{2\exp\left(\frac{y_n - \mu}{\theta}\right)} + \frac{\frac{-2n\theta^3}{\exp\left(\frac{y_n - \mu}{\theta}\right)} + 2\theta^3 + 2(y_n - \mu)\theta^2 + \theta^2 + (y_n - \mu)\theta}{-4\theta\sum_1^n(y_i - \mu) + 4\theta(y_n - \mu)\exp\left(\frac{y_n - \mu}{\theta}\right) + 2(y_n - \mu)\exp\left(\frac{y_n - \mu}{\theta}\right)}
\end{aligned}$$

which will be evaluated by $(\hat{\mu}, \hat{\theta})$, which are MLE's of μ and θ .

3.4 Predictive Inference

3.4.1 Predictions Based on Posterior Density

For $\mu = \ln \alpha$ and $\theta = 1/\beta$, the mean function for the power law process becomes $\lambda(t) = \left(\frac{t}{e^\mu}\right)^{\frac{1}{\theta}}$. Let $N(s_1; s_2)$ denote the number of failures occurring in the interval (s_1, s_2) . Then $N(s_1; s_2)$ has a Poisson distribution with mean

$$\lambda(s_1, s_2) = \left(\frac{s_2}{e^\mu}\right)^{\frac{1}{\theta}} - \left(\frac{s_1}{e^\mu}\right)^{\frac{1}{\theta}}.$$

Hence the probability of r failures occur between the time interval (s_1, s_2) is

$$P[N(s_1; s_2) = r | \mu, \theta] = \frac{1}{r!} [\lambda(s_1, s_2)]^r \exp[-\lambda(s_1, s_2)].$$

Consequently, by plugging in the posterior density $\pi(\mu, \theta | \vec{y})$, the predictive distribution of $N(s_1, s_2)$ is

$$\begin{aligned} P[N(s_1; s_2) = r | \vec{y}] &= \int_0^\infty \int_{-\infty}^\infty P[N(s_1; s_2) = r | \mu, \theta] \pi(\mu, \theta | \vec{y}) d\mu d\theta \\ &= \frac{\Gamma(n+r)}{\Gamma(r+1)} c(\vec{y}) \int_0^\infty \theta^{-n} \exp\left(\frac{\sum y_i}{\theta}\right) (s_2^{1/\theta} - s_1^{1/\theta})^r (s_2^{1/\theta} - s_1^{1/\theta} + e^{y_n/\theta})^{-(n+r)} d\theta. \end{aligned} \quad (3.6)$$

An important special case is $s_1 = t_n = e^{y_n}$, that is, we are interested in the probability of number of failures occurring in some future time (t_n, s_2) . In this case, (3.6) reduces to

$$P[N(t_n : s_2) = r | \vec{y}] = \binom{n+r-1}{r} \sum_{k=0}^r (-1)^k \binom{r}{k} \left[\frac{ny_n - \sum y_i}{(n+k)lns_2 - \sum y_i - ky_n} \right]^{n-1}. \quad (3.7)$$

For the special case $s_1 = 0, s_2 = s$, that is, another equivalent system is to begin operating and we want to predict the number of failures of the new system over interval $(0, s)$, for some s of interest, (3.6) reduces to

$$P[N(0 : s) = r | \vec{y}] = \frac{\Gamma(n+r)}{\Gamma(r+1)} c(\vec{y}) \int_0^\infty \theta^{-n} \left(\prod_{i=1}^n ti/s \right)^{1/\theta} (1 + (t_n/s)^{1/\theta})^{-(n+r)} d\theta. \quad (3.8)$$

The integral in (3.6) and (3.8) can be computed numerically. However, for the case $s > y_n$, (3.8) can be written as an infinite sum, (Bar-lev, S.K. *et al.* (1992) [7])

$$P[N(0 : s) = r | \vec{y}] = \binom{n+r-1}{r} \sum_{k=0}^\infty \binom{-n-r}{k} \left[\frac{ny_n - \sum y_i}{(n+k)lns - \sum y_i - ky_n} \right]^{n-1}.$$

If $s = y_n$, then (3.8) can be reduced to

$$P[N(0 : y_n) = r | \vec{y}] = \binom{n+r-1}{n-1} (1/2)^{n+r}.$$

We shall now discuss the prediction of future failure times. Given current available data t_1, t_2, \dots, t_n , we have to predict the future $(n+r)$ th failure time T_{n+r} . Define $Z_r =$

$T_{n+r} - T_n$, conditional on the observation $T_n = t_n$. Then the prediction of T_{n+r} is equivalent to the prediction of Z_r . From equation (3.7), we have

$$P[N(t_n : s_2) = r | \vec{y}] = \binom{n+r-1}{r} \sum_{k=0}^r (-1)^k \binom{r}{k} (\phi(k))^{n-1}$$

where $\phi(k) = \left[\frac{ny_n - \sum y_i}{(n+k)lns_2 - \sum y_i - ky_n} \right]$ and

$$\begin{aligned} P(Z_r \leq z | \vec{y}) &= P(\text{at least } r \text{ failures in } (t_n, t_n + z] | \vec{y}) \\ &= 1 - P(\text{at most } (r-1) \text{ failures in } (t_n, t_n + z] | \vec{y}) \\ &= 1 - \sum_{j=0}^{r-1} \binom{n+j-1}{j} \sum_{k=0}^j \binom{j}{k} (-1)^k (\phi(k))^{n-1}. \end{aligned}$$

(3.9)

The time for $(n+r)$ th failure can now be estimated by evaluating $E(Z_r)$.

3.4.2 Prediction Inference Based on Estimation of Intensity Function

Let $Q = \frac{\nu(T_n)}{\hat{\nu}(T_n)}$. It is known that $nQ \sim \chi_{(n-1)}^2$ approximately (Lee, L. and Lee, K. (1987)).

$$\hat{\nu}(T_n) = \frac{e^{-\hat{\mu}/\hat{\theta}}}{\hat{\theta}} \cdot T_n^{1/\hat{\theta}-1} \text{ and } T_n = e^{Y_n},$$

where $\hat{\theta}$, $\hat{\mu}$ are classical MLE's of θ and μ respectively.

$$\nu(T_n) = \frac{e^{-\mu/\theta}}{\theta} \cdot T_n^{1/\theta-1},$$

which gives $\mu = \theta \ln\left(\frac{\theta\nu}{t_n^{1/\theta-1}}\right)$. Recall that the noninformative priors are $\pi_0(\mu) = \text{constant} = c$, $\pi_0(\theta) \propto 1/\theta$. Hence, the prior of ν conditional on θ is

$$\pi_0(\nu|\theta) = \pi_\mu(\nu) \cdot \frac{d\mu}{d\nu} = c \cdot \theta \cdot \frac{t_n^{1/\theta-1}}{\theta\nu} \cdot \frac{\theta}{t_n^{1/\theta-1}} = \frac{c\theta}{\nu}.$$

Therefore the prior of ν is

$$\pi_0(\nu) \propto \int \pi_0(\nu|\theta) \cdot \pi_0(\theta) d\theta \propto 1/\nu.$$

If ν is given, from the equation $\hat{\nu} = \nu/Q$ and $nQ \sim \chi_{(n-1)}^2$, the p.d.f. $L(\hat{\nu}|\nu)$ of $\hat{\nu}$ conditional on ν is

$$L(\hat{\nu}|\nu) = \frac{\left(\frac{n\nu}{\hat{\nu}}\right)^{\left(\frac{n-3}{2}\right)} e^{-\frac{n\nu}{2\hat{\nu}}}}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)} \left(\frac{n\nu}{\hat{\nu}^2}\right).$$

By the Bayes rule, the posterior density $\pi(\nu(T_n)|\hat{\nu}(T_n)) \propto L(\hat{\nu}|\nu)\pi_0(\nu)$ and

$$\begin{aligned} \pi(\nu|\hat{\nu}) &= \frac{L(\hat{\nu}|\nu)\pi_0(\nu)}{\int L(\hat{\nu}|\nu)\pi_0(\nu) d\nu} \\ &= \frac{\left(\frac{n\nu}{\hat{\nu}}\right)^{\left(\frac{n-3}{2}\right)} e^{-\frac{n\nu}{2\hat{\nu}}}}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)} \left(\frac{n}{\hat{\nu}}\right). \end{aligned} \tag{3.10}$$

It is concluded from the posterior density (3.10) that $\frac{n\nu}{\hat{\nu}} \sim \chi_{n-1}^2$. We have the Bayesian point estimates for ν is

$$\begin{aligned} \tilde{\nu}_B = E(\nu) &= \frac{n-1}{n} \hat{\nu}, \\ \hat{\nu}_B &= \frac{n-3}{n} \hat{\nu}. \end{aligned}$$

We now use $\tilde{\nu}_B$ to give prediction inference. The probability of the number of failures in time interval (t_n, y) is

$$P[N(t_n : y) = r|\tilde{\nu}] = \frac{e^{-\int_{t_n}^y \tilde{\nu}_B(t) dt} \left(\int_{t_n}^y \tilde{\nu}_B(t) dt\right)^r}{r!}. \tag{3.11}$$

The probability of the $(n+1)th$ failure time is

$$\begin{aligned} P(Z_1 \leq z|\vec{t}) &= Pr(\text{at least 1 failure in } [t_n, t_n + z]|\vec{t}) \\ &= 1 - Pr(\text{no failure in } [t_n, t_n + z]|\vec{t}) \\ &= 1 - \exp\left(-\int_{t_n}^{t_n+z} \tilde{\nu}_B(t) dt\right). \end{aligned} \tag{3.12}$$

One of the drawbacks in Bayesian analysis comes from the requirement of numerical calculation. Predictive inference based on Bayesian estimate $\tilde{\nu}_B$ of intensity function greatly simplifies the computation. In the following section, we utilize a data set to show

the results remain very close as traditional posterior density approach, which usually needs double integral.

3.4.3 Data Comparison

Figure 3.2 displays the predictive probability function of $N(t_n, s_2)$ for the data in table 2.1 with $t_n = 67,344$ and $s_2 = 80,000$. The line with circles represents the predictive probability distribution (3.7) based on the posterior density, while the line with stars represents the predictive probability distribution (3.11) based on a Bayesian estimate of intensity function v . In the time interval $(67,344, 80,000)$, the peak point shows that the most possible number of failures is three with probability of $1/4$.

Figure 3.3 displays the predictive probability function of $(n + 1)th$ failure time Z_1 given (3.8) and (3.12) for the data table 2.1 with $t_n = 67,344$ and $s_2 = 80,000$. Similarly as Figure 3.2, the line with circles represents the predictive probability distribution (3.8) based on the posterior density, while the line with stars represents the predictive probability distribution (3.12) based on a Bayesian estimate of intensity function v . The next failure will almost certainly occur within the next 14,000 seconds and within 2000 seconds there is a probability of about $1/3$ of a failure occurring.

Figure 3.2 and 3.3 show that we have very close results by using two different approaches. However, the one with the Bayesian estimate of intensity function simplify the computation and don't require numerical calculation.

3.5 Posterior Density for Some Parametric Functions

In this section, we shall derive posterior distributions of functions of (θ, μ) which are of particular interest. Those functions are system reliability, expected number of failures in some time interval and intensity function.

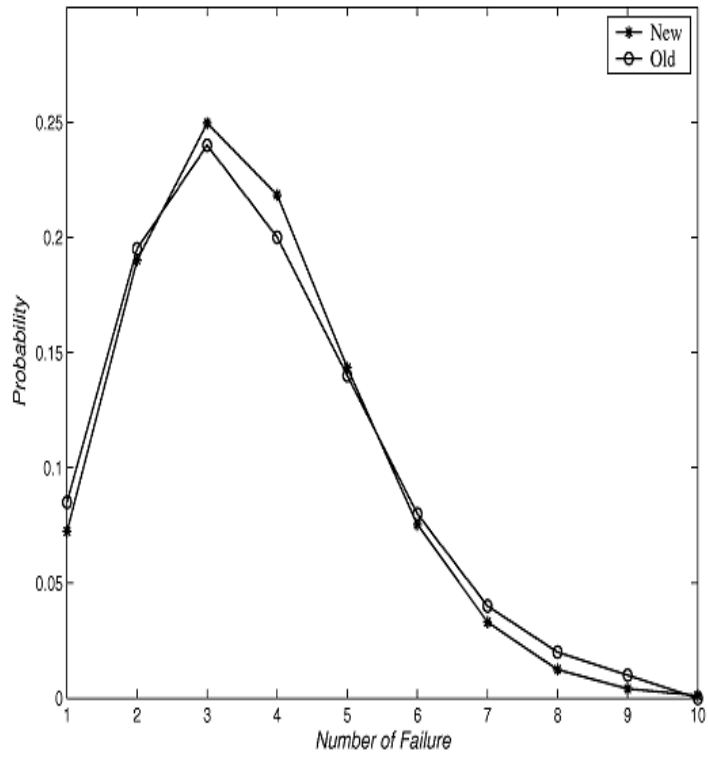


Figure 3.2: The Predictive Probability Function of $N(67,344 : 80,000)$

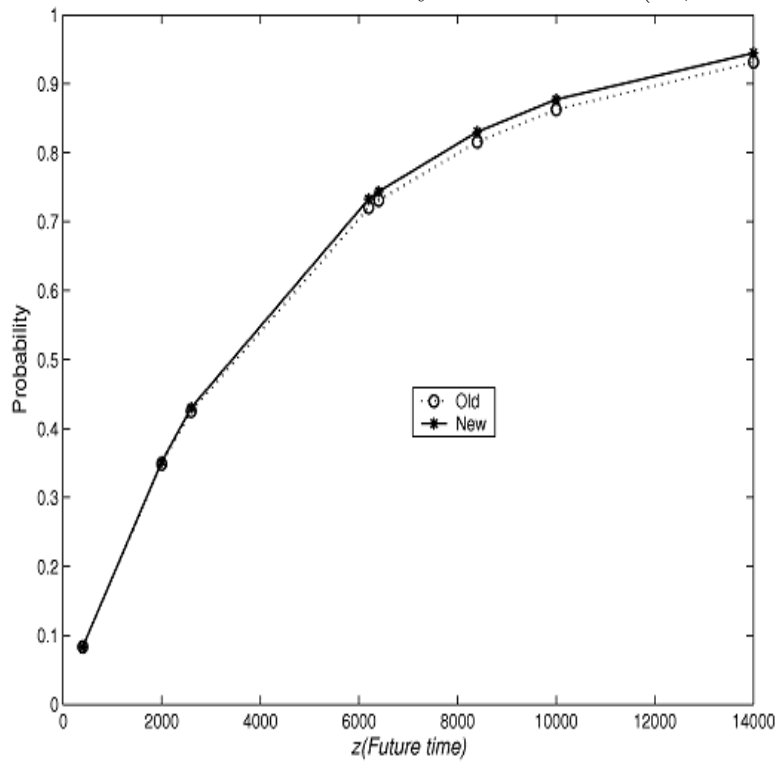


Figure 3.3: The Probability of $n + 1$ th Failure Time

3.5.1 Posterior Density for System Reliability

Recall that the reliability function is defined to be the probability of no failures over a specified time interval. For a given repairable system for which data has been collected, a high reliability over some future time of interest will affect decisions on replacement. Also, a high reliability for some period of interest in reliability growth may imply that it is worthwhile ending the development process. With this in mind, set

$$\begin{aligned} R &= R(y, s) = P[N(y, s) = 0] \\ &= \exp\left\{-e^{\frac{s-\mu}{\theta}} + e^{\frac{y-\mu}{\theta}}\right\}, \end{aligned}$$

which implies

$$\mu = \theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{-\ln r}\right)$$

