BATHTUB FAILURE RATES OF MIXTURES IN RELIABILITY AND THE SIMES INEQUALITY UNDER DEPENDENCE IN MULTIPLE TESTING

by

Jie Wang

B.S., University of Science and Technology of China, 2002

M.A., University of Pittsburgh, 2007

Submitted to the Graduate Faculty of

the Department of Statistics in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

STATISTICS DEPARTMENT

This dissertation was presented

by

Jie Wang

It was defended on

June 2nd 2008

and approved by

Henry W. Block, Department of Statistics

Thomas H. Savits, Department of Statistics

Leon J. Gleser, Department of Statistics

Gong Tang, Department of Biostatistics

Dissertation Advisors: Henry W. Block, Department of Statistics,

Thomas H. Savits, Department of Statistics

BATHTUB FAILURE RATES OF MIXTURES IN RELIABILITY AND THE SIMES INEQUALITY UNDER DEPENDENCE IN MULTIPLE TESTING

Jie Wang, PhD

University of Pittsburgh, 2008

Two topics are presented in this dissertation: (1) obtaining bathtub-shaped failure rates from mixture models; and (2) the Simes inequality under dependence.

The first topic is in the area of reliability theory. Bathtub-shaped failure rates are well-known in reliability due to their extensive applications for many electronic components, systems, products and even biological organisms. Here we derive some the conditions for obtaining bathtub-shaped failure rates distributions from mixtures, which have been utilized to model heterogeneous populations. In particular, we show that the mixtures of a family of exponential distributions and an IFR gamma distribution can yield distributions with bathtub-shaped failure rates.

The second topic is concerned with the area of multiple testing, but uses dependence concepts important in reliability. Simes [1986] considered an improved Bonferroni test procedure based on the so-called Simes inequality. It has been proved that this inequality holds for independent multivariate distributions and a wide class of positively dependent distributions. However, as we show in this dissertation, the inequality reverses for a broad class of negatively dependent distributions. We also make some comments with regard to the Simes inequality and positive dependence.

ACKNOWLEDGMENTS

Many people deserve special thanks and acknowledgment for their guidance, support, and friendship during my time here at Department of Statistics, University of Pittsburgh. First and foremost are my advisors, Dr. Henry W. Block and Dr. Thomas H. Savits. Their willingness and enthusiasm to explore always one step more has made the work of developing this dissertation an absolute joy. For their unbounded insights into statistics, research, education and life, I am eternally grateful. I want to thank Dr. Leon J. Gleser and Dr. Gong Tang for their thoughtful contributions as members of my dissertation committee and for taking a genuine interest in my work. I am also grateful to Dr. Allan R. Sampson for his invaluable comments and tremendous support throughout my journey of pursuing the Ph.D. degree.

I am also indebted to all of the faculty, staff and graduate students whom I have had the privilege of knowing and working with over the past six years. I cannot hope to thank them all by name here. However, there are two with whom I have worked especially closely. I would like to thank Wei Zhang for her thoughts and advice on everything from my research to cooking great meals. Qiang Wu helped me a lot from surviving tough courses to dealing with LaTeX. I am truly grateful. In addition, I want thank Scott Nickleach for sharing his teaching experience when I was a first-time instructor.

I would never have made it here had it not been for the love and support of my family, especially my parents and sister. I cannot begin to thank you, except to say I owe where I am today to you. Finally, to my husband Yang, thank you for your patience, understanding, support, and love; without it, I would not have finished this dissertation.

TABLE OF CONTENTS

1.0 INTRODUCTION

1.1 BATHTUB-SHAPED FAILURE RATES OF MIXTURES

Many populations in reliability theory are considered to be heterogeneous. The reason is that there are usually at least two subpopulations. These heterogeneous populations can simply arise due to physically mixing similar products from different streams, or pooling data to enlarge the sample size. Therefore, we often use the mixture of the subpopulations' lifetime distributions to model the lifetime distribution of the entire population. In this case, the aging behavior of the mixture is often complicated and difficult to predict.

The aging behavior is studied through the failure rate which is a function of time t. It is an important characteristic of lifetime distributions. The failure rate function measures the chance of an object failing at a specified time t. Large failure rates imply more of a chance to fail and small failure rates imply less of a chance to fail. A natural progression of aging or wear out is suggested by an increasing failure rate; a decreasing failure rate suggests an improvement over time. A constant failure rate implies neither wear out nor improvement. Only exponential distributions exhibit constant failure rates.

Learning the aging behavior of an object or aging process or a system gives us valuable information. For example, many electronic components, systems and products exhibit a bathtub-shaped failure rate, as shown in Figure 1. This failure rate consists of three periods: early, useful and wear-out life. The early life shows a decreasing failure rate. The wear-out life shows an increasing failure rate. The useful life shows an approximately constant and minimum failure rate which thus has the lowest chance to fail and consequently the highest reliability to perform. This kind of curve, in engineering, enables the determination of an optimum break-in time, for example, the optimum burn-in time, the optimum warranty period and cost, the optimum preventive replacement time of components, and the spare parts requirements and their production rate.

The aging behaviors of mixtures have long been studied in reliability. A classic paper of Proschan [1963] studied mixtures of exponential distributions and observed that the failure rate of the mixture was decreasing. Furthermore, it is intuitively obvious that the limit of the failure rate of the mixture of exponentials is the failure rate of the strongest exponential, i.e., the one with the smallest failure rate. One of the first papers to infer this result was the paper of Clarotti and Spizzichino [1991]. Block, Mi, and Savits [1993] gave a general version of this result and showed that, subject to mild technical assumptions, the asymptotic failure rate of a mixture is the asymptotic failure rate of the strongest subpopulation. A related result of Block and Joe [1997] shows that for a mixture of lifetimes, where the failure rate of the lifetimes are essentially ratios of polynomials (as most well-known lifetime distributions are), the limiting distribution of the mixture has the same eventual monotonicity as that of the strongest component. For example, if the failure rate of the strongest component is eventually increasing, so is the failure rate of the mixture. A recent paper, Block, Li, and Savits [2003a] gives improved versions of these two results. In this latter paper, the initial behavior of mixtures is also discussed, as are the initial and eventual failure rates for systems of components.

Currently, much is known about the asymptotic and initial behavior of the failure rates of mixtures, but not much is known about the intermediate behavior. The intermediate behavior has been studied mostly in the case of two known distributions. Gurland and Sethuraman [1994], (1995) studied the intermediate behavior of failure rates for distributions mixed with an exponential. Block, Savits, and Wondmagegnehu [2003b] considered mixtures of two increasing linear failure rates. Gupta and Warren [2001] studied mixtures of gamma distributions. Jiang and Murthy [1998] examined mixtures of Weibulls. Wondmagegnehu, Navarro, and Hernandez [2005] discussed a variety of mixtures of Weibulls and exponentials.

In this dissertation we go beyond mixtures of two distributions and consider the intermediate behavior for continuous mixtures of various distributions. In particular, we show how mixtures of whole families of distributions yield a distribution with a bathtub-shaped failure rate. Specifically, we show that when a continuous mixture of exponentials, gammas

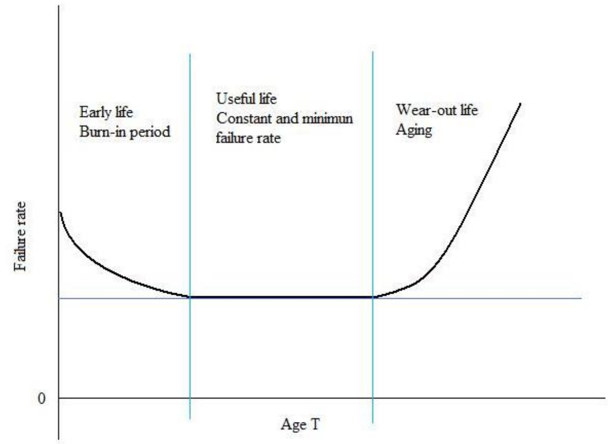

Figure 1: Bathtub-shaped failure rate curve

or Weibulls with decreasing failure rate are mixed with a gamma with increasing failure rate, a bathtub-shaped failure rate distribution may be obtained. We also examine the conditions for obtaining a bathtub-shaped failure rate when a continuous mixture of exponentials is mixed with a continuous mixture of gammas with increasing failure rate.

1.2 THE EFFECT OF DEPENDENCE ON THE SIMES PROCEDURE

When testing multiple hypotheses simultaneously, the Bonferroni procedure is often used. Consider m null hypothesis $H_{01}, H_{02}, \ldots, H_{0m}$ with the observed p-values $p_1, p_2, \ldots, p_m$, respectively. The Bonferroni procedure rejects the overall null hypothesis $H_0 = \cap_{i=1}^{m} H_{0i}$ if any p-value is less than or equal to α/m, where α is the desired significance level of the test for H_0. Furthermore the individual null hypothesis H_{0i} is rejected if the corresponding observed p-value satisfies $p_i \leq \alpha/m$. The Bonferroni inequality then insures that the familywise error rate (the probability of making one or more false rejections among all $H_{01}, H_{02}, \ldots, H_{0m}$

when H_0 is actually true) is controlled at level less than or equal to α, i.e.,

$$Pr(\bigcup_{i=1}^{m} \{P_i \leq \alpha/m\} \,|H_0 \text{ is true}) \leq \alpha. \tag{1.1}$$

P_i, $i = 1, \ldots, m$ are the random p-values defined through the distributions of the test statistics. For instance, if the distributions of X_i are symmetric, then

$$P_i = \begin{cases} F_{H_{0i}}(X_i) & \text{for a left tail test,} \\ \bar{F}_{H_{0i}}(X_i^-) & \text{for a right tail test,} \\ 2\min\left\{F_{H_{0i}}(X_i), \bar{F}_{H_{0i}}(X_i^-)\right\} & \text{for a two-tail test.} \end{cases}$$

where X_i are the test statistics. $F_{H_{0i}}$ and $\bar{F}_{H_{0i}}$ are the cumulative distribution functions and the survival functions of the distribution of X_i when H_{0i} is true, $i = 1, \ldots, m$.