The posterior cumulative distribution of reliability is

$$F(r) = \int_0^\infty Pr\left[\mu \leq \theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{-\ln r}\right) \mid \theta\right] \cdot \pi(\theta) d\theta.$$

Hence, the posterior pdf of reliability is

$$\begin{aligned} f(r \mid \vec{y}) &= \int_0^\infty f_\mu\left[\theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{-\ln r}\right)\right] \cdot \frac{d\left[\theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{-\ln r}\right)\right]}{dr} \cdot \pi(\theta) d\theta \\ &= \int_0^\infty \int_0^\infty C(\vec{y})(1/\theta)^{n+1} \cdot \exp\left\{\frac{\sum_1^n (y_i) - n\theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{-\ln r}\right)}{\theta}\right. \\ &\quad \left. - \exp\left(\frac{y_n - \theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{-\ln r}\right)}{\theta}\right)\right\} d\theta \cdot \frac{-\theta}{r \ln r} \cdot \pi(\theta) d\theta \\ &= (-r \ln r)^{-1} \cdot \frac{ny_n - \sum_1^n y_i}{n-2} \cdot \int_0^\infty C(\vec{y})(1/\theta)^{n+1} \exp\left\{\frac{\sum_1^n y_i}{\theta} + \exp\left(\frac{y}{\theta}\right) \cdot \frac{1}{(e^{s/\theta} - e^{y/\theta}) \ln r}\right\} \\ &\quad \cdot \left(\frac{-\ln r}{e^{s/\theta} - e^{y/\theta}}\right)^n d\theta \\ &= \frac{(-\ln r)^n}{-r \ln r} \int_0^\infty C(\vec{y}) \cdot (1/\theta)^{n+1} \cdot \exp\left(\frac{\sum_1^n y_i}{\theta}\right) r^{(e^{\frac{s-y}{\theta}} - 1)^{-1}} \cdot (e^{s/\theta} - e^{y/\theta})^{-n} d\theta. \end{aligned}$$

In a special case, we consider an equivalent system which is just beginning to operate; i.e., over the time interval $(0, e^s]$

$$f(r \mid \vec{y}) = \frac{(-\ln r)^n}{-r \ln r} \int_0^\infty C(\vec{y}) \cdot (1/\theta)^{n+1} \cdot \exp\left(\frac{\sum_1^n (y_i - s)}{\theta}\right) r^{(e^{\frac{y_n - s}{\theta}})} d\theta.$$

For $s = \ln(t_n)$, we obtain

$$f(r|\vec{y}) = \frac{(-\ln r)^{n-1} r^{-1}}{\Gamma(n-1)(ny_n - \sum_1^n y_i)}.$$

3.5.2 Posterior Density For the Expected Number of Failures in Some Time Interval

Similar to the consideration of the reliability function, we stop modifying system when the expected number of failures in some period of interest gets small enough in a reliability growth model. Here the expected number of failures in a given interval is of interest. The expected number of failures over time interval $(e^y, e^s]$ is also the mean value function in that given time, which is

$$m = m(y, s) = e^{\frac{s-\mu}{\theta}} - e^{\frac{y-\mu}{\theta}},$$

which implies

$$\mu = \theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{m}\right).$$

The posterior cdf of mean value is

$$F(m) = 1 - \int_0^\infty Pr[\mu \leq \theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{m}\right) | \theta] \cdot \pi(\theta) d\theta.$$

Hence, the posterior density of m is

$$\begin{aligned} f(m|\vec{y}) &= - \int_0^\infty f_\mu\left[\theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{m}\right)\right] \cdot \frac{d\left[\theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{m}\right)\right]}{dm} \cdot \pi(\theta) d\theta \\ &= - \int_0^\infty \int_0^\infty C(\vec{y}) \frac{1}{\theta^{n+1}} \cdot \exp\left\{\frac{\sum_1^n (y_i) - n\theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{m}\right)}{\theta}\right. \\ &\quad \left. - \exp\left(\frac{y_n - \theta \ln\left(\frac{e^{s/\theta} - e^{y/\theta}}{m}\right)}{\theta}\right)\right\} d\theta \cdot \frac{-\theta}{m} \cdot \pi(\theta) d\theta \\ &= m^{n-1} \cdot \frac{ny_n - \sum_1^n y_i}{n-2} \cdot \int_0^\infty C(\vec{y}) (1/\theta)^{n+1} \exp\left\{\frac{\sum_1^n y_i}{\theta} + \exp\left[-me^{\left(\frac{y-s}{\theta}-1\right)^{-1}}\right]\right\} \\ &\quad \cdot (e^{s/\theta} - e^{y/\theta})^{-n} d\theta. \end{aligned}$$

3.5.3 Posterior Density for Intensity Function

After the completion of the testing stage of a system, the system is supposed to have a constant failure rate (intensity function) through its useful life time. Thus we are interested in the intensity function value of the time stopping developing. The intensity function is

$$\nu = \nu(y) = \frac{1}{\theta} \exp\left(\frac{y - \mu}{\theta} - y\right),$$

which implies

$$\mu = y - \theta \ln(\nu\theta + y)$$

We can write cdf of ν as

$$F(\nu) = 1 - \int_0^\infty Pr[\mu \leq y - \theta \ln(\nu\theta + y) | \theta] \cdot \pi(\theta) d\theta.$$

Therefore, the posterior density for ν is

$$\begin{aligned} f(\nu | \vec{y}) &= - \int_0^\infty f_\mu[y - \theta \ln(\nu\theta + y)] \cdot \frac{d[y - \theta \ln(\nu\theta + y)]}{d\nu} \cdot \pi(\theta) d\theta \\ &= \int_0^\infty \int_0^\infty C(\vec{y})(1/\theta)^{n+1} \cdot \exp\left\{\frac{\sum_1^n (y_i) - ny}{\theta} + n \ln(\nu\theta + y) - (\nu\theta + y)\right\} d\theta \\ &\quad \cdot \frac{\theta^2}{\nu\theta + y} \pi(\theta) d\theta \\ &= \int_0^\infty C(\vec{y})(1/\theta)^{n+1} \cdot \exp\left[\frac{\sum_1^n (y_i) - ny}{\theta} - (\nu\theta + y)\right] \cdot (\nu\theta + y)^n d\theta \\ &\quad \cdot \int_0^\infty \frac{\theta^2}{\nu\theta + y} \pi(\theta) d\theta. \end{aligned}$$

We obtained posterior densities for reliability function, mean value function and intensity function. Therefore, we can find Bayesian point estimates using numerical calculation. If the integral can't be evaluated in a closed form, many mathematical software packages, such as *Mathematica*, *Maple*, *Matlab*, can do double integration.

3.6 Summary

We applied a logarithm transformation on the shape parameter α and a reciprocal transformation on the scale parameter β . Then we obtained a location parameter μ and

a scale parameter θ . This makes noninformative priors more appropriate. We developed Bayesian procedures for the power law process with failure truncation data based on θ and μ . Basic Bayesian results such as the posterior density, marginal posterior distribution of each parameter and Bayesian point estimates of parameters are obtained. Estimates of θ include the estimate under squared loss function and Bayesian MLE. Posterior variance for the estimate under squared loss and m th moment of θ are also derived. We applied Lindley's approximation to find estimation of μ under squared loss function.

Then we discussed the Bayesian prediction inferences for the expected number of failures and the future failure times. There are two ways to approach this. One is using posterior density; the other one is using a newly developed Bayesian estimation of intensity function. Predictive inference with the second approach greatly simplifies the calculations. We compare our results with the paper of Bar-lev, S.K. *et al.*(1992)[7] by using the data set from Musa(1979). Finally, we also derive posterior densities of system reliability, mean value function and intensity function.

Chapter 4

Empirical Bayes Analysis on the Power Law Process

In this chapter we focus on Empirical Bayes (EB) analysis on the Power Law Process by employing parametric EB priors and nonparametric EB priors. For the parametric EB priors, we apply two-hyperparameter natural conjugate prior and a more generalized three-hyperparameter natural conjugate prior. Those priors were stated in Huang and Bier (1998)[33]. Here we derive an Empirical Bayes procedure to estimate the natural conjugate priors. To compare with the previous chapter, when we completely know the prior, the approach is Bayesian. If we don't know the prior completely, we use Empirical Bayes when assuming parameters of prior are fixed but unknown. Since we have past experience about the parameters of the model, we can employ data to estimate the hyperparameters of priors, hence estimate the priors in parametric Empirical Bayes.

4.1 Parametric Empirical Bayes on the PLP

Parametric empirical Bayes procedures are easier to work with if the intensity function is parametrized as

$$\nu(t) = \eta\beta t^{\beta-1}, \quad t > 0.$$

Throughout most of this dissertation, we have used

$$\nu(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha}\right)^{\beta-1}, \quad \text{for } \alpha > 0, \beta > 0.$$

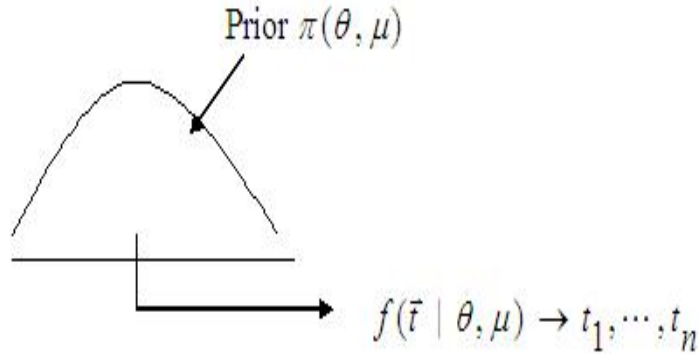


Figure 4.1: Bayesian Model for the PLP

The relationship between η and α is $\eta = \alpha^{-\beta}$.

If k systems are similar, but not identical, a parametric empirical Bayes (PEB) approach can be applied. We assume that the system parameters (in our case the (η_i, β_i) 's) are drawn from some prior distribution $\pi(\eta, \beta)$. Figure 4.1 displays Bayesian framework in the previous chapter. Figure 4.2 displays parametric empirical Bayes model for the PLP. We assume (η_i, β_i) 's make up a random sample selected from the prior. This assumption is reasonable if the k systems are made from the same manufacturing process. Here we employ natural conjugate priors with two and three hyperparameters.

4.1.1 Two Hyperparameters (a, m)

Assume there are k systems. Let \vec{t}_i denote the vector of failure times for system i , and let

$$T = [\vec{t}_1, \vec{t}_2, \dots, \vec{t}_k]$$

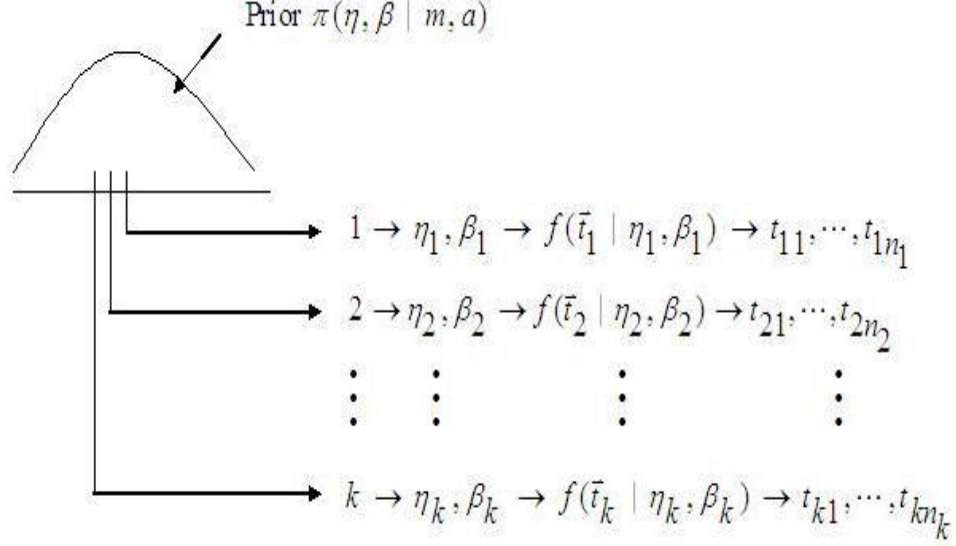


Figure 4.2: Parametric Empirical Bayes Model for the PLP

denote the two-dimensional array of failure times. The likelihood function of the first n_i failure times for system i is

$$L(\vec{t}_i | \eta, \beta) = L(t_{i1}, t_{i2}, \dots, t_{in_i} | \eta, \beta) = \eta^{n_i} \beta^{n_i} \left(\prod_{j=1}^{n_i} t_{ij} \right)^{\beta-1} \exp(-\eta t_{n_i}^{\beta}). \quad (4.1)$$

The natural conjugate prior distribution for the power law failure model is given by

$$\pi_0(\eta, \beta | m, a) = c^{-1} \eta^{m-1} \beta^{m-1} (\exp(-a) t_{n_i}^m)^{\beta-1} \exp(-\eta t_{n_i}^{\beta}), \quad (4.2)$$

here (m, a) are positive hyperparameters. c is a constant and t_{n_i} is a fixed truncated failure time for system i .

$$\begin{aligned} c &= \int \int \eta^{m-1} \beta^{m-1} (\exp(-a) t_{n_i}^m)^{\beta-1} \exp(-\eta t_{n_i}^{\beta}) d\eta d\beta \\ &= \Gamma^2(m) a^{-m} [\exp(-a) t_{n_i}^m]^{-1} \end{aligned} \quad (4.3)$$

Then the marginal distribution of \vec{t}_i given m, a is

$$\begin{aligned} m(\vec{t}_i | m, a) &= \int \int L(t_{i1}, t_{i2}, \dots, t_{in_i} | \eta, \beta) \pi_0(\eta, \beta | m, a) d\eta d\beta \\ &= c^{-1} \int \int \eta^{n_i+m-1} \beta^{n_i+m-1} [\exp(-a) t_{n_i}^m \prod_{j=1}^{n_i} t_{ij}]^{\beta-1} \exp(-2\eta t_{n_i}^{\beta}) d\eta d\beta \end{aligned}$$

$$\begin{aligned}
&= c^{-1} \int 2^{-(m+n_i)} \Gamma(n_i + m) t_{n_i}^{-(m+n_i)\beta} [\exp(-a) t_{n_i}^m \prod_{j=1}^{n_i} t_{ij}]^{\beta-1} \beta^{n_i+m-1} d\beta \\
&= \frac{c^{-1} 2^{-(m+n_i)} \Gamma^2(n_i + m) [a + n_i \ln(t_{n_i}) - \ln \prod_{j=1}^{n_i} (t_{ij})]^{-(n_i+m)}}{\exp(-a) t_{n_i}^m \prod_{j=1}^{n_i} t_{ij}} \\
&= \frac{2^{-(m+n_i)} \Gamma^2(n_i + m) [a + n_i \ln(t_{n_i}) - \ln \prod_{j=1}^{n_i} (t_{ij})]^{-(n_i+m)}}{\prod_{j=1}^{n_i} t_{ij} \Gamma^2(m) a^{-m}}. \tag{4.4}
\end{aligned}$$

Since K systems are independent, we have the marginal distribution of T given m and a is

$$\begin{aligned}
m(T|m, a) &= m(\vec{t}_1|m, a) m(\vec{t}_2|m, a) \cdots m(\vec{t}_k|m, a) \\
&= \prod_{i=1}^k \left\{ \frac{\Gamma^2(n_i + m) [a + n_i \ln(t_{n_i}) - \ln \prod_{j=1}^{n_i} (t_{ij})]^{-(n_i+m)}}{2^{m+n_i} \prod_{j=1}^{n_i} t_{ij} \Gamma^2(m) a^{-m}} \right\}. \tag{4.5}
\end{aligned}$$

In order to obtain MLEs of m and a , we need take a natural logarithm of (4.5)

$$\begin{aligned}
\ln[m(T|m, a)] &= \sum_{i=1}^k \left\{ - (m + n_i) \ln 2 + 2 \ln \Gamma(n_i + m) - (n_i + m) \ln [a + n_i \ln(t_{n_i}) \right. \\
&\quad \left. - \ln \prod_{j=1}^{n_i} (t_{ij})] - \ln \prod_{j=1}^{n_i} t_{ij} - 2 \ln \Gamma(m) + m \ln a \right\}. \tag{4.6}
\end{aligned}$$

Then taking the derivative with respect to a in (4.6) and set it equal to zero, we have

$$\sum_{i=1}^k \left[\frac{n_i + m}{a + n_i \ln(t_{n_i}) - \ln \prod_{j=1}^{n_i} t_{ij}} - \frac{m}{a} \right] = 0. \tag{4.7}$$

Similarly, taking the derivative with respect to m in (4.6) and set it equal to zero, we have

$$\sum_{i=1}^k \left[- \ln 2 + \frac{2\Gamma'(n_i + m)}{\Gamma(n_i + m)} - [a + n_i \ln(t_{n_i}) - \ln \prod_{j=1}^{n_i} (t_{ij})] - 2 \frac{\Gamma'(m)}{\Gamma(m)} + \ln a \right] = 0. \tag{4.8}$$

According to the following well known property of Gamma function:

$$-\frac{\Gamma'(z)}{\Gamma(z)} = \frac{1}{z} + \gamma + \sum_{i=1}^{\infty} \left(\frac{1}{n+z} - \frac{1}{n} \right).$$

Equation (4.8) is simplified to

$$2 \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{1}{m+j-1} + k \ln \frac{a}{2} - ka - \sum_{i=1}^k \left[n_i \ln(t_{n_i}) - \ln \prod_{j=1}^{n_i} (t_{ij}) \right] = 0 \tag{4.9}$$

MLEs of a and m can be obtained by solving equations (4.7) and (4.9) numerically. In general, the likelihood equations do not admit a closed form solution and a numerical method must be employed to approximate the MLEs of the hyperparameters (a, m) . In the special case that we have one system, that is $k = 1$, with observations (t_1, t_2, \dots, t_n) , we are able to obtain the estimate of a in a closed form from Bayesian maximum likelihood approach. The following results shall be considered as Bayesian inference.

$$\hat{a} = \hat{m} \ln t_n - \frac{\hat{m}}{n} \ln \prod_{i=1}^n t_i.$$

We can use Newton-Raphson method to get MLE for m .

The posterior distribution

$$\pi(\eta, \beta | \vec{t}, n, \hat{m}, \hat{a}) = \frac{1}{m(\vec{t} | \hat{m}, \hat{a})} c^{-1} \eta^{n+\hat{m}-1} \beta^{n+\hat{m}-1} [\exp(-\hat{a}) t_n^{\hat{m}} \prod_{i=1}^n t_i]^{\beta-1} \exp(-2\eta t_n^\beta). \quad (4.10)$$

Hence, the prior distribution and posterior distribution are from the same family, priors are natural conjugate priors. The posterior mean for η is

$$\begin{aligned} \tilde{\eta}_B &= E(\eta) = \int \int \eta \pi(\eta, \beta) d\beta d\eta \\ &= \frac{c^{-1} 2^{n+\hat{m}+1} \Gamma(n + \hat{m} + 1) \Gamma(\hat{m} + n) [\hat{a} - \ln \prod_{i=1}^n t_i + (n + 1) \ln t_n]^{-(n+\hat{m})}}{m(\vec{t} | \hat{m}, \hat{a}) [\exp(-\hat{a}) t_n^{\hat{m}} \prod_{i=1}^n t_i]} \\ &= \frac{\Gamma(n + \hat{m} + 1)}{2\Gamma(\hat{m} + n)} \left\{ \frac{\hat{a} + (n + 1) \ln t_n - \ln \prod_{i=1}^n t_i}{\hat{a} + n \ln t_n - \ln \prod_{i=1}^n t_i} \right\}^{-(n+\hat{m})}. \end{aligned} \quad (4.11)$$

The posterior mean for β is

$$\begin{aligned} \tilde{\beta}_B &= E(\beta) = \int \int \beta \pi(\eta, \beta) d\eta d\beta \\ &= \frac{c^{-1} 2^{n+\hat{m}} \Gamma(n + \hat{m} + 1) \Gamma(\hat{m} + n) [\hat{a} - \ln \prod_{i=1}^n t_i + n \ln t_n]^{-(n+\hat{m}+1)}}{m(\vec{t} | \hat{m}, \hat{a}) [\exp(-\hat{a}) t_n^{\hat{m}} \prod_{i=1}^n t_i]} \\ &= \frac{\Gamma(n + \hat{m} + 1)}{\Gamma(n + \hat{m}) (\hat{a} + n \ln t_n - \sum_{i=1}^n \ln t_i)}. \end{aligned} \quad (4.12)$$

4.1.2 Three Hyperparameters (a, m, y_m)

The situation is similar as in the previous subsection. But here we take a more general conjugate prior. Recall the likelihood function:

$$L(\vec{t}_i|\eta, \beta) = L(t_{i1}, t_{i2}, \dots, t_{in}|\eta, \beta) = \eta^{n_i} \beta^{n_i} \left(\prod_{j=1}^{n_i} t_{ij} \right)^{\beta-1} \exp(-\eta t_{n_i}^\beta).$$

The natural conjugate prior distribution for the power law failure model is given by

$$\pi_0(\eta, \beta|m, a, y_m) \propto \eta^{m-1} \beta^{m-1} \left(\prod_{i=1}^m y_i^m \right)^{\beta-1} \exp(-\eta y_m^\beta). \quad (4.13)$$

The parameters $y_1 \dots y_m$ can be interpreted as a pseudo-data set, where m is the number of failures and y_i is the time of the i th failure. For simplicity and without loss of generality, we can choose $\prod_{i=1}^m y_i^m = \exp(-a) y_m^m$. Thus the natural conjugate prior becomes

$$\pi_0(\eta, \beta|m, a, y_m) = c^{-1} \eta^{m-1} \beta^{m-1} (\exp(-a) y_m^m)^{\beta-1} \exp(-\eta y_m^\beta), \quad (4.14)$$

here (m, a, y_m) are positive hyperparameters. c is a constant and

$$\begin{aligned} c &= \int \int \eta^{m-1} \beta^{m-1} (\exp(-a) y_m^m)^{\beta-1} \exp(-\lambda y_m^\beta) d\eta d\lambda \\ &= \Gamma^2(m) a^{-m} [\exp(-a) y_m^m]^{-1}. \end{aligned} \quad (4.15)$$

Then the marginal distribution of \vec{t}_i given m, a, y_m is

$$\begin{aligned} m(\vec{t}_i|m, a, y_m) &= \int \int L(t_{i1}, t_{i2}, \dots, t_{in}|\eta, \beta) \pi_0(\eta, \beta|m, a, y_m) d\eta d\beta \\ &= c^{-1} \int \int \eta^{n_i+m-1} \beta^{n_i+m-1} [\exp(-a) y_m^m \prod_{i=1}^{n_i} t_{ni}]^{\beta-1} \exp[-\eta(t_{n_i}^\beta + y_m^\beta)] d\eta d\beta \\ &= \frac{1}{c} \int \Gamma(n_i + m) [t_{n_i}^\beta + y_m^\beta]^{-(m+n_i)} [\exp(-a) y_m^m \prod_{j=1}^{n_i} t_{ij}]^{\beta-1} \beta^{n_i+m-1} d\beta \end{aligned} \quad (4.16)$$

Since K systems are independent, the likelihood distribution of T given m and a is

$$m(T|m, a) = m(\vec{t}_1|m, a) m(\vec{t}_2|m, a) \cdots m(\vec{t}_k|m, a). \quad (4.17)$$

Again the likelihood equations do not admit a closed form solution and a numerical method must be employed to approximate the MLEs of (a, m, y_m) . However, suppose

we only have observations (t_1, t_2, \dots, t_n) from one system, which means we only have a random sample of size one (η, θ) from the prior $\pi(\eta, \beta|a, m, y_m)$, our inference shall be regarded as Bayesian maximum likelihood approach. Hence, for this special case, we have the posterior distribution of (η, β) is

$$\pi(\eta, \beta|\vec{t}, n, \hat{m}, \hat{a}, \hat{y}_m) = \frac{c^{-1}\eta^{n+\hat{m}-1}\beta^{n+\hat{m}-1}[\exp(-\hat{a})\hat{y}_m^{\hat{m}}\prod_{i=1}^n t_i]^{\beta-1}\exp(-\eta(\hat{y}_m^\beta + t_n^\beta))}{m(\vec{t}|\hat{m}, \hat{a}, \hat{y}_m)}. \quad (4.18)$$

The posterior mean for η is

$$\begin{aligned} E(\eta) &= \int \int \eta \pi(\eta, \beta) d\beta d\eta \\ &= \frac{c^{-1}\Gamma(n + \hat{m} + 1) \int \beta^{n+\hat{m}-1} [\exp(-\hat{a})\hat{y}_m^{\hat{m}}\prod_{i=1}^n t_n]^{\beta-1} (\hat{y}_m^\beta + t_n^\beta)^{-(n+\hat{m})} d\beta}{m(\vec{t}|\hat{m}, \hat{a}, \hat{y}_m)} \\ &= \frac{\Gamma(n + \hat{m} + 1) \int \beta^{n+\hat{m}-1} [\exp(-\hat{a})\hat{y}_m^{\hat{m}}\prod_{i=1}^n t_n]^{\beta-1} (\hat{y}_m^\beta + t_n^\beta)^{-(n+\hat{m}+1)} d\beta}{\Gamma(n + \hat{m}) \int \beta^{n+\hat{m}-1} [\exp(-\hat{a})\hat{y}_m^{\hat{m}}\prod_{i=1}^n t_n]^{\beta-1} (\hat{y}_m^\beta + t_n^\beta)^{-(n+\hat{m})} d\beta}. \end{aligned} \quad (4.19)$$

The posterior mean for β is

$$\begin{aligned} E(\beta) &= \int \int \beta \pi(\eta, \beta) d\eta d\beta \\ &= \frac{c^{-1}\Gamma(n + \hat{m}) \int \beta^{n+\hat{m}} [\exp(-\hat{a})\hat{y}_m^{\hat{m}}\prod_{i=1}^n t_n]^{\beta-1} (\hat{y}_m^\beta + t_n^\beta)^{-(n+\hat{m})} d\beta}{m(\vec{t}|\hat{m}, \hat{a}, \hat{y}_m)} \\ &= \frac{\int \beta^{n+\hat{m}} [\exp(-\hat{a})\hat{y}_m^{\hat{m}}\prod_{i=1}^n t_n]^{\beta-1} (\hat{y}_m^\beta + t_n^\beta)^{-(n+\hat{m})} d\beta}{\int \beta^{n+\hat{m}-1} [\exp(-\hat{a})\hat{y}_m^{\hat{m}}\prod_{i=1}^n t_n]^{\beta-1} (\hat{y}_m^\beta + t_n^\beta)^{-(n+\hat{m})} d\beta}. \end{aligned} \quad (4.20)$$

It should be addressed that the problem with parametric Emperical Bayes (PEB) is that we assume that the estimates of the prior parameters are the prior parameters themselves. The PEB approach does not account for uncertainty in the estimates of these hyper-parameters. Variation in these estimates would lead to more variation in the estimates of function of parameters, such as intensity and reliability etc.

4.1.3 Prior Plots

The joint prior density is given by (4.18). By take integral with respect to η ,

$$\begin{aligned} \pi_0(\beta|m, a, y_m) &= \int c^{-1}\eta^{m-1}\beta^{m-1}(\exp(-a)y_m^m)^{\beta-1}\exp(-\eta y_m^\beta)d\eta, \\ &= \frac{a^m \beta^{m-1} \exp(-a\beta)}{\Gamma(m)} \end{aligned} \quad (4.21)$$

marginal prior distribution of β has a Gamma distribution with mean m/a and variance m/a^2 . The conditional prior distribution of η given β is

$$\begin{aligned}\pi_0(\eta|\beta) &= \frac{\pi_0(\eta, \beta)}{\pi_0(\beta)} \\ &= \frac{c^{-1}\eta^{n+m-1}\beta^{n+m-1}[\exp(-a)y_m^m \prod_{i=1}^n t_i]^{\beta-1}\exp(-\eta(y_m^\beta + t_n^\beta))}{m(\vec{t}|m, a, y_m)} \frac{\Gamma(m)}{a^m \beta^{m-1} \exp(-a\beta)} \\ &= \frac{y_m^{\beta m} \eta^{m-1} \exp(-\eta y_m^\beta)}{\Gamma(m)}\end{aligned}$$

which is Gamma distribution with mean m/y_m^β and variance $m/y_m^{2\beta}$. Figure 4.1 to Figure 4.8 are the prior density plots given different value of hyperparameters.