The Bonferroni procedure is widely used because it is simple and requires no distributional assumptions. However, in some cases, this procedure may be too conservative and may lack power. Therefore, many modified procedures have been proposed to make the test less conservative and more powerful.

Simes [1986] suggested a procedure which rejects the overall hypothesis H_0 if $p_{(j)} \leq j\alpha/m$ for some $j = 1, \ldots, m$, where $p_{(j)}$ is the jth smallest observed p-value. Simes proved that

$$Pr(\bigcup_{j=1}^{m} \left\{P_{(j)} \leq j\alpha/m\right\} |H_0 \text{ is true}) \leq \alpha \tag{1.2}$$

holds with equality for independent statistics and conjectured it might also be true for a large family of multivariate distributions of test statistics. When the inequality (1.2) holds, the modified procedure controls the familywise error rate at level α. In addition, Simes provided some simulation results by comparing the modified procedure with the Bonferroni procedure. The simulation shows that, for those specified cases, the modified procedure is less conservative and considerably more powerful. In fact, this is always true when comparing the two procedures since that

$$\bigcup_{i=1}^{m} \{P_i \leq \alpha/m\} = \left\{P_{(1)} \leq \alpha/m\right\} \subset \bigcup_{j=1}^{m} \left\{P_{(j)} \leq j\alpha/m\right\}.$$

Thus if an overall null hypothesis has been rejected by the Bonferroni procedure, it must also be rejected by the Simes procedure. On the contrary, if it has been rejected by the Simes procedure, it is not necessary to be rejected by the Bonferroni procedure too. As a result, a test using the Simes procedure is more powerful and less conservative than a test using the Bonferroni procedure. However, as Simes realized, the inequality (1.2) is not true in general, because counterexamples can be found.

Samuel-Cahn [1996] showed that, for one-sided tests with positively correlated bivariate normal test statistics and two-sided tests with bivariate normal test statistics, (1.2) holds; but for one-sided test with negatively correlated bivariate normal test statistics, the inequality (1.2) is reversed. Sarkar [1998] proved that (1.2) is true for MTP_2 (multivariate totally positive of order two) random variables. He also mentioned that it was believed that for some negatively dependent multivariate distributions, the inequality would not hold.

Another popular modified Bonferroni-type procedure is the BH procedure proposed by Benjamini and Hochberg [1995]. The BH procedure rejects all $H_{(1)}, \ldots, H_{(k)}$, where k is the largest j for which $P_{(j)} \leq \frac{j}{m}\alpha$. Instead of the familywise error rate, the BH procedure tries to control the false discovery rate (see Chapter 4). Benjamini and Hochberg [1995] proved that the BH procedure controls the false discovery rate for independent test statistics and Benjamini and Yekutieli [2001] showed that under certain types of positive dependence, the procedure controls the false discovery rate. However, little is known about the BH procedure under any type of negative dependence structure.

Thus, as we can see from above, the dependence structure of test statistics plays an important role in these modifications. In this dissertation, we show that, subject to mild conditions on the marginal distributions of test statistics, the Simes inequality (1.2) holds for a broad class of positively dependent multivariate distributions, but reverses for a large family of negatively dependent multivariate distributions. We also demonstrate that (1.2) holds for a certain type of positively dependent multivariate t.

1.3 CONTENTS OF THE DISSERTATION

In Chapter 2, we provide a review of lifetime distributions. Section 2.1 covers some basic characteristics of lifetime distributions and, in particular, discusses the exponential, gamma and Weibull distributions. Section 2.2 introduces mixtures by presenting definitions, failure rate functions and a brief literature review. Section 2.3 gives a variety of shapes of failure rates and sufficient conditions to identify these shapes.

In Chapter 3, we state the results concerning when bathtub-shaped failure rates are obtained. Section 3.1 concerns mixing a continuous mixture of exponentials, gammas or Weibulls with decreasing failure rate and a gamma with increasing failure rate. Section 3.2 expands the result to mixtures produced by a continuous mixture of exponentials, gammas or Weibulls with decreasing failure rate and a mixture of gammas with increasing failure rate.

In Chapter 4, we provide a brief review of multiple testing procedures and statistical dependence. Section 4.1 reviews some general concepts in multiple testing. Section 4.2 covers the classic Bonferroni procedure. Section 4.3 briefly introduces the Simes procedure and the Simes inequality which are the main concern of our work. Section 4.4 provides the definitions for a few dependence structures, including totally positive of order 2, positively (negatively) dependent through stochastic ordering and condition N.

In Chapter 5, we present the results for controlling familywise error rate under dependence when using the Simes procedure. Section 5.1 and Section 5.2 give results under the positive dependence and negative dependence, respectively.

Chapter 6 presents some thoughts about future research. In Section 6.1, we consider further situations which might yield bathtub-shaped failure rates. In Section 6.2, we discuss the possibility of controlling the false discovery rate (instead of family-wise error rate) under dependence.

2.0 LIFETIME DISTRIBUTIONS

2.1 INTRODUCTION TO LIFETIME DISTRIBUTIONS

We begin with a discussion of lifetimes. These lifetimes could be of machines or biological organisms. Despite its engineering or biological complexity, we consider an object as a whole, that is, we do not consider its structure and its constituent parts at this point. A failure occurs when the object ceases to function. Usually, the lifetime is modeled as a nonnegative random variable with certain distribution.

2.1.1 Some key characteristics

Let T be the lifetime of a unit, $T \geq 0$. The cumulative distribution function is given by

$$F(t) = Pr(T \leq t).$$

If T is continuous, the distribution of T can also be characterized by its probability density function

$$f(t) = \frac{dF(t)}{dt}.$$

In reliability, we also use the survival function, or reliability function defined by

$$\bar{F}(t) = 1 - F(t) = Pr(T > t).$$

When T is continuous, the failure rate function is defined by

$$r(t) = \frac{f(t)}{\bar{F}(t)}.$$

Roughly speaking, the failure rate represents the rate at which an object fails instantaneously given that it has survived up to time t.

2.1.2 Some lifetime distributions

Next we introduce the exponential, gamma and Weibull distributions. All three are continuous distributions which are standard lifetime distributions used in reliability. We assume time t is always nonnegative in the following discussion unless specified otherwise.

2.1.2.1 Exponential distribution The probability density function of an exponential distribution is given by

$$f(t) = \lambda e^{-\lambda t}, \quad \lambda > 0.$$

The survival function is

$$\bar{F}(t) = e^{-\lambda t}, \quad \lambda > 0$$

and the failure rate is $r(t) = \lambda$. The exponential distribution is the only distribution having a constant failure rate. The exponential distributions are a special case of many other families of distributions, such as the gamma and Weibull. In the reliability literature, the exponential distribution plays a central role.

2.1.2.2 Gamma distribution The gamma distribution involves two parameters: the shape parameter $\alpha > 0$ and the scale parameter $\lambda > 0$. The density function of a gamma is given by

$$f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t},$$

where $\Gamma(\alpha)$ is the gamma function.

The reliability function is given by

$$\bar{F}(t) = \int_t^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} s^{\alpha-1} e^{-\lambda s} ds.$$

In general, there is no a simple form for the survival function unless α is a positive integer. Then it can be written as

$$\bar{F}(t) = \sum_{i=0}^{\alpha-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t}.$$

and the failure rate function is then given by

$$r(t) = \frac{\frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1}}{\sum_{i=0}^{\alpha-1} \frac{(\lambda t)^i}{i!}}.$$

The gamma distribution arises naturally as the time-to-first failure distribution for a system with standby exponentially distributed backups. If there are $n - 1$ standby backup units and the system and all backups have the same exponential lifetimes with parameter λ, then the total lifetime has a gamma distribution with parameter $\alpha = n$ and λ. The gamma is a flexible life distribution model that may offer a good fit to some sets of failure data.

2.1.2.3 Weibull distribution The Weibull distribution is one of the most commonly used families of distributions in reliability engineering because of its many shapes. It is a very flexible life distribution model also with shape and scale parameters, $\beta > 0$ and $c > 0$ respectively. The density function is given by

$$f(t) = \beta c^{\beta} t^{\beta-1} e^{-(ct)^{\beta}}.$$

The reliability function is

$$\bar{F}(t) = e^{-(ct)^{\beta}}.$$

The failure rate function is given by

$$r(t) = \beta c^{\beta} t^{\beta-1}.$$

Because of its flexible shape and ability to model a wide range of failure rates, the Weibull has been used successfully in many applications.