1. Figure 4.3 and Figure 4.4 gives contour and three-dimension graph for $m = 2$ $a = 2$ $y_m = 2$. Only draw x-axis η in $[0,4]$ and y-axis β in $[0,4]$.

2. If only m increases, the other parameters are fixed, all the means and variance increase. The graph moves always from x-axis and y-axis, and more spread out. As shown in Figure 4.5 and Figure 4.6 when $a = 2$ $m = 6$ $y_m = 2$.

3. If only y_m increases, the other parameters are fixed, mean m/y_m^β decreases and variance $m/y_m^{2\beta}$ decreases the graph moves close to y -axis and more concentrated. As shown in figure 4.7 and figure 4.8 $m = 2$ $a = 2$ $y_m = 6$.

4. If only a increases, the other parameters are fixed, mean m/a and variance m/a^2 decrease. The graph moves close to x-axis and more concentrate. As shown in Figure 4.9 and Figure 4.10 when $a = 6$ $m = 2$ $y_m = 2$.

4.2 Nonparametric Prior on the Power Law Process

We now assume $\theta_j, j = 1, 2, \dots, m$ from m systems are drawn from a prior distribution $\pi_0(\theta)$. The goal of density estimation is to approximate the probability density function $\pi_0(\theta)$. Assume we have m independent, identically distributed observations $\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_m$ which are obtained from m systems using classical MLEs in the previous chapter.

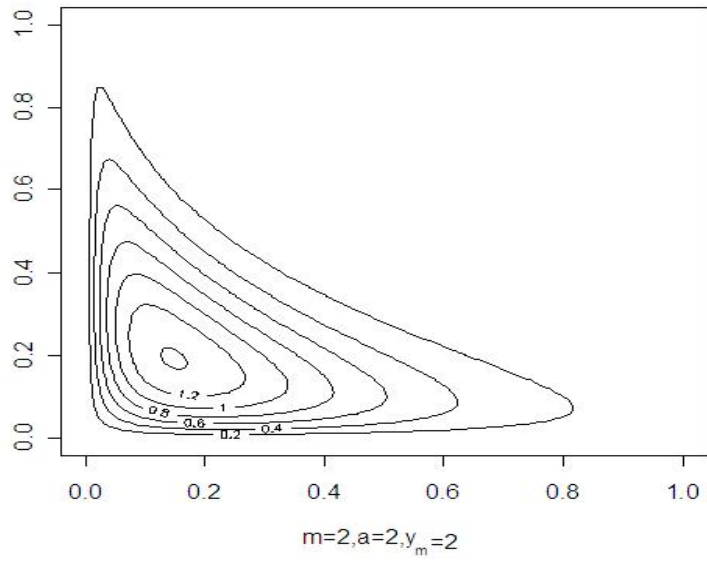


Figure 4.3: Prior Contour Plot $a = 2$ $m = 2$ $y_m = 2$

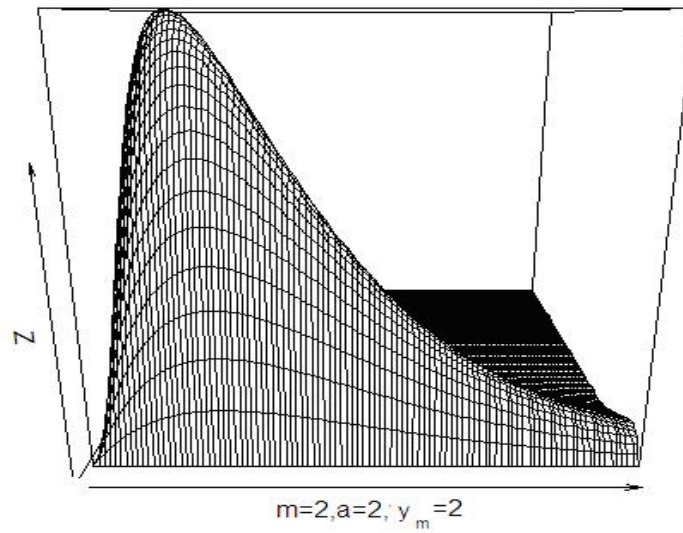


Figure 4.4: Prior Three-Dimension Plot $a = 2$ $m = 2$ $y_m = 2$

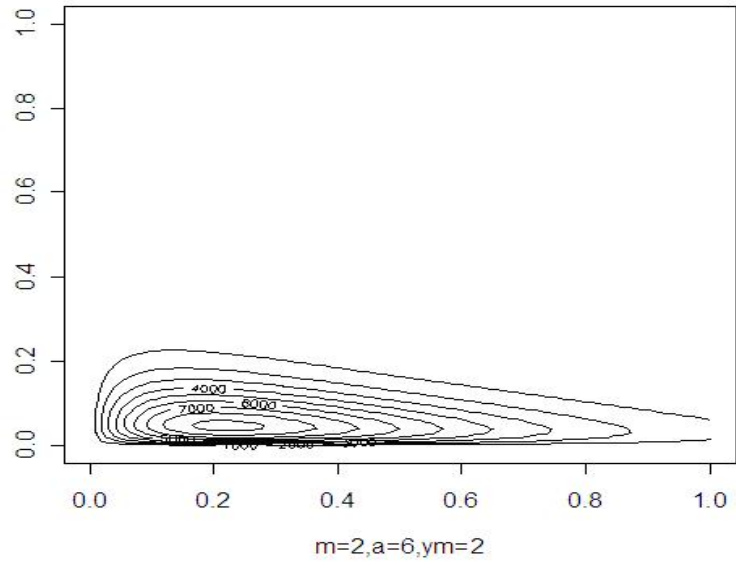


Figure 4.5: Prior Contour Plot $a = 2$ $m = 6$ $y_m = 2$

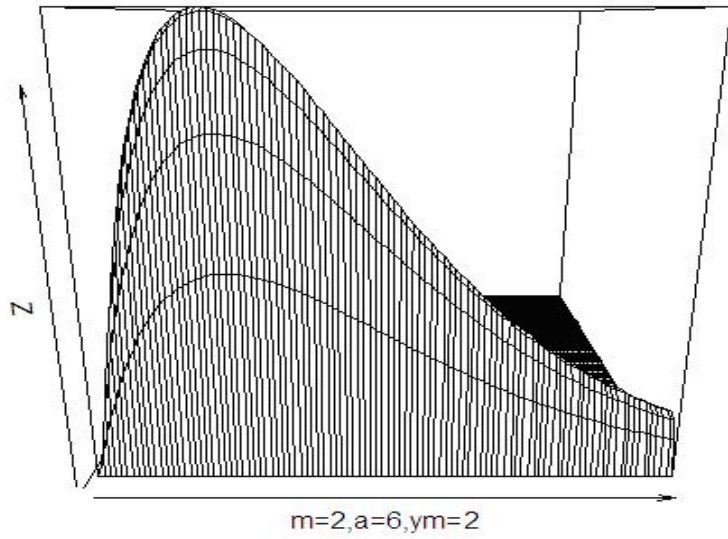


Figure 4.6: Prior Three Dimension Plot $a = 2$ $m = 6$ $y_m = 2$

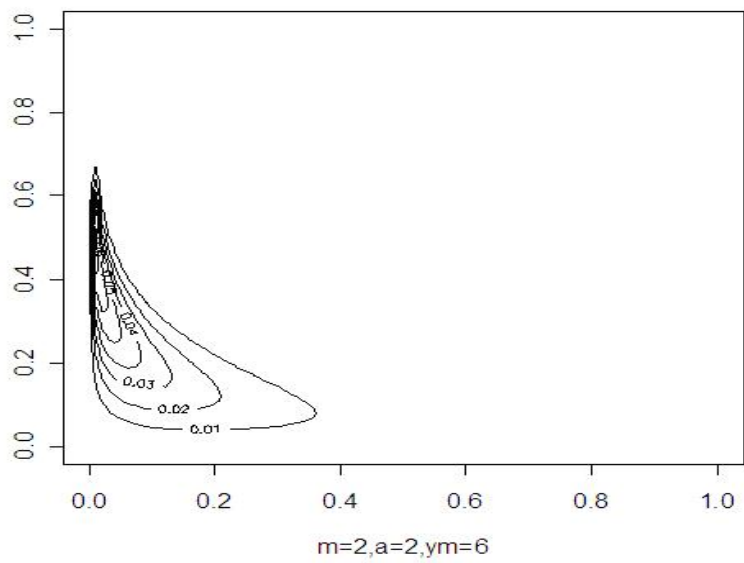


Figure 4.7: Prior Contour Plot $a = 2$ $m = 2$ $y_m = 6$

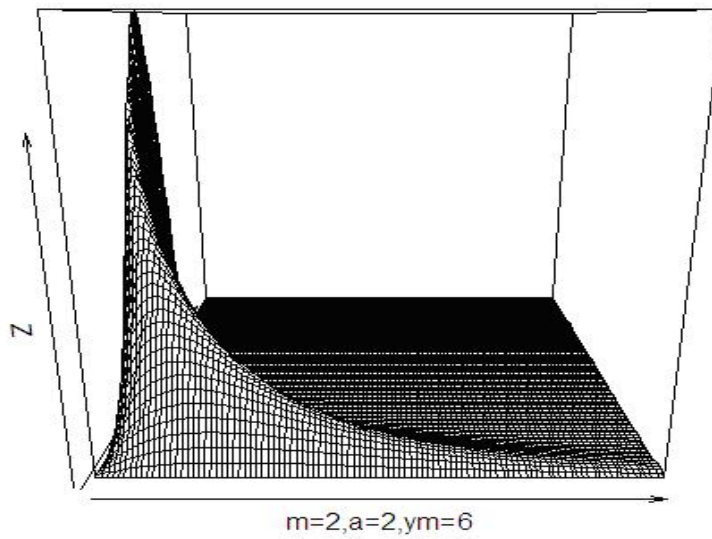


Figure 4.8: Prior Three Dimension Plot $a = 2$ $m = 2$ $y_m = 6$

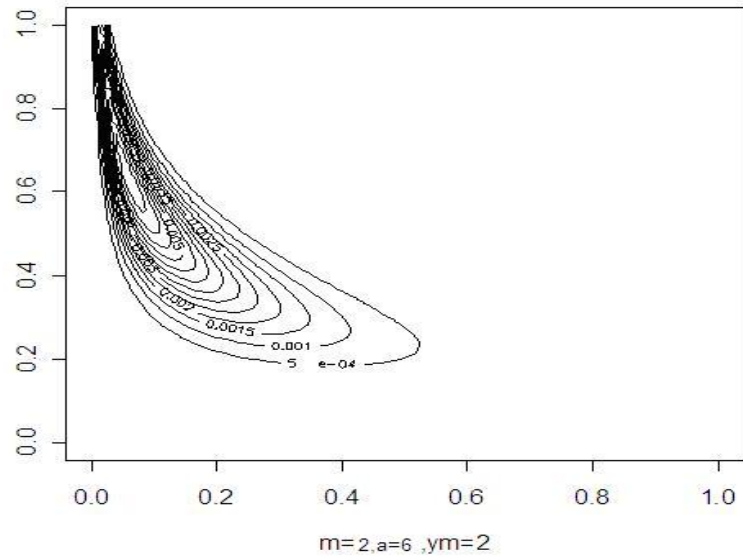


Figure 4.9: Prior Contour Plot $a = 6$ $m = 2$ $y_m = 2$

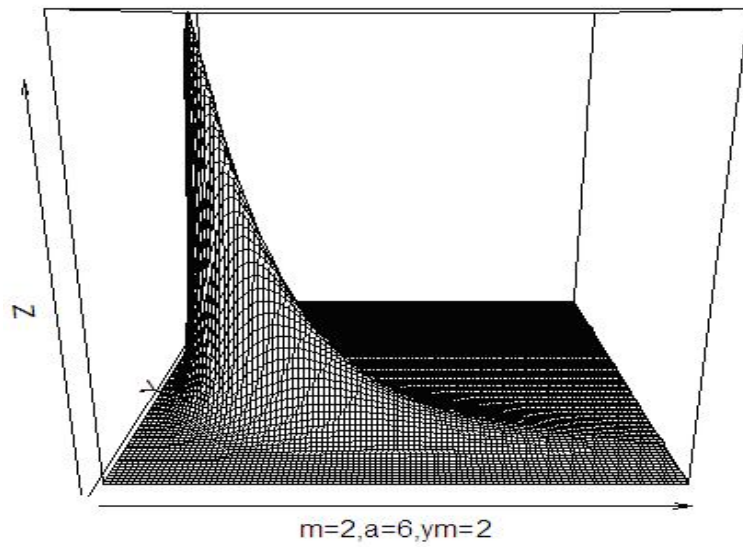


Figure 4.10: Prior Three-Dimension Plot $a = 6$ $m = 2$ $y_m = 2$

The kernel density estimator $\hat{\pi}_0(\theta)$ for the estimation of the density value $\pi_0(\theta)$ is defined as

$$\hat{\pi}_0(\theta) = \frac{1}{mh} \sum_{j=1}^m K\left(\frac{\hat{\theta}_j - \theta}{h}\right)$$

where $K(\bullet)$ denotes the kernel function, and h denotes the bandwidth, or the smoothing parameter. The bandwidth controls the amount of smoothing. If h is large, there is a lot of smoothing, and if h is small there is less smoothing. $\hat{\pi}_0(\theta)$ is a nonparametric probability density estimation to $\pi_0(\theta)$. A number of possible kernel functions are listed in the following table:

Table 4.1: Commonly Used Kernel Functions

Kernel	$K(u)$
Uniform	$\frac{1}{2}I(u \leq 1)$
Triangle	$(1 - u)I(u \leq 1)$
Epanechnikov	$\frac{3}{4}(1 - u^2)I(u \leq 1)$
Quartic	$\frac{15}{16}(1 - u^2)^2I(u \leq 1)$
Triweight	$\frac{35}{32}(1 - u^2)^3I(u \leq 1)$
Gaussian	$\frac{1}{\sqrt{2\pi}}exp(-\frac{1}{2}u^2)$
Cosinus	$\frac{\pi}{4} \cos(\frac{\pi}{2})I(u \leq 1)$

Kernel function has the following properties: (i) $K(u) = K(-u)$, (ii) $\int_{-\infty}^{\infty} K(u)du = 1$, (iii) $\int_{-\infty}^{\infty} uK(u)du = 0$ and (vi) $\int_{-\infty}^{\infty} u^2K(u)du \neq 0$.