2.1.3 Shapes of failure rate functions

The failure rate function is an important concept in reliability. Failure rate functions often falling into one of three categories are considered: (a) monotonic failure rates, where the failure rate curve is either increasing or decreasing ; (b) bathtub failure rates, where the curve has a bathtub or a U shape; and (c) generalized bathtub failure rates, where the failure rate curve is a polynomial, or has roller-coaster shape or some other generalization. Many lifetime distributions may be categorized with respect to the shape of their failure rate functions:

Definition 1. *Let $r(t)$ be the failure rate function of a lifetime distribution. It is* *(i) an IFR (increasing failure rate) distribution if* $r(t)$ *is nondecreasing in* t*;*

(ii) a DFR (decreasing failure rate) distribution if $r(t)$ *is nonincreasing in* t*;*

(iii) a BT (bathtub-shaped) distribution if there exits a $t_0 > 0$ *such that* $r(t)$ *is nonincreasing for* $0 \le t \le t_0$ *and nondecreasing for* $t \ge t_0$*;*

(iv) an UBT (upside-down bathtub-shaped) distribution if there exits a $t_0 > 0$ *such that* $r(t)$ *is nondecreasing for* $0 \le t \le t_0$ *and nonincreasing for* $t \ge t_0$*.*

Remark. Since the IFR and DFR can be viewed as special cases of the BT or UBT, to distinguish these special cases from the general cases, we use the words "degenerate" and "nondegenerate". For example, a degenerate BT distribution is either an IFR or a DFR distribution; a nondegenerate BT distribution has a complete bathtub-shaped failure rate.

If $r(t)$ is continuous and differentiable on $[0, \infty)$, the following provide sufficient conditions for classifying the shape of failure rate function.

(a) If $r'(t) = 0$ for all $t \ge 0$, then the lifetime T follows an exponential distribution with a constant failure rate.

(b) If $r'(t) \ge 0$ for all $t \ge 0$, then the distribution of T has an increasing failure rate.

(c) If $r'(t) \le 0$ for all $t \ge 0$, then the distribution of T has a decreasing failure rate.

(d) Suppose there exists $t_0 > 0$ such that $r'(t) \le 0$ for all $0 \le t < t_0$, $r'(t_0) = 0$ and $r'(t) \ge 0$ for all $t > t_0$, then the distribution of T has a bathtub-shaped failure rate.

(e) Suppose there exists $t_0 > 0$ such that $r'(t) \ge 0$ for all $0 \le t < t_0$, $r'(t_0) = 0$ and $r'(t) \le 0$ for all $t > t_0$, then the distribution of T has an upside-down bathtub-shaped failure rate.

Glaser [1980] provided some sufficient conditions for using the function $\eta(t) = -f'(t)/f(t)$ to help identify the shape of failure rate functions. In particular, it has been useful in classifying distributions, such as gammas, whose failure rate functions are not of a simple form.

Theorem 2.1.1. *[Glaser [1980]] Assume that* $f(t)$ *is continuous and twice differentiable on* $(0, \infty)$*. Define* $l(t) = 1/r(t) = \bar{F}(t)/f(t)$ *and* $\eta(t) = -f'(t)/f(t)$*.*

(a) If $\eta'(t) > 0$ *for all* $t > 0$*, then the distribution of* T *has an increasing failure rate.*

(b) If $\eta'(t) < 0$ *for all* $t > 0$*, then the distribution of* T *has a decreasing failure rate.*

(c) Suppose there exists $t_0 > 0$ such that $\eta'(t) < 0$ for all $t \in (0, t_0)$, $\eta'(t_0) = 0$ and $\eta'(t) > 0$ for all $t > t_0$.

(i) If there exists $y_0 > 0$ such that $l'(y_0) = 0$, then the distribution of T has a bathtub-shaped failure rate.

(ii) If there does not exist $y_0 > 0$ such that $l'(y_0) = 0$, then the distribution of T has an increasing failure rate.

(c) Suppose there exists $t_0 > 0$ such that $\eta'(t) > 0$ for all $t \in (0, t_0)$, $\eta'(t_0) = 0$ and $\eta'(t) < 0$ for all $t > t_0$.

(i) If there exists $y_0 > 0$ such that $l'(y_0) = 0$, then the distribution of T has an upside-down bathtub-shaped failure rate.

(ii) If there does not exist $y_0 > 0$ such that $l'(y_0) = 0$, then the distribution of T has a decreasing failure rate.

We apply Theorem 2.1.1 to a gamma distribution and derive the conditions for IFR or DFR.

Example 1. Let $f(t)$ be the density function of a gamma distribution with shape parameter α and scale parameter λ:

$$f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}$$

and

$$\eta(t) = -\frac{f'(t)}{f(t)} = \lambda - \frac{\alpha - 1}{t}.$$

Therefore,

$$\eta'(t) = \frac{\alpha - 1}{t^2}.$$

If $\alpha > 1$, $\eta'(t) > 0$ for all $t > 0$ and it is an IFR distribution. If $\alpha < 1$, $\eta'(t) < 0$ for all $t > 0$ and it is a DFR distribution. If $\alpha = 1$, it is an exponential distribution. The failure rate is then the constant λ.

2.2 DISTRIBUTIONS OF MIXTURES

When there are at least two subpopulations, the lifetime of the entire population can be modeled by a mixture. Consider a family of lifetime distributions of subpopulations with densities $\{f_\omega : \omega \in \Omega\}$, where $(\Omega, \mathcal{F}, P)$ is a probability space with $f_\omega(t)$ a measurable function of (ω, t). The density and survival functions of the mixture are given by

$$f(t) = \int_\Omega f_\omega(t) P(\,d\omega)$$

and

$$\bar{F}(t) = \int_\Omega \bar{F}_\omega(t) P(\,d\omega)$$

respectively. The failure rate functions of the mixture is given by

$$r(t) = \frac{f(t)}{\bar{F}(t)} = \frac{\int_\Omega f_\omega(t) P(\,d\omega)}{\int_\Omega \bar{F}_\omega(t) P(\,d\omega)}.$$

When mixing two subpopulations, the density function of the mixture is more simply given by

$$f(t) = pf_1(t) + qf_2(t)$$

and the survival function is by

$$\bar{F}(t) = p\bar{F}_1(t) + q\bar{F}_2(t),$$

where $p > 0$, $q > 0$ and $p + q = 1$. Here p and q are the "weights" of the corresponding subpopulations. The failure rate function of the mixture is

$$r(t) = \frac{f(t)}{\bar{F}(t)} = \frac{pf_1(t) + qf_2(t)}{p\bar{F}_1(t) + q\bar{F}_2(t)}.$$

If we set

$$\omega(t) = \frac{p\bar{F}_1(t)}{p\bar{F}_1(t) + q\bar{F}_2(t)},$$

then we can express $r(t)$ as

$$r(t) = \omega(t) r_1(t) + (1 - \omega(t)) r_2(t).$$

Example 2. [Lai and Xie [2006]] Mixing two DFR distributions having density functions $f_1(t)$ and $f_2(t)$. The first derivative of $r(t)$ is

$$r'(t) = \omega(t)r_1'(t) + (1-\omega(t))r_2'(t) + \omega'(t)\left[r_1(t) - r_2(t)\right].$$

Since

$$\begin{aligned}
\omega'(t) &= \left[\frac{p\bar{F}_1(t)}{p\bar{F}_1(t) + q\bar{F}_2(t)}\right]' \\
&= \frac{-pf_1(t)(p\bar{F}_1(t) + q\bar{F}_2(t)) + p\bar{F}_1(t)(pf_1(t) + qf_2(t))}{(p\bar{F}_1(t) + q\bar{F}_2(t))^2} \\
&= \frac{pq\bar{F}_1(t)f_2(t) - pqf_1(t)\bar{F}_2(t)}{(p\bar{F}_1(t) + q\bar{F}_2(t))^2} \\
&= -\frac{pq\bar{F}_1(t)\bar{F}_2(t)}{(p\bar{F}_1(t) + q\bar{F}_2(t))^2}\left(\frac{f_1(t)}{\bar{F}_1(t)} - \frac{f_2(t)}{\bar{F}_2(t)}\right) \\
&= -\omega(t)(1-\omega(t))\left[r_1(t) - r_2(t)\right],
\end{aligned}$$

thus

$$r'(t) = \omega(t)r_1'(t) + (1-\omega(t))r_2'(t) - \omega(t)(1-\omega(t))\left[r_1(t) - r_2(t)\right]^2.$$

Because the two sub-distributions have DFR, $r_1'(t) \leq 0$ and $r_2'(t) \leq 0$; and it is clear that $\omega(t) > 0$ and $1 - \omega(t) > 0$, thus $r'(t) \leq 0$, the mixture has DFR.

The function $\eta(t)$ is

$$\eta(t) = -\frac{f'(t)}{f(t)} = -\frac{pf_1'(t) + qf_2'(t)}{pf_1(t) + qf_2(t)}.$$

If $\eta(t)$ is differentiable on $[0, \infty)$, then

$$\eta'(t) = \frac{q^2A(t) + p^2B(t) + pqC(t)}{\left[pf_1(t) + qf_2(t)\right]^2},$$

where

$$A(t) = \left[f_2'(t)\right]^2 - f_2''(t)f_2(t),$$

$$B(t) = \left[f_1'(t)\right]^2 - f_1''(t)f_1(t)$$

and

$$C(t) = 2f_1'(t)f_2'(t) - f_1''(t)f_2(t) - f_1(t)f_2''(t).$$

Example 3. Mixing two exponential distributions with parameters λ_1 and λ_2, $\lambda_1 \neq \lambda_2$.

The density functions are

$$f_1(t) = \lambda_1 e^{-\lambda_1 t} \quad \text{and} \quad f_2(t) = \lambda_2 e^{-\lambda_2 t}.$$

Then $A(t) = 0$, $B(t) = 0$ and

$$C(t) = -\lambda_1 \lambda_2 (\lambda_1 - \lambda_2)^2 \exp\left[-(\lambda_1 + \lambda_2)t\right] < 0.$$

Thus $\eta(t) < 0$ for all t. According to Theorem 2.1.1, the mixture has DFR.

3.0 MIXTURES WITH BATHTUB-SHAPED FAILURE RATES

It was recognized by several authors that distributions with bathtub-shaped failure rates could arise as simple mixtures. For example, Gupta and Warren [2001] showed that a certain mixture of two gammas, one with an increasing failure rate (IFR) and the other with a decreasing failure rate (DFR) could have a failure rate which is bathtub. An even simpler example, given by Block, Li, and Savits [2003a], shows that a mixture of an exponential and an IFR gamma can have a bathtub-shaped failure rate. It turns out that both of these examples are a special case of a much more general result. Mixing an IFR gamma with a host of different DFR distributions turns out to have a failure rate with bathtub shape. In Section 3.1, we provide the theorem which shows that mixing an IFR gamma with a continuous mixture of exponentials yields a distribution with a bathtub-shaped failure rate. In Section 3.2, we consider mixing a continuous mixture of IFR gammas and a continuous mixture of exponentials.

3.1 MIXING AN IFR GAMMA AND A MIXTURE OF EXPONENTIALS

We list a few facts about the gamma distribution which are needed in the proof. The density of the gamma with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$ is given by

$$g(t|\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1}\exp(-\lambda t) \quad \text{for} \quad t > 0 \tag{3.1}$$

and we use the notation $g(t|\alpha)$ when $\lambda = 1$. We denote the survival functions by $\bar{G}(t|\alpha,\lambda)$ and $\bar{G}(t|\alpha)$ for the $\lambda = 1$ case. It is easy to check that

$$e^t \bar{G}(t|\alpha) = \frac{1}{\Gamma(\alpha)} \int_0^\infty (t+v)^{\alpha-1} e^{-v}\, dv$$

and

$$\frac{d}{dt}\left[e^t g(t|\alpha)\right] = e^t g(t|\alpha - 1), \tag{3.2}$$

$$\frac{d}{dt}\left[e^t \bar{G}(t|\alpha)\right] = e^t \bar{G}(t|\alpha - 1). \tag{3.3}$$

3.1.1 Main result

Theorem 3.1.1. *Consider a gamma distribution with density $g(t|\alpha, \lambda_0)$ with $\alpha > 2$ and $\lambda_0 > 0$ and a family of exponentials with parameters $\lambda > \lambda_0$. Let P be a probability measure whose support set S is a subset of (λ_0, ∞). The resulting mixture, with density*

$$f(t) = p \int \lambda \exp(-\lambda t) P(\,d\lambda) + q g(t|\alpha, \lambda_0) \tag{3.4}$$

where $p + q = 1$, $p > 0$, $q > 0$, has a bathtub-shaped failure rate. If P has a finite first moment, then the failure rate is nondegenerate bathtub.

Remark. The assumption on P that its support set $S \subset (\lambda_0, \infty)$ implies that

$$\inf \{x : P((-\infty, x]) > 0\} > \lambda_0.$$

Proof. It is enough to prove the theorem for $\lambda_0 = 1$ and then to rescale. The proof can be done for most parameter choices by directly examining the failure rate $r(t) = f(t)/\bar{F}(t)$ where $\bar{F}(t)$ is the survival function. However the proof is easiest and most complete using Theorem 2.1.1 which examines the function $\eta(t) = -f'(t)/f(t)$ for $t > 0$. In this case using the gamma notation in (3.1) and the differential equations (3.2) and (3.3) we find that

$$\eta(t) = \frac{p \int \lambda^2 \exp\left[-(\lambda - 1)t\right] P(\,d\lambda) + q e^t g(t|\alpha) - q e^t g(t|\alpha - 1)}{p \int \lambda \exp\left[-(\lambda - 1)t\right] P(\,d\lambda) + q e^t g(t|\alpha)},$$

and

$$\eta'(t) = \frac{q^2 A(t) + p^2 B(t) + pq C(t)}{\left\{p \int \lambda \exp\left[-(\lambda - 1)t\right] P(\,d\lambda) + q e^t g(t|\alpha)\right\}^2},$$

where

$$\begin{aligned} A(t) &= e^t \left[g(t|\alpha - 1) - g(t|\alpha - 2)\right] e^t g(t|\alpha) - e^t \left[g(t|\alpha) - g(t|\alpha - 1)\right] e^t g(t|\alpha - 1) \\ &= \frac{\alpha - 1}{\left[\Gamma(\alpha)\right]^2} t^{2(\alpha - 2)}, \end{aligned}$$

$$
\begin{aligned}
B(t) &= \int \lambda^2 \exp\left[-(\lambda-1)t\right] P(\,d\lambda) \int \xi(\xi-1) \exp\left[-(\xi-1)t\right] P(\,d\xi) \\
&\quad - \int \lambda^2(\lambda-1) \exp\left[-(\lambda-1)t\right] P(\,d\lambda) \int \xi \exp\left[-(\xi-1)t\right] P(\,d\xi) \\
&= -\int \left\{ \int_{[\xi,\infty)} \lambda\xi(\lambda-\xi)^2 \exp\left[-(\lambda+\xi-2)t\right] P(\,d\lambda) \right\} P(\,d\xi)
\end{aligned}
$$

and

$$
\begin{aligned}
C(t) &= -\int \lambda^2(\lambda-1) \exp\left[-(\lambda-1)t\right] P(\,d\lambda) e^t g(t|\alpha) \\
&\quad + \int \lambda \exp\left[-(\lambda-1)t\right] P(\,d\lambda) \left[e^t g(t|\alpha-1) - e^t g(t|\alpha-2)\right] \\
&\quad - \int \lambda^2 \exp\left[-(\lambda-1)t\right] P(\,d\lambda) e^t g(t|\alpha-1) \\
&\quad + \int \lambda(\lambda-1) \exp\left[-(\lambda-1)t\right] P(\,d\lambda) \left[e^t g(t|\alpha) - e^t g(t|\alpha-1)\right] \\
&= -\int \lambda(\lambda-1)^2 \exp\left[-(\lambda-1)t\right] P(\,d\lambda) e^t g(t|\alpha) \\
&\quad -2 \int \lambda(\lambda-1) \exp\left[-(\lambda-1)t\right] P(\,d\lambda) e^t g(t|\alpha-1) \\
&\quad - \int \lambda \exp\left[-(\lambda-1)t\right] P(\,d\lambda) e^t g(t|\alpha-2).
\end{aligned}
$$

Note that for all $t > 0$, $A(t) \geq 0$, $B(t) \leq 0$ and $C(t) \leq 0$.

To show that η is bathtub-shaped, we need only show that the function

$$h(t) = q^2 A(t) + p^2 B(t) + pq C(t)$$

has only one sign change and it is from negative to positive. We first consider the sign of

$$h(0+) = q^2 A(0+) + p^2 B(0+) + pq C(0+).$$

For $\alpha > 2$, $A(0+) = 0$ and so $h(0+) \leq 0$. If P is nondegenerate, then

$$h(0+) \leq p^2 B(0+) = -p^2 \int \left\{ \int_{[\xi,\infty)} \lambda\xi(\lambda-\xi)^2 P(\,d\lambda) \right\} P(\,d\xi) < 0.$$

If P is degenerate at $\lambda_1 > \lambda_0 = 1$, then $B(t) = 0$ and so $h(t)$ behaves like $q^2 A(t) + pqC(t)$ as $t \downarrow 0$, where

$$\begin{aligned}
&\Gamma(\alpha)C(t) \\
= \quad & -\exp\left[-(\lambda_1 - 1)t\right]\left\{\lambda_1(\lambda_1 - 1)^2 t^{\alpha-1} + 2(\alpha-1)\lambda_1(\lambda_1 - 1)t^{\alpha-2} + \right. \\
& \qquad \left. (\alpha-1)(\alpha-2)\lambda_1 t^{\alpha-3}\right\}.
\end{aligned}$$

For $2 < \alpha < 3$, $C(0+) = -\infty$; for $\alpha = 3$, $C(0+) = -(\alpha-1)(\alpha-2)\lambda_1/\Gamma(\alpha)$. Thus we conclude that $h(0+) < 0$ when $2 < \alpha \leq 3$. For $\alpha > 3$, however, $h(0+) = 0$ since $C(0+) = 0$. In this case, we can write $h(t)$ as

$$\begin{aligned}
& h(t) \\
= \quad & t^{\alpha-3}\left\{q^2 \frac{\alpha-1}{[\Gamma(\alpha)]^2} t^{\alpha-1} - pq\frac{e^{-(\lambda_1-1)t}}{\Gamma(\alpha)}\left[\lambda_1(\lambda_1-1)^2 t^2 + 2(\alpha-1)\lambda_1(\lambda_1-1)t + \lambda_1(\alpha-1)(\alpha-2)\right]\right\}.
\end{aligned}$$

Consequently $\frac{h(t)}{t^{\alpha-3}} \to -pq(\alpha-1)(\alpha-2)\lambda_1/\Gamma(\alpha)$ as $t \downarrow 0$ and so $h(t) < 0$ in a neighborhood of $t = 0+$. Thus we conclude that $h(t) < 0$ for small values of $t > 0$ for all $\alpha > 2$.