To obtain the prior density $\hat{\pi}_0(\theta)$, we need to select a kernel function $K(\bullet)$ and bandwidth h . In our work, we employ the most common used kernel function: Gaussian kernel. It is differentiable everywhere and is given by

$$K(\theta) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\theta^2}$$

To determine an optimal kernel density and bandwidth, we need minimize the mean integrated squared error (MISE) which is defined as

$$MISE(K, h) = E \int_{-\infty}^{\infty} (\hat{\pi}_0(\theta) - \pi_0(\theta))^2 dx.$$

For a fixed Gaussian kernel, estimation of bandwidth h can be obtained by unimodel Silverman's method (See details in chapter 5 of [51]). To estimate nonparametric density for μ , we apply the same procedure as above. Hence, nonparametric kernel prior density for θ and

$$P(\theta) = \frac{1}{n\hat{h}_1} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\theta - \hat{\theta}_i}{\hat{h}_1}\right)^2\right\}$$

nonparametric kernel prior density for μ and

$$P(\mu) = \frac{1}{n\hat{h}_2} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\mu - \hat{\mu}_i}{\hat{h}_2}\right)^2\right\}.$$

This is the simplest idea that employs nonparametric Empirical Bayes on the Power Law Process. Further work can be done to address more optimal nonparametric priors.

4.3 Further Research

It is fundamental basis of Bayesian decision theory that statistical inference should start with the determination of three factors: the distribution family for the observations, the prior distribution for the parameters and the loss associated with decisions. Further work can be done to check the robustness of the priors we used, such as noninformative priors, natural conjugate EB priors and nonparametric Empirical Bayes priors. We can slightly change the prior and see what happens to the decision. Two commonly used measures are the range of posterior decision and comparing Bayes risk criteria.

4.4 Summary

We worked on Empirical Bayes (EB) analysis on the Power Law Process by employing parametric EB priors and nonparametric EB priors. For the parametric EB

priors, we apply two-hyperparameter natural conjugate prior and a more generalized three-hyperparameter natural conjugate prior. Those priors were mentioned in Huang and Bier (1998)[33]. Here we derive Empirical Bayes procedure to estimate the natural conjugate priors. We employed past experience to estimate priors through data. We considered a special case when we only have one system. Under that case, the analysis becomes a Bayesian Maximum Likelihood approach. Hence, we also showed some results from Bayesian perspective in this chapter. According to nonparametric EB priors, we have k estimates of parameter from k systems, then we construct a nonparametric prior with normal kernel function and an optimal bandwidth.

Chapter 5

Microarray Analysis: Normalization and Transformation of Probe-Level Data

This chapter provides an overview of microarray from statistics perspective. We give a description of the target data sets. We also summarize and compare several methods of transformation and normalization for probe-level data.

5.1 Overview of Microarray Analysis

DNA microarray technology is a tool for studying how large numbers of genes interact with each other and enables the simultaneous analysis of thousands of sequences of DNA for genetic and genomic research. Microarray technology has been developing rapidly over the last several years.

Statistical and data-analytic techniques are involved in all stages of microarray experimentation and analysis. A task map in microarray data analysis is given in Figure 5.1. This chapter covers the application of several existing methods in probe level analysis of oligonucleotide arrays. In the next chapter, which includes a more important issue, several classical and Bayesian statistical methods are applied to analyze differentially expressed genes on expression level data produced from Dr. Haura's laboratory.

To select differentially expressed genes across different conditions is the first level of gene expression analysis. The second level considers the terms such as common functionalities, interactions and co-regulation. Therefore clustering is an important issue. The third level aims to find the underlying regulatory regions and gene networks that

ultimately are responsible for the observed patterns.

5.2 Data Description

We obtain two sets of data from the experiments in Dr. Haura's laboratory. The first set consists of DNA probe-level data, which is the base data set obtained by scanning hybridized cDNA. The second set is gene expression level data. Gene expression level data are computed from probe level data. There are various methods developed for the gene expression index computation. (Irizarry et al.(2003); Lemon et al.(2002); Holder et al.(2001); Naef et al.(2001); Zhou and Abagyan (2002); Affymetrix Inc. (2001b); Zhang et al.(2002)) In the following section we will give a brief comparison of four main methods and corresponding software.

Here is an introduction of probe level data. It is the raw data before having gene expressions. The first type of probe is referred to a perfect match (PM). Each PM probe is paired with a mismatch (MM) probe. These two probes are referred to as a probe pair. Each gene expression is represented by 11-20 probe pairs as shown in Figure 5.2 and usually a value representing the average difference between PM and MM. The purpose of the MM probe design is to measure non-specific binding and background noise. After scanning the arrays hybridized to labeled RNA samples, intensity values PM_{ij} and MM_{ij} are recorded for arrays $i = 1, \dots, I$ and probe pairs $j = 1, \dots, J$ for any given probe set.

In our data set, the expression level data contain 22215 genes. This arranges the data set consisting of 22215 rows. We perform treatments on five samples and use another five samples as control. This yields dataset consisting of 10 columns. Those five control units are denoted by GFP (GFP protein) and five experimental units are denoted by STAT3 (Stat3: a member of the family of signal transducers and activators of transcription).

A scatterplot of experiment gene expression (STAT3) against control gene expres-

Data-analytic Tasks in Microarray Experimentation

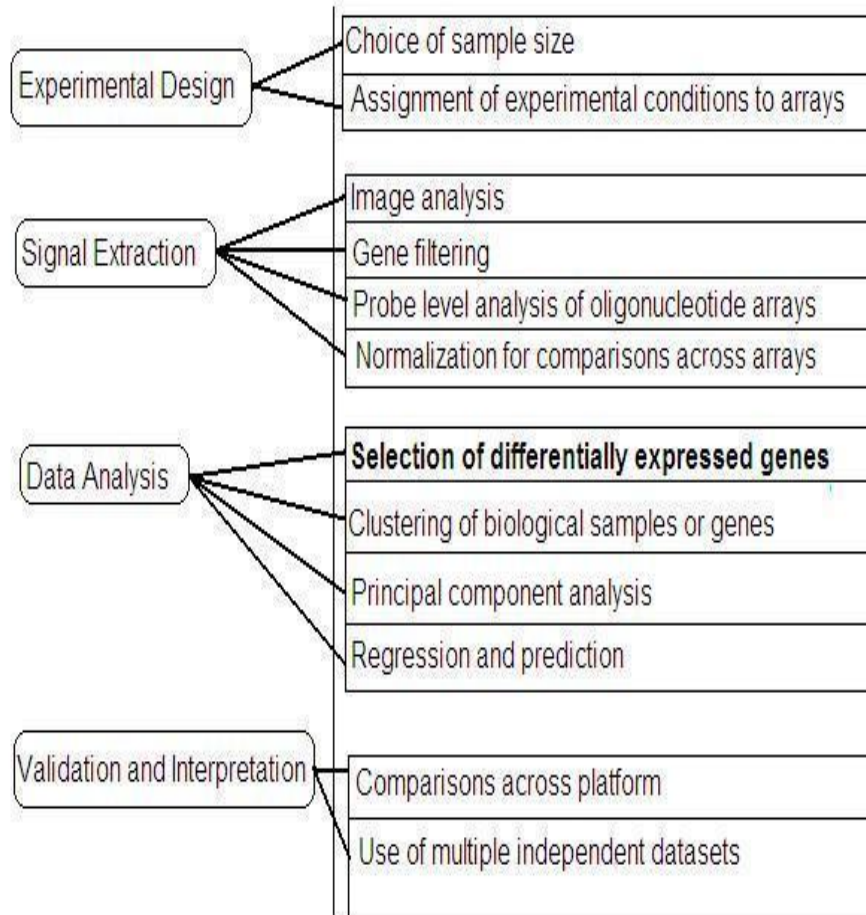


Figure 5.1: Data-analytic Tasks in Microarray Experimentation

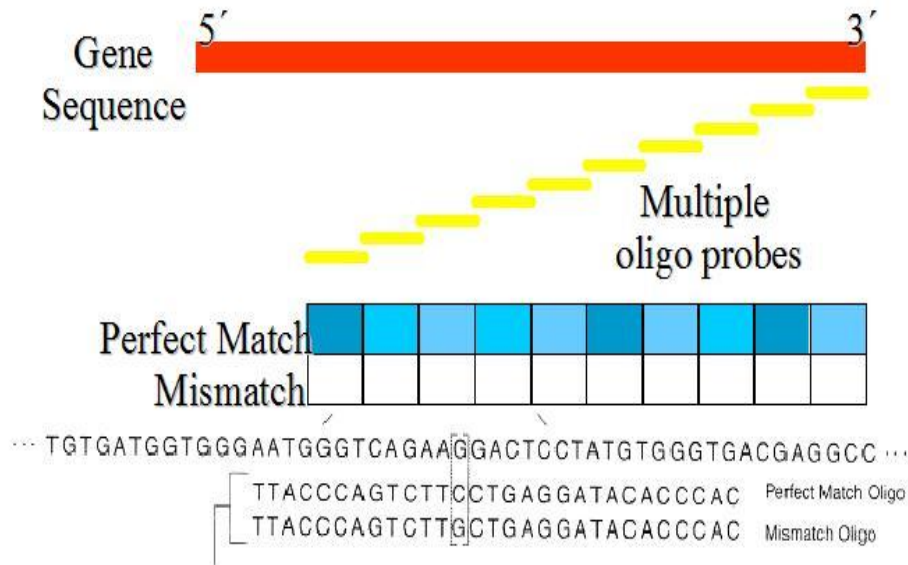


Figure 5.2: Signal Extraction in Probe Level Data

sion (GFP) is shown in Figure 5.3. It gives a brief idea about the value of gene expression. Each dot represents a gene. The regression is supposed to be $y = x$ based on the assumption that up-regulated genes and down-regulated genes with similar average intensity roughly canceled out or otherwise most genes remain unchanged. This assumption is usually true in large genome studies.(Dudoit et al.(2002)[25])

5.3 Probe-Level Analysis of Oligonucleotide Arrays

An important step in microarray analysis is the normalization of raw data. For Affymetrix gene chips, summarizing 11 to 20 probe pairs into one measure of expression is an essential step. The normalization aims to account for system technical difference in measurement process and control for many experimental sources of variability. Measurement differences consistently between chips are due to image analysis (identifying and quantifying each spot on the array), different amount of RNA, hybridization conditions

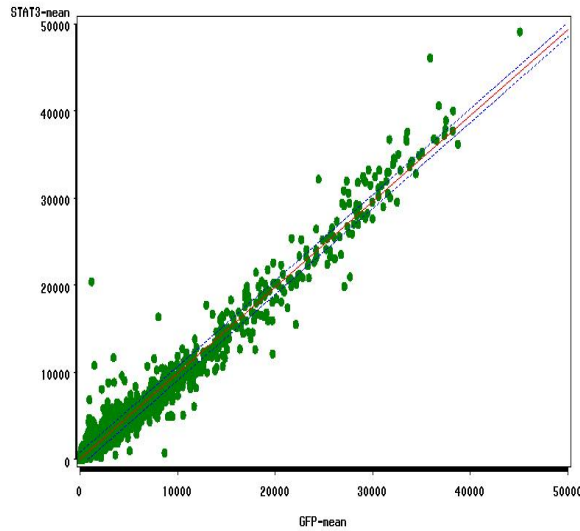


Figure 5.3: STAT3 v.s. GFP Mean Value: $STAT3 = 0.98745 \cdot GFP$

(temperature, time, mixing, etc) and scanner setting (Laser and detector, chemistry of the fluorescent label) etc.

Different approaches - all underlying some specific model assumptions - have been proposed. Two key elements should be specified in each approach: mathematical model and normalization method. We will discuss a standard method used in the Affymetrix Microarray Suite 5.0 software (MAS 5.0) with comparison to other three preprocessing algorithms: the robust multichip analysis (RMA); model based expression index (MBEI); a variance stabilization method (VSN). RMA and MBEI(dChip) are available within the Bioconductor project.

5.3.1 MAS 5.0 – Microarray Analysis Suite Version 5.0

Microarray Analysis Suit Version 5.0 (MAS 5.0) is produced by Affymetrix Inc. 2002. H. Lee Moffitt Cancer Center is applying this package to extract signals from scanning image. Here is an introduction about this approach. The mathematical model

is

$$Signal = TukeyBiweight\{\log_2(PM_j - MM_j^*)\}$$

for probe pair j .

MM_j^* is an adjusted MM_j that is never bigger than PM_j . Tukey Biweight is a robust average procedure with weights. The mean is calculated to identify center of data. Distance of each data point from the mean is calculated. The distance determines how each value is weighted in the average, i.e. outlier far from the median contribute little to the average.

MAS 5.0 offers only a global normalization procedure for the summarized probe sets. It adjusts the trimmed mean signal to a specified target signal value, in some case 500. Expression measures for each probe set are calculated with Tukey Biweight before normalization. MAS 5.0 assigns to each probe set an expression call and also offers the possibility of performing data scaling.

The drawbacks for MAS 5.0 depend on two facts. Average of different probes isn't really meaningful since probes have intrinsically different hybridization characteristics. The MAS 5.0 method doesn't learn based on cross-chip performance of individual probes.

5.3.2 MBEI – Model-Based Expression Index (dChip 2001)

MBEI accounts for individual probe-specific effects, automatic detection and handling of outliers and image artifacts. dChip is a software package produced by Li and Wong(2001) associated with Model-Based Expression Index method. This model is based on the observation that the variation of a specific probe across multi arrays could be considerably smaller than the variance across probes in a probe set and some probes are outliers. To take this into account, the following multiplicative model was proposed as

$$PM_{ij} - MM_{ij} = \theta_i \phi_j + e_{ij},$$

which indicates a strong probe affinity effect (ϕ_j). Distribution of error (e_{ij}) is assumed to be independent of signal strength. Furthermore, dChip allows assessment of a standard error (SE) for each probe set intensity measure, which is an indicator of hybridization quality to the probe set. Standard errors of ϕ_j (probe pair j) are useful for discarding probe sets with low hybridization quality. Standard errors of θ_i (array i) are used to identify array outliers.

Normalization methods in MBEI are invariant set normalization, quantile normalization and cyclic Lowess. For the invariant set normalization, a set of non-differentially expressed genes are selected by their invariant ranks of the probe intensities. Those genes can be regarded as baselines. The invariant set normalization produces a better fitting of the replicates comparing to the MAS 5.0 scaling procedure. Lowess technique is to apply a nonlinear curve to the scatter plots of the probe pair differences of a baseline array against all the other arrays and then force the curve to the line $y = x$.

5.3.3 RMA – Robust Multichip Analysis

A log scale linear additive model Robust Multichip Analysis (RMA) was proposed by Bolstad, Irizarry, Speed(2002). This method analyzes data for a set of chips using only PM and ignoring MM. The mathematical model is

$$\log(PM_{ij}) = \log(\theta_i) + \log(\phi_j).$$

A robust linear fitting procedure, such as median polish, was used to estimate the log scale expression values θ_i . In practice, $\log(PM_{ij} - BG) = \log(a_i) + \log(b_j) + \log(e_{ij})$. Signal $\log(PM_{ij})$ represents the transformation that background corrects and normalizes. Thus background value is important here. Recent results suggest that subtracting MM as a way of correcting for non-specific binding is not always appropriate. Unadjusted MM value may add more noise.