Also, since the support S of P is by definition a closed set, it does not contain a neighborhood of 1 and so it follows that $B(t) \to 0$ and $C(t) \to 0$ as $t \to \infty$. Since $A(t) \to \infty$ as $t \to \infty$ for $\alpha > 2$, we conclude that $h(t) > 0$ for large t and consequently there exists at least one positive root of $h(t) = 0$.

To show there is only one such root, it suffices to show that if $h(t_0) = 0$ for $0 < t_0 < \infty$, then $h'(t_0) > 0$. So let

$$0 = h(t_0) = q^2 A(t_0) + p^2 B(t_0) + pqC(t_0)$$

and consider

$$h'(t_0) = q^2 A'(t_0) + p^2 B'(t_0) + pqC'(t_0).$$

Using the fact that

$$q^2 A'(t_0) = q^2 \frac{2(\alpha-1)(\alpha-2)}{\Gamma^2(\alpha)} t_0^{2\alpha-5} = q^2 \frac{2(\alpha-2)}{t_0} A(t_0)$$

and

$$q^2 A(t_0) = -p^2 B(t_0) - pqC(t_0),$$

it follows after some simplification that

$$
\begin{aligned}
h'(t_0) &= \frac{2(\alpha-2)}{t_0}\left[-p^2 B(t_0)\right] \\
&+p^2 \int \left\{ \int_{[\xi,\infty)} \lambda\xi(\lambda-\xi)^2(\lambda+\xi-2)\exp\left[-(\lambda+\xi-2)t_0\right] P(\,d\lambda) \right\} P(\,d\xi) \\
&+\frac{pq}{\Gamma(\alpha)} t_0^{\alpha-1} \int \lambda(\lambda-1)^3 \exp\left[-(\lambda-1)t_0\right] P(\,d\lambda) \\
&+\frac{pq}{\Gamma(\alpha)}(3\alpha-5) t_0^{\alpha-2} \int \lambda(\lambda-1)^2 \exp\left[-(\lambda-1)t_0\right] P(\,d\lambda) \\
&+\frac{pq}{\Gamma(\alpha)} 3(\alpha-1)(\alpha-2) t_0^{\alpha-3} \int \lambda(\lambda-1) \exp\left[-(\lambda-1)t_0\right] P(\,d\lambda) \\
&+\frac{pq}{\Gamma(\alpha)}(\alpha-1)^2(\alpha-2) t_0^{\alpha-4} \int \lambda \exp\left[-(\lambda-1)t_0\right] P(\,d\lambda).
\end{aligned}
$$

Hence $h'(t_0) > 0$.

As noted previously, we can thus conclude that $h(t) < 0$ for $0 < t < t_0$ and $h(t) > 0$ for $t > t_0$. This implies that $\eta(t)$ is bathtub-shaped. According to Theorem 2.1.1, the failure rate is thus either increasing or nondegenerate bathtub. Suppose now that P has a finite first moment. Then from Remark 2.4 of Block, Li, and Savits [2003a], we deduce that

$$
r'(0+) = -p\int \lambda^2 P(\,d\lambda) + p^2\left[\int \lambda P(\,d\lambda)\right]^2 < 0.
$$

Hence $r(t)$ must be nondegenerate bathtub.

Remarks.

(i) The results of Block, Li, and Savits [2003a] allow us (under certain conditions) to conclude that the failure rate of the mixture (3.4) initially decreases and then ultimately increases to λ_0; however, this is not enough to conclude that $r(t)$ has a bathtub shape.

(ii) We can also show that the failure rate is nondegenerate bathtub if P is nondegenerate with support a subset of $(2\lambda_0, \infty)$.

3.1.2 Applications of main result

In Section 3.1.1, we considered a mixture of the form

$$f(t) = ph(t) + qg(t)$$

where $g(t) = g(t|\alpha, \lambda_0)$ was the density of a gamma distribution with shape parameter $\alpha > 2$ and scale parameter $\lambda_0 > 0$, and $h(t) = \int \lambda \exp(-\lambda t) P(\,d\lambda)$ was a continuous mixture of exponential densities with P being a probability distribution having support a subset of (λ_0, ∞).

We note that the survival function $\bar{H}(t)$ of $h(t)$ is given by

$$\bar{H}(t) = \int_{(\lambda_0, \infty)} \exp(-\lambda t) P(\,d\lambda). \tag{3.5}$$

According to Feller [1966] (Theorem 1, p.415), it follows that $\bar{H}(t)$ must be a completely monotone function (i.e. $(-1)^n \frac{d^n}{dt^n} \bar{H}(t) \geq 0$ for $t > 0$). Conversely, if a survival function $\bar{H}(t)$ is completely monotone, then there is a unique probability measure P on $[0, \infty)$ such that the representation (3.5) holds.

Thus we can reformulate Theorem 3.1.1 as follows:

Theorem 3.1.2. *Any survival function $\bar{F}(t)$ of the form*

$$\bar{F}(t) = p\bar{H}(t) + q\bar{G}(t|\alpha, \lambda_0),$$

where $p + q = 1$, $p > 0$, $q > 0$ $\lambda_0 > 0$, $\alpha > 2$ and $\bar{H}(t)$ is a completely monotone function whose associated probability measure P in the representation (3.5) has support a subset of (λ_0, ∞), has a bathtub-shaped failure rate (may be degenerate). If, in addition $-\bar{H}'(0+) = \int \lambda P(\,d\lambda) < \infty$, then it is nondegenerate bathtub.

Our main results in this section are applications of Theorem 3.1.2 and the following simple lemma.

Lemma 3.1.3. *Let* (Θ, Q) *be a probability space and* $\{\phi_\theta : \theta \in \Theta\}$ *be a family of completely monotone functions jointly measurable in* $(\theta, t) \in [0, \infty)$*. Then*

$$\phi(t) = \int_\Theta \phi_\theta(t) Q(\,d\theta)$$

is completely monotone.

Proof. By (3.5), we know that for every $\theta \in \Theta$, there exists a unique probability measure P_θ on $[0, \infty)$ such that

$$\phi_\theta(t) = \int_{[0,\infty)} \exp(-\lambda t) P_\theta(\,d\lambda).$$

Hence,

$$\phi(t) = \int_\Theta \left[\int_{[0,\infty)} \exp(-\lambda t) P_\theta(\,d\lambda) \right] Q(\,d\theta) = \int_{[0,\infty)} \exp(-\lambda t) R(\,d\lambda) \tag{3.6}$$

where R is a probability measure on $[0, \infty)$ given by

$$R(A) = \int_\Theta P_\theta(A) Q(\,d\theta)$$

for all Borel subsets A of $[0, \infty)$ (c.f., Meyer, 1966, Theorem T16, p.16). Thus the representation (3.6) shows that ϕ is completely monotone.

Our results in this section involve specific mixtures which give rise to bathtub distributions. These results are of the following types:

1) mixtures of DFR gammas with an IFR gamma are bathtub;
2) mixtures of DFR "Weibulls" with IFR gamma are bathtub;
3) mixtures of certain bathtub distributions are bathtub.

The above are quite remarkable since recent research has shown that mixtures of distributions with even the simplest failure rate functions can lead to vastly different monotonic behavior. For example, in Block, Savits, and Wondmagegnehu [2003b] it was shown that the mixture of two distributions with increasing linear failure rates can have four changes of monotonicity. Moreover, results concerning the behavior of mixtures of more than two distributions are very sparse in the literature, while the above results involve possibly continuous mixtures.

We first show that arbitrary mixtures of DFR gamma distributions with an IFR gamma distribution have a bathtub-shaped failure rate.

Theorem 3.1.4. *Consider an arbitrary mixture of DFR gamma densities* $g(t|\beta, \xi)$ *with* $0 < \beta \leq 1$, $\xi > 0$ *and let* $\lambda_0 > 0$. *Let* Q *be a probability measure on* $\boldsymbol{R}^2$ *with support a subset of* $(0, 1] \times (\lambda_0, \infty)$. *Then the mixture distribution*

$$f(t) = p \int_{\boldsymbol{R}^2} g(t|\beta, \xi)\, dQ(\beta, \xi) + qg(t|\alpha, \lambda_0) \tag{3.7}$$

has a bathtub-shaped failure rate for $\alpha > 2$.

Proof. According to Gleser [1989], any DFR gamma is a mixture of exponentials, i.e., if $0 < \beta \leq 1$ and $\xi > 0$, then the survival function can be written as

$$\bar{G}(t|\beta, \xi) = \int_{[0,\infty)} \exp(-\lambda t) P_{(\beta,\xi)}(\, d\lambda),$$

where $P_{(\beta,\xi)}$ is a probability measure on $[0, \infty)$ having density

$$p_{(\beta,\xi)}(\lambda) = \frac{(\lambda - \xi)^{-\beta} \xi^{\beta}}{\lambda \Gamma(1 - \beta)\Gamma(\beta)} I_{[\xi,\infty)}(\lambda).$$

But from our Lemma 3.1.3, it follows that we can write

$$\bar{G}(t) = \int_{\mathbf{R}^2} \bar{G}(t|\beta, \xi)\, dQ(\beta, \xi) = \int_{[0,\infty)} \exp(-\lambda t)\, dR(\lambda),$$

where $R(A) = \int_{\mathbf{R}^2} P_{(\beta,\xi)}(A)\, dQ(\beta, \xi)$ for all Borel subsets $A \subset [0, \infty)$. Using our assumptions on Q, it is not hard to show that R has support which is a subset of (λ_0, ∞). The result then follows from Theorem 3.1.2.

Remark. Since the probability measure R above does not have a finite first moment, we cannot use Theorem 3.1.2 to conclude that the failure rate is nondegenerate bathtub. However, if the support of Q is a subset of $(0, 1] \times (2\lambda_0, \infty)$, it follows that the support of R is a subset of $(2\lambda_0, \infty)$. Since R is clearly nondegenerate, we conclude from the Remark following Theorem 3.1.1 that the failure rate of the mixture is nondegenerate bathtub.

Our next result is similar to Theorem 3.1.4 but with DFR 'Weibull' distributions. These 'Weibull' distributions we consider are a variant of the usual Weibull since the failure rate of the usual nondegenerate (i.e., not exponential) DFR Weibull decreases to zero. The variant we consider has the survival function ($\xi > 0$)

$$\bar{H}(t|\beta, c, \xi) = e^{-\xi t} \bar{W}(t|\beta, c) \tag{3.8}$$

where $\xi > 0$ and $\bar{W}(t|\beta, c)$ is the survival function of the usual DFR Weibull distribution with scale parameter $c > 0$ and shape parameter β with $0 < \beta < 1$, i.e.,

$$\bar{W}(t|\beta, c) = \exp\left[-(ct)^{\beta}\right], \quad t > 0. \tag{3.9}$$

For this variant of the Weibull distribution, the failure rate decreases to ξ as $t \to \infty$. We call this a 'Weibull' distribution with parameter (β, c, ξ).

Theorem 3.1.5. *Consider any family of DFR 'Weibull' distributions of the form* (3.8) *and let* $\lambda_0 > 0$. *Let* Q *be a probability measure on* $\boldsymbol{R}^3$ *with support a subset of* $(0, 1] \times (0, \infty) \times (\lambda_0, \infty)$. *Then the mixture with survival function*

$$\bar{F}(t) = p \int_{\mathbf{R}^3} \bar{H}(t|\beta, c, \xi)\, dQ(\beta, c, \xi) + q\bar{G}(t|\alpha, \lambda_0)$$

has a bathtub shaped failure rate for $\alpha > 2$

Proof. Jewell [1982] claims that any DFR Weibull distribution is a mixture of exponentials. Thus for $\bar{W}$ given in (3.9), we can write

$$\bar{W}(t|\beta, c) = \int_{[0,\infty)} \exp(-\lambda t) P_{(\beta,c)}(\, d\lambda),$$

for some probability measure $P_{(\beta,c)}$ on $[0, \infty)$. Hence

$$\bar{H}(t|\beta, c, \xi) = \int_{[0,\infty)} \exp(-(\lambda + \xi)t) P_{(\beta,c)}(\, d\lambda) = \int \exp(-\gamma t) P_{(\beta,c,\xi)}(\, d\gamma)$$

where $P_{(\beta,c,\xi)}$ is the shifted version of $P_{(\beta,c)}$, i.e., for every bounded Borel function ψ on $[0, \infty)$,

$$\int_{[0,\infty)} \psi(\gamma) P_{(\beta,c,\xi)}(\, d\gamma) = \int_{[0,\infty)} \psi(\lambda + \xi) P_{(\beta,c)}(\, d\lambda).$$

Note that the support of $P_{(\beta,c,\xi)}$ is contained in $[\xi, \infty)$. By our Lemma 3.1.3 we can write

$$\bar{H}(t) = \int_{\mathbf{R}^3} \bar{H}(t|\beta, c, \xi)\, dQ(\beta, c, \xi) = \int_{[0,\infty)} \exp(-\gamma t) R(\, d\gamma)$$

with

$$R(A) = \int_{\mathbf{R}^3} P_{(\beta,c,\xi)}(A)\, dQ(\beta, c, \xi).$$

Hence the result again follows from Theorem 3.1.2 since R has support which is a subset of (λ_0, ∞).

Remark. As noted in the proof, Jewell [1982] claims without proof that a DFR Weibull survival function is completely monotone. The easiest proof that we know is a nice application of Faa di Bruno and we include it here. It suffices to consider the case $c = 1$ and $0 < \beta \leq 1$ and let $\phi(t) = \exp(-t^\beta)$, $t > 0$. We can write ϕ as a composition $\phi(t) = a\,[b(t)]$ where $a(t) = e^{-x}$ and $b(t) = t^\beta$. According to Faa Di Bruno's formulas (cf. Constantine and Savits [1996])

$$\frac{d^n}{dt^n}\phi(t) = \sum_{k=1}^{n} \frac{d^k}{dx^k} a\,[b(t)] \sum_{p(n,k)} n! \prod_{i=1}^{n} \frac{\left[\frac{d^i}{dt^i} b(t)\right]^{\lambda_i}}{(\lambda_i)!(i!)^{\lambda_i}}$$

where $p(n,k) = \{(\lambda_1, \ldots, \lambda_n) : \lambda_i \geq 0, \sum_{i=1}^{n} \lambda_i = k, \sum_{i=1}^{n} i\lambda_i = n\}$. Since

$$\frac{d^k}{dx^k} a\,[b(t)] = (-1)^k \phi(t),$$

to show that $(-1)^n \frac{d^n}{dt^n}\phi(t) \geq 0$ for $n \geq 1$, $t > 0$, it suffices to show that each term

$$(-1)^{n+k} \prod_{i=1}^{n} \frac{\left[\frac{d^i}{dt^i} b(t)\right]^{\lambda_i}}{(\lambda_i)!(i!)^{\lambda_i}} \geq 0.$$

But

$$(-1)^{i+1} \frac{d^i}{dt^i} b(t) = (-1)^{i+1} \beta(\beta - 1) \ldots (\beta - i + 1) t^{\beta - 1} \geq 0$$

for $0 < \beta \leq 1$. Since $(\lambda_1, \ldots, \lambda_n) \in p(n,k)$,

$$\sum_{i=1}^{n} (i+1)\lambda_i = n + k$$

and we are done.

Finally we have a closure result for any bathtub distributions which arise in any of the above manners (i.e., from previous Theorems)

Theorem 3.1.6. *The mixture of any bathtub distributions which arise as a mixture of exponentials, DFR gammas or DFR 'Weibulls' with the same IFR gamma having shape parameter $\alpha > 2$ has a bathtub-shaped distribution.*

Proof. Since this will be a mixture of exponentials with an IFR gamma having shape parameter $\alpha > 2$, it will have a bathtub shape.

3.1.3 The $\alpha = 2$ case

The result for the special case when $\alpha = 2$ is similar to Theorem 3.1.1, but a slightly different proof is required.

Theorem 3.1.7. *Consider a gamma distribution with density $g(t|\alpha, \lambda_0)$ with $\alpha = 2$ and $\lambda_0 > 0$ and a family of exponentials with parameters $\lambda > \lambda_0$. Let P be a probability measure whose support set S is a subset of (λ_0, ∞). The resulting mixture, with density*

$$f(t) = p \int \lambda \exp(-\lambda t) P(\,d\lambda) + qg(t|2, \lambda_0) \tag{3.10}$$

where $p + q = 1$, $p > 0$, $q > 0$, has a bathtub-shaped failure rate $r(t)$. If P has a finite first moment, then the failure rate is nondegenerated bathtub if

$$p \int \lambda^2 P(\,d\lambda) - p^2 \left[\int \lambda P(\,d\lambda)\right]^2 > q;$$

otherwise, it is increasing.