Normalization methods include quantile normalization and curve fitting normalization. The quantile method tries to make the same distribution of probe intensities for each array in a set of arrays. The method is bound to the idea that a quantile-quantile plot shows that the distribution of two data vectors is the same if the plot is a straight diagonal. The idea can be extended to n dimensions.

5.3.4 VSN – Variance Stabilization of Network

VSN is a normalization procedure produced by Huber *etal*(2002)[34] and also a method to preprocess DNA microarray expression data. In probe level data analysis, VSN uses the same mathematical model as in RMA except for the normalization. And normalization method is variance stabilizing transformations.

As the name states, variance stabilization transformation removes the dependence of the variance on the total intensity. This gives genes with higher intensities an equal chance of being ranked high as genes with lower intensity.

5.3.5 Comparison

To compare the probe level transformation and normalization methods, several standards should be considered, such as precision, consistency, specificity, sensitivity and accuracy. Precision means the reproducibility of measurement, as estimated by standard error across replicate chips. Specificity means the proportion of the signal that originates from the intended transcript (i.e. cross hybridization). Sensitivity gives lowest transcript concentration for an acceptable accuracy. Accuracy measures the distance of measurement to true value.

Li & Wong demonstrated that the multiplicative model has a more sensible model to analyze data from high density oligonucleotide array experiments compared to MAS

5.0. MBEI(dchip) also is more suitable for any further analysis than MAS 5.0 estimation does a reasonable job on probe-set that are bright. dChip and RMA does a better job on genes that are less abundant.

According to Irizarry *et al*, RMA has a better precision than MAS 5.0 and dChip based on higher squared coefficient correlation, especially for low expression levels. Concerning in the amount of true positives identified using spiked-in experiments, RMA performs slightly better than dChip, but much better than MAS 5.0. On the basis of published data, RMA also shows better sensitivity and specificity with respect to dChip and MAS 5.0. The advantage of RMA and VSN are two-fold: first, we are able to detect more of the spike-in genes while getting less false positives; secondly, the resulting data is easier to analyze. The strong intensity dependency of MAS 5.0 data has disappeared.

RMA and VSN perform similar on some data set. However, quantile normalization in RMA performs faster. Log transformation in RMA is more interpretable than arcsine transformation in VSN. At the moment RMA appears to be the best method available. However, it is also necessary to check model assumption for any given data.

5.4 Summary

Microarray analysis is a fairly new research area and just developed in past few years. Here we first gave a structure how statistical techniques are involved in all stages of microarray analysis. Then we introduced how the real data set that we are analyzing look like. We summarized and compared several methods in transformation and normalization of probe level data. It is treated as low level analysis in microarray and the results are usually obtained by existing software packages. Following to this step, we will have gene expression data, which will be analyzed in next chapter.

Chapter 6

Statistical Methods of Selecting Differentially Expressed Genes

In this chapter, the object is to find differentially expressed genes in 22,215 genes. The data are nonpaired five-control (GFP protein) and five-experiment (STAT3) gene expression. Both classical and Bayesian methods are applied. Classical methods include fold change, T-test, Wilcoxon Rank-sum test, Local Z-score (Chen,Z.[17]) and SAM(Storey 2002). Empirical Bayes methods include EBarrays(Newton, 2002[42]), LIMMA(Smyth G.K.(2003)[55]) and Cybor-T (Baldi and Long(2001)[5]). We mainly discuss two classical methods SAM (Storey 2002), local Z-score (Z. Chen[17]) and one parametric Empirical Bayes method (Newton, 2002). Several partial lists of differentially expressed genes are shown. Classical method intends to control false discovery rate, while Empirical Bayes method EBarrays aims to classify genes by expression patterns using posterior probability.

6.1 Select Differentially Regulated Genes Using Classical Statistics Methods

6.1.1 Fold Change Method

Fold change method is the simplest and most intuitive approach. However, the fold threshold is chosen arbitrarily. We may get too many or too few genes. Usually variance of gene expression data in low intensity is large, in high intensity is small. Figure 6.1 is a

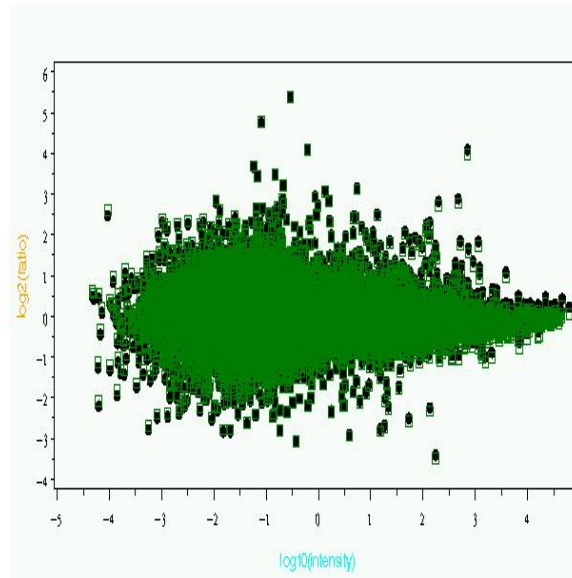


Figure 6.1: Lowess Regression for Five Experiments

scatter plot of Log-ratio against log-intensity while vertical axis represents

$$\text{Log}(\text{ratio}) = \log_2(\text{STAT3}/\text{GFP})$$

and horizontal axis represents

$$\text{Log}(\text{intensity}) = \log_{10}(\text{STAT3} * \text{GFP}).$$

It has a funnel shape. By using fold change method, this leads to high false positives at the low intensity end and missing true positives at the high intensity. To improve the sensitivity, local Z-score will be illustrated later.(Z. Chen[17])

In microarray analysis, we have large set of genes. Before the gene profiles of RNA samples can be analyzed and interpreted, the GFP and STAT3 intensities must be normalized relative to one another so that the STAT3/GFP ratio provides an unbiased representation of relative expression. Per-chip normalization is essentially a type of scaling to adjust the total or average intensity of each array. Per-gene normalization compares the results for a single gene across all the samples.

Since most genes will not change, Figure 6.1 is supposed to center at zero. Based on this assumption, normalization is necessary to balance the expression intensities so that meaningful biological comparisons can be made. The following are two normalization methods concerning gene-expression data from a single array hybridization (reviewed by John Quackenbush(2002)[46]). Note that Log-ratio for each gene is denoted by T_i , $T_i = \log_2(STAT3_i/GFP_i)$ and Log-intensity of each gene is denoted by A_i , $A_i = \log_{10}(GFP_i * STAT3_i)$

- Total intensity normalization

Let $T'_i = T_i - \text{Log}_2(Ntotal)$ where $Ntotal = \Sigma Stat3 / \Sigma GFP$. Thus, mean of T'_i is equal to zero. This is equivalent to subtracting a constant from the logarithm of the expression ratio.

- Normalization using regression techniques

A basic normalization method is print-tip LOWESS normalization. LOWESS stands for LOcally WEighted polynomial regreSSion(Dudoit et al.(2002)). We set $y_i = \text{Log}_2(Stat3_i/GFP_i)$ and $x_i = \text{Log}_{10}(Stat3_i * GFP_i)$ for each gene. Then we make a regression such that $y_i = m * x_i + b$, obtain an estimate $\hat{y}(x_i)$. Then use this estimate to plot scatter graph : $T'_i = T_i - \hat{y}(x_i)$ on A , where T_i and A_i are defined as above. T'_i will be brought to be centered at zero by the regression line.

From figure 6.1, it is obvious that normalization is not a key issue in our data since Log-ratios in our data have centered at zero. The slope of regression line is approximately zero so that there is only a slight change after normalization.

6.1.2 T-test

A univariate statistical test T-test is used to select differentially expressed genes. In a T -test, the empirical means m_c and m_t and variances s_c^2 and s_t^2 are used to calculate

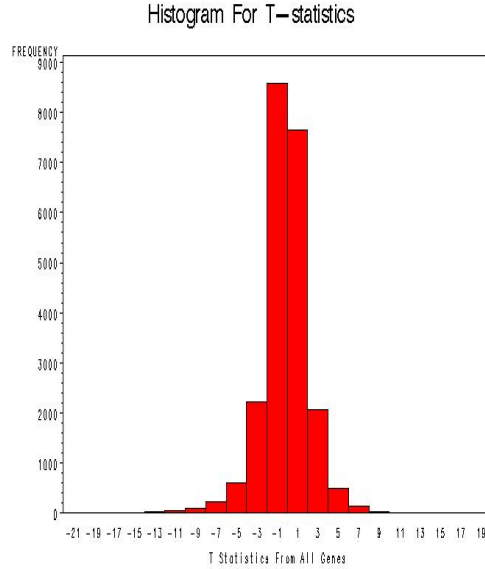


Figure 6.2: Histogram of T-statistics

the normalized distance between two populations in the form:

$$t = (m_c - m_t) / \sqrt{\frac{s_c^2}{n_c} + \frac{s_t^2}{n_t}}. \quad (6.1)$$

Here, for each population, $m = \Sigma x_i / n$ and $s^2 = \Sigma (x_i - m)^2 / (n - 1)$ are the estimates for the mean and standard deviation. It is well known that t follows approximately a Student distribution, with

$$f = \frac{[(s_c^2/n_c) + (s_t^2/n_t)]^2}{(s_c^2/n_c)^2/(n_c - 1) + (s_t^2/n_t)^2/(n_t - 1)} \quad (6.2)$$

degrees of freedom. When t exceeds a certain threshold depending on the confidence level selected, the two populations are considered to be different. Because in the t -test the distance between the population means is normalized by the empirical standard deviations, this has the potential for addressing some of the shortcomings of the simple fixed fold-threshold approach. The fundamental problem with the t -test for array data is that the repetition number n_c and/or n_t are often small since experiments remain costly or tedious to repeat. Small populations of size $n = 1, 2$, or 3 are still very common and lead to poor estimates of the variance. Thus a better framework is needed to address

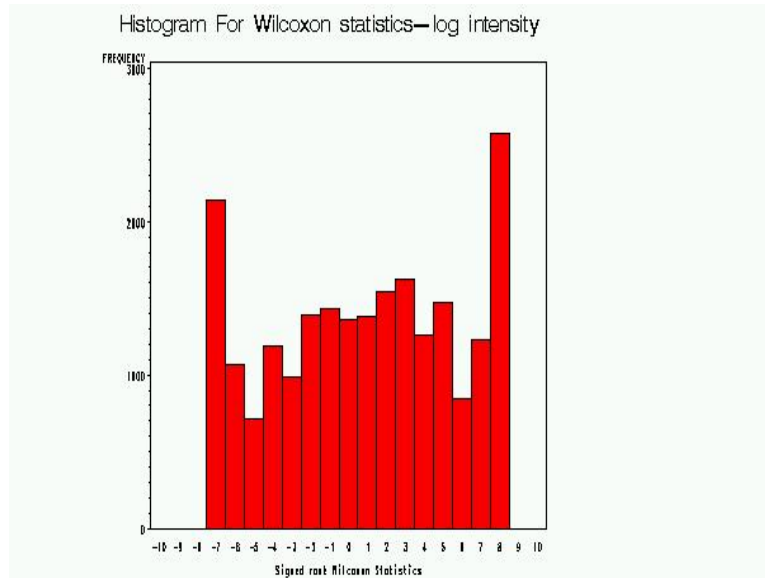


Figure 6.3: Histogram of Log-expression Wilcoxon Statistics

these shortcomings. Later we will describe several Bayesian probabilistic frameworks for array data, which can effectively address the problem of poor replicates.

Figure 6.2 gives a histogram of 22215 genes' T-statistics. From the graph, we can conclude that most gene expression do not change since T-statistics is concentrated at the center.

6.1.3 Wilcoxon Rank-Sum Test

Wilcoxon Rank-Sum test is a nonparametric test for two independent samples and is equivalent to Mann-Whitney U test . The statistic T_j is concerned about a difference in means, where Wilcoxon Rank-Sum statistic is more concerned about general distribution. For example, if all five experiments expression value is slightly greater than five controls expression value by chance, it will give a significant statistic value when the truth is not. Thus T-statistics is more powerful. Figure 6.3 gives a histogram of Wilcoxon statistics for all the genes.

6.1.4 SAM

SAM stands for Significance Analysis of Microarrays and is an outperformed method in identifying differentially expressed genes in DNA microarrays (Newton,2002[42]). In this section, we give a brief context how SAM works . The result table and plot for our data will also be interpreted.

Statistically speaking, identifying differentially expressed genes is a multiple hypothesis testing which tests all genes simultaneously and decides which genes are differentially expressed. There are four key steps in SAM as following:

STEP 1. A Statistic is formed for each gene by

$$d_j = \frac{\bar{x}_{j2} - \bar{x}_{j1}}{s_j + s_o} \text{ for gene } j.$$

where \bar{x}_{j1} and \bar{x}_{j2} are the average gene expression for gene j under control and experiment respectively. s_j is the pooled standard deviation for gene j . A small positive constant s_o is added to the denominator in order to ensure that the variance of d_j is independent of gene expression. The value for s_o was chosen to minimize the coefficient of variation of d_j , which is computed as a function of s_j in moving windows across the data.

STEP 2. Calculate null distribution for statistics. Each gene has a null distribution.

The null distribution can most easily be calculated by permutating the group labels. For example, we label our data as (1,1,1,1,1,2,2,2,2,2). To assess null distribution, we do a random permutation of the sample labels and recompute the statistics and count how many exceed a threshold, say, ± 2 . Redo this, say, 200 times. We can find the average number of genes exceeding ± 2 and use this number to estimate false discovery rate(FDR) in step 4.