Proof. Without loss of generality, we need only consider the case $\lambda_0 = 1$. As in the proof of Theorem 3.1.1, we show that $\eta(t)$ is bathtub, or equivalently, that $h(t)$ is bathtub. In this case, $h(t) = q^2 A(t) + p^2 B(t) + pqC(t)$, where

$$\begin{aligned}
A(t) &= 1, \\
B(t) &= -\int \left\{ \int_{[\xi,\infty)} \lambda\xi(\lambda - \xi)^2 \exp\left[-(\lambda + \xi - 2)t\right] P(\,d\lambda) \right\} P(\,d\xi), \\
C(t) &= -\left\{ t \int \lambda(\lambda - 1)^2 \exp\left[-(\lambda - 1)t\right] P(\,d\lambda) + 2 \int \lambda(\lambda - 1) \exp\left[-(\lambda - 1)t\right] P(\,d\lambda) \right\}.
\end{aligned}$$

It is easy to see that $h(t) \to q^2$ as t increases to ∞, and thus $h(t) > 0$ for large t. Also, as in the proof of Theorem 3.1.1, one can show that if $h(t_0) = 0$ for some $0 < t_0 < \infty$, then $h'(t_0) > 0$. Consequently, $h(t)$ can have at most one zero. Moreover, if it has a zero, the sign of $h(t)$ goes from $-$ to $+$. We thus conclude that $\eta(t)$ is bathtub. By Theorem 2.1.1, it follows that the failure rate is either increasing or nondegenerate bathtub.

Suppose now that P has a finite first moment. In this case, we can examine the value of $r'(0+)$:

$$r'(0+) = -p \int \lambda^2 P(\,d\lambda) + p^2 \left[\int \lambda P(\,d\lambda)\right]^2 + q.$$

Thus the failure rate is increasing if $r'(0+) \geq 0$, while it is nondegenerate bathtub if $r'(0+) < 0$.

3.2 MIXING A MIXTURE OF IFR GAMMAS AND A MIXTURE OF EXPONENTIALS

Mixtures of IFR gammas are not generally IFR. Gupta and Warren [2001] have given an example of a mixture of two IFR gammas where the failure rate has two changes of monotonicity. Mixing this with a mixture of exponentials would most certainly give a distribution which does not have a bathtub shape. However, if we carefully select a class of IFR gammas, a bathtub-shaped failure rate distribution still may be obtained.

Theorem 3.2.1. *Consider a family of gamma distributions with density $g(t|\alpha, \lambda_0)$ with $\lambda_0 > 0$ and a family of exponentials with parameters $\lambda > \lambda_0$. Let Q be a probability measure whose support set S_Q is a subset of (α_0, α_0+1) with $\alpha_0 \geq 7/3$ and P be a probability measure whose support set S_P is a subset of (λ_0, ∞). The resulting mixture, with density*

$$f(t) = p \int \lambda \exp(-\lambda t) P(\,d\lambda) + q \int g(t|\alpha, \lambda_0) Q(\,d\alpha) \tag{3.11}$$

where $p + q = 1$, $p > 0$, $q > 0$, has a bathtub-shaped failure rate. If P has a finite first moment, then the failure rate is nondegenerate bathtub.

Proof. The proof we give here is similar to the proof of Theorem 3.1.1. We prove the theorem for $\lambda_0 = 1$ since it can be rescaled for others. It is clear that

$$\frac{d}{dt}\left[e^t \int g(t|\alpha) Q(\,d\alpha)\right] = e^t \int g(t|\alpha - 1) Q(\,d\alpha), \tag{3.12}$$

The function $\eta(t) = -f'(t)/f(t)$ then can be written as

$$\eta(t) = \frac{p \int \lambda^2 \exp\left[-(\lambda-1)t\right] P(\,d\lambda) + qe^t \int g(t|\alpha) Q(\,d\alpha) - qe^t \int g(t|\alpha-1) Q(\,d\alpha)}{p \int \lambda \exp\left[-(\lambda-1)t\right] P(\,d\lambda) + qe^t \int g(t|\alpha)} Q(\,d\alpha).$$

The first derivative of $\eta(t)$ is given by

$$\eta'(t) = \frac{q^2 A(t) + p^2 B(t) + pqC(t)}{\left\{p \int \lambda \exp\left[-(\lambda-1)t\right] P(\,d\lambda) + qe^t \int g(t|\alpha) Q(\,d\alpha)\right\}^2},$$

where

$$
\begin{aligned}
A(t) &= e^t \left[\int g(t|\alpha-1)Q(\,d\alpha) - \int g(t|\alpha-2)Q(\,d\alpha) \right] e^t \int g(t|\alpha)Q(\,d\alpha) \\
&\quad - e^t \left[\int g(t|\alpha)Q(\,d\alpha) - \int g(t|\alpha-1)Q(\,d\alpha) \right] e^t \int g(t|\alpha-1)Q(\,d\alpha) \\
&= \int\int \frac{t^{\alpha_1+\alpha_2-4}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} (\alpha_1-\alpha_2+1)(\alpha_2-1)Q(\,d\alpha_1)Q(\,d\alpha_2),
\end{aligned}
$$

$$
\begin{aligned}
B(t) &= \int \lambda^2 \exp\left[-(\lambda-1)t\right] P(\,d\lambda) \int \xi(\xi-1)\exp\left[-(\xi-1)t\right] P(\,d\xi) \\
&\quad - \int \lambda^2(\lambda-1)\exp\left[-(\lambda-1)t\right] P(\,d\lambda) \int \xi \exp\left[-(\xi-1)t\right] P(\,d\xi) \\
&= -\int \left\{ \int_{[\xi,\infty)} \lambda\xi(\lambda-\xi)^2 \exp\left[-(\lambda+\xi-2)t\right] P(\,d\lambda) \right\} P(\,d\xi)
\end{aligned}
$$

and

$$
\begin{aligned}
C(t) &= -\int \lambda^2(\lambda-1)\exp\left[-(\lambda-1)t\right] P(\,d\lambda) \int e^t g(t|\alpha)Q(\,d\alpha) \\
&\quad + \int \lambda \exp\left[-(\lambda-1)t\right] P(\,d\lambda) \left[\int e^t g(t|\alpha-1)Q(\,d\alpha) - \int e^t g(t|\alpha-2)Q(\,d\alpha) \right] \\
&\quad - \int \lambda^2 \exp\left[-(\lambda-1)t\right] P(\,d\lambda) \int e^t g(t|\alpha-1)Q(\,d\alpha) \\
&\quad + \int \lambda(\lambda-1)\exp\left[-(\lambda-1)t\right] P(\,d\lambda) \left[\int e^t g(t|\alpha)Q(\,d\alpha) - \int e^t g(t|\alpha-1)Q(\,d\alpha) \right] \\
&= -\int \lambda(\lambda-1)^2 \exp\left[-(\lambda-1)t\right] P(\,d\lambda) \int e^t g(t|\alpha)Q(\,d\alpha) \\
&\quad -2\int \lambda(\lambda-1)\exp\left[-(\lambda-1)t\right] P(\,d\lambda) \int e^t g(t|\alpha-1)Q(\,d\alpha) \\
&\quad - \int \lambda \exp\left[-(\lambda-1)t\right] P(\,d\lambda) \int e^t g(t|\alpha-2)Q(\,d\alpha) \\
&= -\int\int \frac{\lambda t_{\alpha-3}\exp\left[-(\lambda-1)t\right]}{\Gamma(\alpha)} \left[(\lambda-1)^2t^2 + 2(\lambda-1)(\alpha-1)t \right. \\
&\qquad \left. +(\alpha-1)(\alpha-2)\right] P(\,d\lambda)Q(\,d\alpha).
\end{aligned}
$$

Since the support set of the probability measure Q is a subset of (α_0, α_0+1), $(\alpha_1-\alpha_2+1)(\alpha_2-1) > 0$. Therefore $A(t) > 0$. Meanwhile, it is easy to check that $B(t) < 0$ and $C(t) < 0$.

Next we show that the function

$$h(t) = q^2 A(t) + p^2 B(t) + pqC(t)$$

has only one sign change and it is from negative to positive. Consider the sign of

$$h(0+) = q^2 A(0+) + p^2 B(0+) + pqC(0+).$$

For $\alpha_1, \alpha_2 \in (\alpha_0, \alpha_0 + 1)$ with $\alpha_0 \geq 7/3$, $A(0+) = 0$ and so $h(0+) \leq 0$. If P is nondegenerate, then

$$h(0+) \leq p^2 B(0+) = -p^2 \int \left\{ \int_{[\xi,\infty)} \lambda\xi(\lambda - \xi)^2 P(\,d\lambda) \right\} P(\,d\xi) < 0.$$

If P is degenerate at $\lambda_1 > \lambda_0 = 1$, then $B(t) = 0$ and so $h(t)$ behaves like $q^2 A(t) + pqC(t)$ as $t \downarrow 0$, where

$$\begin{aligned} C(t) \;&=\; -\int \frac{\lambda_1 \exp\left[-(\lambda_1 - 1)t\right]}{\Gamma(\alpha)} \left[(\lambda_1 - 1)^2 t^{\alpha-1} + 2(\lambda_1 - 1)(\alpha - 1)t^{\alpha-2} \right. \\ &\qquad \left. + (\alpha - 1)(\alpha - 2)t^{\alpha-3}\right] Q(\,d\alpha). \end{aligned}$$

For a specific α, if $7/3 \leq \alpha_0 < \alpha < 3$, as $t \downarrow 0$,

$$(\lambda_1 - 1)^2 t^{\alpha-1} + 2(\lambda_1 - 1)(\alpha - 1)t^{\alpha-2} + (\alpha - 1)(\alpha - 2)t^{\alpha-3} \to \infty.$$