STEP 3. Choose rejection regions and compare the statistics from observation to statistics from null distribution. $d(j)$ is the order statistics for d_j such that $d(1) \leq d(2) \cdots \leq d(J)$. $\bar{d}(j)$ is the estimate of the expected order statistics from K permutations. K can be chosen by user, for example, 200 in our data. Δ is a threshold chosen by user in software,

for example, we use 0.53245 in our data. We define reject regions as:

$$d(j) - \bar{d}(j) \geq \Delta \text{ or } d(j) - \bar{d}(j) \leq -\Delta.$$

STEP 4. Find the estimate of False Discovery Rate (FDR). False Discovery Rate is the error rate that we call truly unchanged genes differentially expressed. The FDR was proposed by Benjamini and Hochberg (1995). An estimate of FDR (Storey(2002)[56]) is

$$F\hat{D}R(\Delta) = \frac{R^0(\Delta)}{R(\Delta)} \times \hat{\pi}_0(\Delta')$$

where $R^0(\Delta)$ is the average number of significant genes from K random permutation of labels, which can be interpreted as average number of false discovered genes.

$$R^0(\Delta) = \frac{1}{K} \sum_1^K \#\{d_j : d_j \leq t_1(\Delta) \text{ or } d_j \geq t_2(\Delta)\}.$$

$R(\Delta)$ is the number of significant genes we discovered based on a threshold Δ .

$$R(\Delta) = \#\{d_j : d_j \leq t_1(\Delta) \text{ or } d_j \geq t_2(\Delta)\}$$

and $\hat{\pi}_0(\Delta')$ is an estimate of the overall proportion of true null hypothesis and depends on another threshold Δ' .

$$\hat{\pi}_0(\Delta') = \frac{J - R(\Delta')}{J - R_0(\Delta')}.$$

Table 6.1 displays a significant gene list by employing VSN normalized gene expression. Figure 6.4 displays a typical result plot by using SAM package. We input our data type as two class and unpaired data. Since we use VSN normalized gene expression, data has been in log scale. Number of permutations is set to one hundred to calculate the d-statistic under null hypothesis. The critical value Δ is adjusted to 2.06105 in order to obtain a reasonable false discovery rate to 0.05747. Score (d) is the statistic value from the observations. Numerator of d , denominator ($s + s_0$) of d and fold change are also given. q-value is the lowest pFDR (positive false discovery rate) at which the gene is called significant. It is similar to p-value, but interpreted as the probability that a false positive given its statistic is as or more extreme than the observed statistic.

Significant: 1033
Median # false significant: 0.59364

SAM Plot

Delta 2.06105
Fold Change

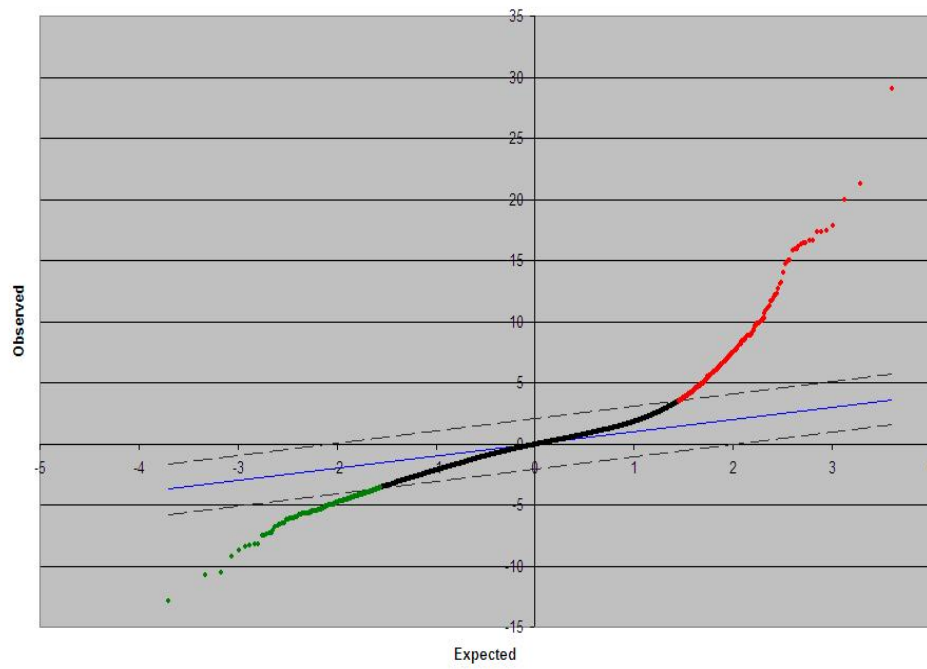


Figure 6.4: SAM Plot

Significant Genes List							
Input Parameters							
Imputation Engine				10-Nearest Neighbor Imputer			
Data Type				Two class, unpaired data			
Data in log scale?				True			
Number of Permutations				100			
Blocked Permutation?				False			
RNG Seed				37571352			
(Delta, Fold Change)				(2.06105,)			
Computed Quantities							
Computed Exchangeability Factor S0				0.024499213			
S0 percentile				0.02			
False Significant Number (Median, 90%				(0.59364,2.37455)			
False Discovery Rate (Median, 90%				(0.05747,0.22987)			
PioHat				0.59346			
578 Positive Significant Genes							
Row	Gene Name	Gene ID	Score	Numerator	Denominator	Fold Change	q-value
8487	208992-s-at	gb: BC000627.1	29.0263	3.5768	0.1232	12.0184	0.0398
4516	204988-at	gb: NM_005141.1	21.2830	1.8281	0.0858	3.5462	0.0398
5178	205650-s-at	gb:NM.021871.1	19.9642	2.1003	0.1052	4.2850	0.0398
21449	222088_s_at	gb:AA778684	17.8435	1.5972	0.0895	3.0238	0.0398
15609	216236_s_at	gb:AL110298.1	17.4420	1.3602	0.0780	2.5724	0.0398
18977	219612_s_at	gb:NM.000509.3	17.3390	2.4297	0.1401	5.3553	0.0398
2289	202760_s_at	gb:NM.007203.1	17.3031	1.5197	0.0878	2.8728	0.0398
5177	205649_s_at	gb:NM.000508.2	16.6140	1.5350	0.0924	2.8874	0.0398
3032	203504_s_at	gb:NM.005502.1	16.5730	1.4455	0.0872	2.7329	0.0398
18520	219155_at	gb:NM.012417.1	16.4220	0.7138	0.0435	1.6401	0.0398

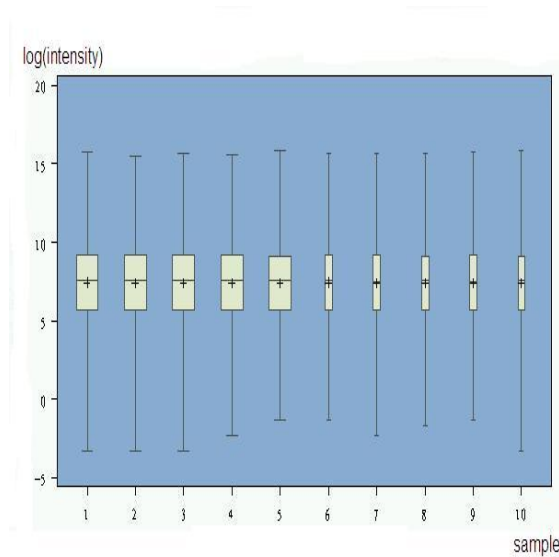


Figure 6.5: Boxplot of Gene Expression for Five Experiments and Five Controls

6.1.5 Local Z-score Method

As mentioned in fold change method, the scatter plot of Log-ratio against Log-intensity is funnel-shaped. Gene expression intensities have high variation in low intensity and lower variation in high intensity. This is an extremely serious problem in our data. If only use a cut-off threshold, we will get a large quantity of false positive genes in low intensity and miss differentially expressed genes in high intensity. Local Z-score method balances the weights on the low and high intensity. The data we are analyzing is gene intensity (MAS 5.0) from five experiments (STAT3) and five controls (GFP), which has 10 columns and 22215 rows. Steps in Calculating Local Z-score is shown as following:

Step 1. Normalization across array: each column is divided by its median. Figure 6.5 displays boxplots for ten-array gene expression before normalization. After step 1, median will be exactly same.

Step 2. Get Combined Experiment Intensity and Combined Control Intensity: Fifth root of product of five experiments (controls).

Step 3. Calculate LOG(RATIO) and LOG(INTENSITY) based on combined experiment intensity and combined control intensity.

Step 4. Lowess Normalization to LOG(RATIO) (in our case, ratio is centered at 0 already)

Step 5. Graph LOG(RATIO) versus LOG(INTENSITY)

Step 6. Split data into 4 local parts based on LOG(INTENSITY) (by quartiles of LOG(INTENSITY))

Step 7. Each part, calculate the local Z-score :

$$Z - \text{score} = \frac{\text{Log}(\text{ratio})}{\text{local SD}}.$$

We call genes whose z-score is greater than 2 differentially expressed genes. In our result table, we have 159 genes from part I (black), 167 genes from part II (blue), 152 genes from part III (green) and 170 genes from part IV (red).

Main advantage of local z-score is that it partially solves the problem that the change of low-intensity genes is more significant than high-intensity genes. In local z-score method, it gives a good shot to find significant genes in high intensity. One-cut of fold change will give us too many low-intensity genes. All level of intensity get relatively equal chance.

6.2 Select Differentially Expressed Genes by Bayesian and Empirical Bayes Methods

There is a kind of information sharing among genes. The data from other genes provide some information about typical variability in the system. This can benefit our analysis because experiments often involve tens of thousand of genes but only tens of microarrays, so the amount of information per genes can be relatively low. Empirical Bayes (EB) Methods are well-suited to high dimensional inference problems and thus provide a

natural approach to microarray data analysis. In general, classical methods focus on controlling false discovery rate, while Empirical Bayes method "EBarrays" aims to classify genes by expression patterns using posterior probability. (Newton(2002)[42])

6.2.1 EBarrays

Instead of applying statistical inference on individual genes as in classical methods, Empirical Bayes analysis takes account of information that shares among genes. Michael A. Newton and Christina Kendziorski (2002) developed an EB framework for selecting differentially expressed genes and **EBarrays** is the software package which is available in an R library. Output in **EBarrays** provides the posterior probabilities of differential expression across multiple conditions. This section focuses on an overview of how EBarrays works under two conditions.

In our data, data can be described in two patterns:

$$(1, 1, 1, 1, 1, 1, 1, 1, 1, 1) \text{ and } (1, 1, 1, 1, 1, 2, 2, 2, 2, 2).$$

The first pattern presents equivalent expression (EE) and the second pattern presents differential expression (DE). Results from **EBarrays** provide the posterior probabilities of two patterns as illustrated in Table 6.2. P_1 is the posterior probability that a gene has an EE pattern and P_2 is the posterior probability that a gene has a DE pattern.

Here is the mathematical structure:

A distribution for equivalent expression (EE_j) for gene j , sample $i=1,2,\dots,n$:

$$f_o(x_j) = \int (\prod_{i=1}^n f_{obs}(x_{ji}|\mu))\pi(\mu) d\mu.$$

A distribution for differential expression (DE_j):

$$f_1(x_j) = f_o(x_{j1})f_o(x_{j2}).$$

The posterior probability of differential expression calculated by Bayes' rule is:

$$\frac{pf_1(x_j)}{pf_1(x_j) + (1 - p)f_o(x_j)}.$$

Table 6.2: EBarrays Result Table

Name	P_1	P_2	Fold Change	Link	Description
117_at	7.01E-44	1.00E+00	0.417294305	3310	X51757/FEATURE=cds
200612_s_at	2.12E-04	1.00E+00	0.72426676	163	gb:NM_001282.1
200613_at	2.61E-04	1.00E+00	0.728831113	1173	gb:NM_004068.1
200632_s_at	2.74E-29	1.00E+00	2.053329445	10397	gb:NM_006096.1
200635_s_at	3.00E-06	1.00E+00	1.429124935	5792	gb:AU145351
200636_s_at	9.75E-06	1.00E+00	1.418685758	5792	gb:NM_002840.1
200637_s_at	2.63E-09	1.00E+00	1.525636934	5792	gb:AI762627
200697_at	1.58E-26	1.00E+00	1.987423539	3098	gb:NM_000188.1
200699_at	3.32E-06	1.00E+00	0.695327778	11014	gb:BE962456
200768_s_at	3.62E-04	1.00E+00	1.376725964	4144	gb:BC001686.1
200769_s_at	2.92E-04	1.00E+00	1.378797068	4144	gb:NM_005911.1

Two particular specifications of the general mixture models are described in EBarray: Gamma-Gamma model and Lognormal-Normal model. In Gamma-Gamma model, it assumes observation component is a Gamma distribution with shape parameter $\alpha > 0$ and a mean value μ_j , while marginal distribution for μ_j is an Inverse Gamma with hyperparameters. In Lognormal-Normal model, it assumes observation component is a Lognormal distribution with a mean value μ_j and common variance σ^2 , marginal distribution μ_j is a normal distribution with hyperparameters. It is critical to check model assumption before we apply it. Figure 6.6 displays a good fit of gene expression from Li-Wong normalization to Gamma-Gamma model.

An important advantage in EBarrays is that the data from other genes provide some information about the typical variability in the system through marginal distri-

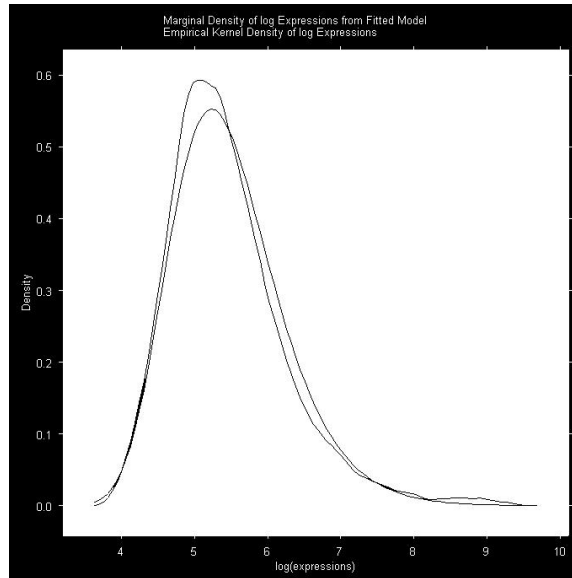


Figure 6.6: EBarrays Fit GG Model - From Li Wong Normalized Data

bution. The general framework provided by EB analysis is quite flexible. Probability distributions are specified in several layers and account for multiple sources of variation. The posterior probability of differential expression is a very useful inference in EBarrays. This transforms evidence to the familiar scale of probability. Posterior probability calculation carries over naturally to comparisons among multi conditions. It is easier to be interpreted to non-statistician.

The methods that treat genes as separate fixed effects may have reduced efficiency compared to methods that treat the genes as arising from some population since they do not take advantage of the level of information sharing among genes. Furthermore, classifying genes into expression patterns by the posterior probability is an optimal procedure in the context of the mixed model: it minimized the expected number of errors. In classical testing, the goal is to bound the false discovery rate and maximize the power.(Newton(2002)[42])

A main drawback is that the data may not satisfy the assumption, for instant, the distributions of expression across genes or within array do not have a normal or gamma

distribution, which occurs very often.

6.2.2 LIMMA

LIMMA is a software package associated with the paper titled "Linear Models and Empirical Bayes Methods For assessing differential expression in microarray experiments" produced by Gordon K.Smyth (2003)[55].

This paper extended and reset the hierarchical model of Lonnstedt and Speed (2002) [40] in the context of general linear models. Consistent and closed form estimators are derived through the marginal distribution of the observed statistics. The advantage of this method that the estimator obtained lower false discovery is shown in a simulation study.

A Bayes inferential approach is proposed in terms of moderated t-statistic in which posterior variances are substituted for the sample-variances. The moderated t-statistic has the advantage over the ordinary t-statistic that very small sample variances are heavily balanced while larger sample variances are moderated to a less relative degree. The moderated t-statistic approach has the advantage over the posterior odds that the number of hyperparameters which are needed to be estimated is reduced and knowledge of non-null prior for the fold change is not required. The moderated t-statistic is distributed independently of the sample variances so that inference about the variance and effect hyperparameters can be considered separately. Moreover, the inferential approach extends easily to provide tests involving two or more patterns through the use of moderated F-statistics. Table 6.3 gives a partial result for our data from LIMMA package. M is the fold change. t is the ordinary t-statistics. B is the moderated t-statistic associated with p-value.

Table 6.3: LIMMA Result Table

Name	M	t	P.Value	B	Description
208992_s_at	3.576775568	43.75188021	1.51E-12	26.00787978	gb:BC000627.1
204988_at	1.826076349	31.30802095	9.49E-11	22.89101456	gb:NM_005141.1
205650_s_at	2.100255784	29.96882705	1.19E-10	22.42797956	gb:NM_021871.1
222088_s_at	1.597187059	26.40506191	5.19E-10	21.02410681	gb:AA778684
202760_s_at	1.519678276	25.54032036	5.88E-10	20.64054422	gb:NM_007203.1
216236_s_at	1.36024007	25.21315777	6.06E-10	20.49054964	gb:AL110298.1
205649_s_at	1.535027452	24.67935255	6.42E-10	20.2398186	gb:NM_000508.2
216238_s_at	2.307502018	24.66346006	6.42E-10	20.23223735	gb:BG545288

6.2.3 Cyber-T

The idea of using a t-statistic with a Bayesian adjusted denominator was proposed by Baldi and Long (2001) who also developed a useful Cyber-T program. Independent normal distributions are modeled for log-expression value. It is reasonable to assume mean and variance are dependent based on the inspection of typical microarray data sets. In this method, the prior of mean conditional on variance has a normal distribution with two hyperparameters. The prior of variance is assumed to have an inverse Gamma with another two hyperparameters. They obtained the posterior density, which has same form as the joint prior density. Bayesian estimates for parameter and hyperparameters are obtained. Simulation shows that these point estimates, combined with a t-test, provide a systematic inference approach that compares favorably with simple t-test or fold methods, and partly compensate for the lack of replication.

However, the work was limited though to two-sample control versus treatment designs and the model didn't distinguish between differentially and non-differentially expressed genes. They also didn't develop consistent estimators for the hyperparameters. The degrees of freedom associated with the prior distribution of the variances was set to a

default value while the prior variance was simply equal to locally pooled sample variances.

6.3 Other Work in Microarray

In spite of the fact that differential expression can be applied to a large number of genes, it remains within the restriction of the old one-gene-at-a-time model. Most genes act related with other genes. The patterns of expression across multiple genes and experiments are critical in DNA microarray analysis. To detect such patterns, clustering must be introduced.

At this level, instead of assuming genes are independent, researchers are interested in genes covariance, at whether there exists multi-gene patterns, cluster of genes that share the same behavior over or across different treatments. Multi-gene expression patterns can be used to characterize diseases and discriminate, for example, different kinds of cancers. Various clustering methods (Sebastiani 2002) have been proposed, including k-means, hierarchical clustering. Clustering methods can be applied not only to genes, but also to conditions, DNA sequences, and other related data. Most popular package in clustering (microarray data) is from Michael Eisen's lab and here is the URL.<http://rana.lbl.gov/EisenSoftware.htm> .

Array data is inherently high-dimensional, hence dimensionality reduction and visualization are particularly useful. Principal component analysis (PCA)(Mike West 2002) and clustering are the most important and widely used methods. PCA can be viewed as a method to compress and visualize data. It provides an optimal linear dimension reduction technique in the mean-square sense.

6.4 Summary

We worked on the selection of differentially expressed genes among 22,215 genes and obtained several gene lists as results. Both classical and Bayesian methods are applied. Classical methods consist of fold change, T-test, Wilcoxon Rank-sum test, Locally Z-score and SAM(Storey 2002). Empirical Bayes methods consist of EBarrays(Newton, 2002[42]), LIMMA(Smyth G.K.(2003)[55]) and Cybor-T (Baldi and Long(2001)[5]). SAM (Storey 2002), local Z-score (Z. Chen[17]) and a parametric Empirical Bayes method (Newton, 2002) are discussed with more detail. We illustrated the procedures by a form of clear steps. We also showed the main advantage and drawback of each method and explained result table. Higher level of microrray analysis, such as clustering, PCA, can be done as further work.

Bibliography

- [1] Anderson,P. and Borgan,O., *Statistical Models Based on Counting Processes*.(1992)
- [2] Ascher,H. and Feingold,H., *Repairable Systems Reliability*.(1984)
- [3] Bain, L.J., *Statistical Analysis of Reliability and Life Testing Models*(1978)
- [4] Bain,L.J. and Engelhardt,M., *Statistical Analysis of Reliability and Life Testing Models*(1991)New York: 2nd edition Marcel Dekker.
- [5] Baldi,P. and Hatfield,G.W., *DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling*.(2002)
- [6] Baldi,P. and Long,A.D., A Bayesian Framework for the Analysis of Microarray Expression Data: Regularized t-Test and Statistical Inferences of Gene Changes. *Bioinformatics*.**17**,(2001) 509-519.
- [7] Bar-Lev S.K., Lavi,I. and Reiser,B., Bayesian Inference for The power law process, *Ann. Inst. Statist. Math.***Vol. 44, No. 4** (1992) 623-639.
- [8] Beiser,J.A. and Rigdon,S.E., Bayes Prediction for the Number of Failures of a Repairable System. *IEEE Transaction on Reliability*.**46**291-295(1997)
- [9] Berger, J., *Statistical Decision Theory and Bayesian Analysis*, 2nd Edition, Springer-Verlag, New York, Berlin, and Heidelberg.(1985)
- [10] Box,G.E.P. and Tiao, G.C., Bayesian Inference in Statistical Analysis, *Addison-Wesley, Reading, Massachusetts*(1973).

- [11] Broberg,P., Statistical Methods for Ranking Differentially Expressed Genes, *Genome Biology*(2003)
- [12] Calabria, R. Guida,M. and Pulcini,G., Some Modified Maximum Likelihood Estimators for the Weibull Process. *Reliability Engineering and System Safety*. **23**51-58.(1987)
- [13] Calabria,R., Guida, M. and Pulcini, G., Bayes Estimation of Prediction Intervals for a Power Law Process, *Comm. Statist. Theory Methods*.**19**,(1990)3023-3035.
- [14] Calabria,R., Guida,M. and Pulcini,G., Power Bounds for a Test of Equality of Trends in k Independent Power Law Processes. *Comm. Statist. Theory Methods*.**21**(1992)3275-3290.
- [15] Calabria,R. and Pulcini,G., Bayes Inference for the Modulated Power Law Process, *Comm. Statist. Theory Methods*.**26(10)**,(1997),2421-2438.
- [16] Chen,Z. and Rao,A.N.V., Bayesian Inference on Power Law Process, *Fourth International Conference on Dynamic Systems and Applications at Atlanta*(submitted) May 21-24 (2003)
- [17] Chen,Z. and Rao, A.N.V., Methods of Selecting Differentially Expressed Genes, *Fourth World Congress of Nonlinear Analysts WCNA-2004*(in preparation) June 30-July 7 (2004)
- [18] Chen,Z., Computing Probabilities Using the Multiplication Rule, Presentation at Wilkes Barre, PA(2004)
- [19] Chen,Z., Hypothesis Testing: Calculating and Minimizing Type II Error Probability, Presentation at Fort Myers, FL(2004)

- [20] Cox D.R. and Lewis. Wiley,P.A.W., *The Statistical Analysis of Series of Events.* (1966)
- [21] Crow, L.H., Reliability Analysis for Complex Repairable Systems. *Reliability and Biometry*
- [22] Crow, L.H., Confidence Interval Procedures for the Weibull Process with Applications to Reliability Growth. *Technometrics.* **24**(1982)67-72.
- [23] *Data Analysis Tools for DNA Microarrays*(2003)
- [24] Duane, J.T., Learning Curve Approach to Reliability Monitoring, *IEEE Trans. Aerospace Electron System.***2**(1964)563-566
- [25] Dudoit, S. et al., Statistical Methods for Identifying Differentially Expressed Genes in Replicated cDNA Microarray Experiments. *Statistica Sinica.***12**(1)(2002)111-139.
- [26] Engelhardt, M. and Bain, L.J., On the Mean Time Between Failures for Repairable Systems. *IEEE Transactions on Reliability*(1986)419-422.
- [27] Engelhardt, M. and Bain, J.L., Statistical Analysis of a Compound Power Law Model for Repairable Systems, *IEEE Transactions on Reliability, R-***36** (1987) 392-396.
- [28] Engelhardt, M. and Bain, J.L., Prediction Intervals for the Weibull Process, *Technometrics.* **20** (1978) 167-169.
- [29] Gnedenko,B.V. and Ushakov,I.A., *Probilistic Reliability Engineering.*(1995)
- [30] Guida,M., Calabria, R. and Pulcini,G., Bayes Inference for a Non-homogeneous Poisson Process with Power Law Intensity. *IEEE Transactions on Reliability. R-***38**(1989)603-609.

- [31] Hatmann,O. and Samans,B. Microarray Data Normalization and Transformation: Comparing MAS, VSN and RMA for Affymetrix GeneChips. *Technical Report: Affymetrix GeneChip normalization with BioConductor*(2003)
- [32] Higgins, J.J. and Tsokos, C.P., A quasi-Bayes estimate of the failure intensity of a reliability-growth model, *IEEE Transactions on Reliability. R-***(30)** (1981) 471-475.
- [33] Huang,Y. and Bier,V., A Natural Conjugate Prior for the Non-Homogeneous Poisson Process with a Power Law Intensity Function, *Commun. Statist.***27(2)**525-551.(1998)
- [34] Huber,W., Heydebreck,A., Sultmann,H., Poustka,A. and Vingron,M., Variance Stabilization Applied to Microarray Data Calibration and to the Quantification of Differential Expression. *Bioinformatics.***18**96-104.(2002)
- [35] Kyparisis,J. and Singpurwalla, N.D., Bayesian Inference for the Weibull Process with Applications to Assessing Software Reliability Growth and predicting Software Failures. *Computer Science and Statistics: Proceedings of the Sixteenth Symposium on the Interface.*(1985)
- [36] Lee,L. and Lee,K. , Some Results on Inference for the Weibull Process, *Technometrics.***20**,(1978)41-45.
- [37] Lee, M., Kuo,F., Whitmore,G. and Sklar, J., Importance of Replication in Microarray Gene Expression Studies: Statistical Methods and Evidence from Repetitive cDNA Hybridizations. *PNAS*(2000)
- [38] Lewin, A., Richardson, S., Marshall, C., Glazier, A. and Aitman, T., Bayesian Modelling of Differentially Gene Expression(2003).
- [39] Lindqvist, B., Statistical Modeling and Analysis of Repairable Systems, *Mathematical Methods in Reliability.***September**,(1997)16-19.

- [40] Lonnstedt,I. and Speed,T.P., Replicated Microarray Data. *Stat. Sinica.* **12**,31-36(2002)
- [41] Musa J.D., *Software Reliability Data*, deposited in IEEE Computer Society Repository,New York,1979.
- [42] Newton M. and Kendzioriski, Parametric Empirical Bayes Methods for Microarrays. *The Analysis of Gene Expression Data: Methods and software.*(2002)254-271.
- [43] Parmingiani, G., Garrett,E.S., Irizarry,R.A. and Zeger,S.L., *The Analysis of Gene Expression Data: Methods and Software.*(2003)
- [44] Qiao, H. and Tsokos, C. , Best Efficient Estimates of the Intensity Function of the Power Law Process. *Journal of Applied Statistics*,**25**,111-120(1998)
- [45] Qiao,H. and Tsokos,C.P., Reliability Growth: MTBF vs. Intensity Function of the Weibull Process.
- [46] Quackenbush, J., Microarray Data Normalization and Transformation, *Nature Genetics Supplement*,**32**,(December 2002)496-501.
- [47] Rigdon,S.E. and Basu,A.P., The Effect of Assuming a Homogeneous Poisson Process When the True Process is a Power Law Process. *Journal of Quality Technology*,**22**,111-117(1990)
- [48] Rigdon,S.E. and Basu, A.P., The Power Law Process: a Model for the Reliability of Repairable systems. *Journal of Quality Technology***21**,251-260.(1990)
- [49] Rigdon,S.E. and Basu, A.P., Estimating the Intensity Function of a Power Law Process at the Current Time: Time Truncated Case. *Communications in Statistics Simulation and Computation***19**,1079-1104.(1990)

- [50] Rigdon,S.E. and Basu,A.P., *Statistical Methods for the Reliability of Repairable system*(2000)
- [51] Roberts,H.W. and Rao,A.N.V., Predicting the Performance of Software Systems via the Power Law Process, *Dissertation of University of South Florida*(2000)
- [52] Singpurwalla, N. and Wilson, S., *Statistical Methods in Software Engineering*.(1999)41-48.
- [53] Speed,T., *Interdisciplinary Statistics: Statistical Analysis of Gene Expression Microarray Data*.(2003)
- [54] Sinha,S.K., John Wiley and Sons, *Reliability and Life Testing*.(1986).
- [55] Smyth G. K. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology***3**(2003).
- [56] Storey, J.D. and Tibshirani, R., SAM Thresholding and False Discovery Rates for Detecting Differential Gene Expression in DNA Microarray. *The Analysis of Gene Expression Data*(2002)272-290.
- [57] Thompson,Jr.,W.A. On the Foundations of Reliability. *Technometrics***23**(1981).
- [58] Tsokos,C.P., Reliability Growth: Nonhomogeneous Poisson Process. *Recent Advances in Life-Testing and Reliability*319-334,CRC,Boca Raton,FL.(1995)
- [59] Tsokos,C. and Rao,A.N.V., Estimation of Failure Intensity for the Weibull Process, *Journal of Engineering Reliability and System Safety*(1995)
- [60] Ushakov,I., *Handbook of Reliability Engineering*. (1994)

About the Author

Ms. Zhao Chen got her bachelor degree in International Finance from Hunan University in 1997. Then she worked as an accountant for two years in China. She started her master program in the mathematics department of the University of South Florida in August 1999. She received her master degree in mathematics in May 2001. She was admitted to doctoral program specialized in Statistics in Fall 2001.

During her study at the University of South Florida, she received financial support as a teaching assistant. She has instructed various mathematics and statistics courses. Ms. Chen attended a number of seminars and presented her research interest in two international conferences and weekly seminar in the department. She also worked as a volunteer instructor in USF Urban Scholars Program. From last July to the present, she works as a biostatistician in H. Lee Moffitt Cancer Center, specialized in microarray and protein analysis.