If $\alpha = 3$, as $t \downarrow 0$,

$$(\lambda_1 - 1)^2 t^{\alpha-1} + 2(\lambda_1 - 1)(\alpha - 1)t^{\alpha-2} + (\alpha - 1)(\alpha - 2)t^{\alpha-3} \to (\alpha - 1)(\alpha - 2).$$

If $\alpha > 3$, however, as $t \downarrow 0$,

$$(\lambda_1 - 1)^2 t^{\alpha-1} + 2(\lambda_1 - 1)(\alpha - 1)t^{\alpha-2} + (\alpha - 1)(\alpha - 2)t^{\alpha-3} \to 0.$$

But

$$\frac{(\lambda_1 - 1)^2 t^{\alpha-1} + 2(\lambda_1 - 1)(\alpha - 1)t^{\alpha-2} + (\alpha - 1)(\alpha - 2)t^{\alpha-3}}{t^{\alpha-3}} \to (\alpha - 1)(\alpha - 2).$$

Therefore

$$(\lambda_1 - 1)^2 t^{\alpha-1} + 2(\lambda_1 - 1)(\alpha - 1)t^{\alpha-2} + (\alpha - 1)(\alpha - 2)t^{\alpha-3} > 0$$

in a neighborhood of $t = 0+$. So $C(t) < 0$ as $t \downarrow 0$ and thus we conclude $h(0+) < 0$.

Since the support S_P of P is by definition a closed set, it does not contain a neighborhood of 1 and so it follows that $B(t) \to 0$ and $C(t) \to 0$ as $t \to \infty$. Since $A(t) \to \infty$ as $t \to \infty$ for $\alpha > \alpha_0 \geq 7/3$, we conclude that $h(t) > 0$ for a large t and consequently there exists at least one positive root of $h(t) = 0$.

To show there is only one such root, it suffices to show that if $h(t_0) = 0$ for $0 < t_0 < \infty$, then $h'(t_0) > 0$. Let

$$0 = h(t_0) = q^2 A(t_0) + p^2 B(t_0) + pqC(t_0)$$

and consider

$$h'(t_0) = q^2 A'(t_0) + p^2 B'(t_0) + pqC'(t_0).$$

Since

$$\begin{aligned}
q^2 A'(t_0) &= q^2 \int\int \frac{t_0^{\alpha_1+\alpha_2-4}(\alpha_1+\alpha_2-4)(\alpha_1-\alpha_2+1)(\alpha_2-1)}{t_0\Gamma(\alpha_1)\Gamma(\alpha_2)} Q(\,d\alpha_1)Q(\,d\alpha_2) \\
&> q^2 \frac{2(\alpha_0-2)}{t_0} \int\int \frac{t_0^{\alpha_1+\alpha_2-4}(\alpha_1-\alpha_2+1)(\alpha_2-1)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} Q(\,d\alpha_1)Q(\,d\alpha_2) \\
&= q^2 A(t_0) = \frac{2(\alpha_0-2)}{t_0}\left[-p^2 B(t_0) - pqC(t_0)\right].
\end{aligned}$$

It follows after simplification that

$$\begin{aligned}
h'(t_0) &> p^2 \frac{2(\alpha_0-2)}{t_0} \int\int_{[\xi,\infty)} \lambda\xi(\lambda-\xi)^2 \exp\left[-(\lambda+\xi-2)t_0\right] P(\,d\lambda)P(\,d\xi) \\
&\quad + p^2 \int\int_{[\xi,\infty)} \lambda\xi(\lambda-\xi)^2(\lambda+\xi-2) \exp\left[-(\lambda+\xi-2)t_0\right] P(\,d\lambda)P(\,d\xi) \\
&\quad + pq \int\int \frac{\lambda t_0^{\alpha-4} \exp\left[-(\lambda-1)t_0\right] U(t_0)}{\Gamma(\alpha)} P(\,d\lambda)Q(\,d\alpha),
\end{aligned}$$

where

$$\begin{aligned}
U(t_0) &= (\lambda-1)^3 t_0^3 + (\lambda-1)^2(2\alpha_0+\alpha-5)t_0^2 \\
&\quad + (\lambda-1)(\alpha-1)(4\alpha/-0-\alpha-6)t_0 + (\alpha-1)(\alpha-2)(2\alpha_0-\alpha-1).
\end{aligned}$$

When $\alpha_0 \geq 7/3$, $U(t_0) > 0$. Then it is clear that $h'(t_0) > 0$.

We can thus conclude that $h(t) < 0$ for $0 < t < t_0$ and $h(t) > 0$ for $t > t_0$. This implies that $\eta(t)$ is bathtub-shaped. According to Theorem 2.1.1 the failure rate is therefor either increasing or nondegenerate bathtub. Suppose now that P has a finite first moment. Then from Remark 2.4 of Block, Li, and Savits [2003a], we deduce that

$$r'(0+) = -p \int \lambda^2 P(d\lambda) + p^2 \left[\int \lambda P(d\lambda) \right]^2 < 0.$$

Hence $r(t)$ must be nondegenerate bathtub.

Theorem 3.2.1 can be rewritten as the following.

Theorem 3.2.2. *Let*

$$\bar{K}(t|\alpha_0, \lambda_0) = \int g(t|\alpha, \lambda_0) Q(d\alpha).$$

Any survival function $\bar{F}(t)$ of the form

$$\bar{F}(t) = p\bar{H}(t) + q\bar{K}(t|\alpha_0, \lambda_0),$$

where $p + q = 1$, $p > 0$, $q > 0$ $\lambda_0 > 0$, $\alpha_0 > 7/3$ and $\bar{H}(t)$ is a completely monotone function whose associated probability measure P in the representation (3.5) has support a subset of (λ_0, ∞) and $\bar{K}(t|\alpha_0, \lambda_0)$ whose associated probability measure Q has support a subset of $(\alpha_0, \alpha_0 + 1)$, has a bathtub-shaped failure rate (may be degenerate). If, in addition $-\bar{H}'(0+) = \int \lambda P(d\lambda) < \infty$, then it is nondegenerate bathtub.

Like Theorem 3.1.2, this reformulation offers an easy access to apply the theorem to the following cases as in Section 3.1.2.

1) mixtures of DFR gammas with mixtures of IFR gammas are bathtub;
2) mixtures of DFR "Weibulls" with mixtures of IFR gammas are bathtub.
3) mixtures of certain bathtub distributions are bathtub.

All the applications are analogous to those in Section 3.1.2.

For a heterogeneous population, if the subpopulations are considered to be a class of DFR gamma or "Weibull" distributions and an IFR gamma distribution or a class of IFR gamma distributions, then the lifetime of the entire population may have a bathtub-shaped failure rate when the conditions described in the theorem are satisfied. Thus, based on the bathtub failure rate obtained, important decision can be made, such as the optimum burn-in time. (Burn-in is a procedure used for eliminating weak components in a mixed population and a widely used engineering method.)

4.0 MULTIPLE TESTING PROCEDURES

4.1 STATISTICAL HYPOTHESES TESTING

Statistical hypothesis testing is a method of using statistics to make a statistical decision about population parameters. The usual process of hypothesis testing consists of four steps.

1. Formulate the null hypothesis H_0 and the alternative hypothesis H_α.
2. Identify the test statistic X.
3. Compute the p-value.
4. Compare the p-value to an acceptable significance level, α. If p-value$< \alpha$, then the null hypothesis is rejected.

Two possible errors can be made when testing a hypothesis. A type I error occurs when a true null hypothesis is rejected (a false negative). A type II error occurs when a false null hypothesis is not rejected (a false positive). The probability of a type I error is called the significance level and is denoted by α; the probability of a type II error is denoted by β. Since type I errors are regarded as more serious than type II errors, α is usually taken to be smaller than β.

If we have our choice between several tests, we would like to choose a test that has small probability of both types of errors. However, reducing the probability of the type I error and reducing the probability of the type II error work against each other. We need to strike an appropriate balance between them. A standard method is to seek a test with small probability of type II error among a set of tests whose probabilities of Type I error are less than or equal to an acceptable level α_0 (these tests are called size-α_0 tests).

The power of a test is the probability that the test will reject a false null hypothesis (that it will not make a Type II error). Thus, the power of a test equals 1 minus the probability of a Type II error, i.e., $1 - \beta$. A test is said to be conservative if the probability of type I error is less than or equal to a presumed level. Hence, in other words, we seek a test which is conservative and has relative large power.

We introduce some standard notation first. Consider m null hypotheses $H_{01}, \ldots, H_{0m}$ and their alternative hypotheses $H_{\alpha 1}, \ldots, H_{\alpha m}$ with test statistics $X_1, \ldots, X_m$, p-values $p_1, \ldots, p_m$ and random p-values $P_1, \ldots, P_m$.

Suppose m_0 of the null hypotheses are true and $m_1 = m - m_0$ are false. Define the number of null hypotheses of the each category as in the following table:

Table 1

Number of	Number not rejected	Number rejected	Total number
True null hypotheses	U	V	m_0
Non-true null hypotheses	T	S	m_1
Total number	$m - R$	R	m

The total number of hypotheses m is assumed to be known, but m_0 and m_1 are not known. R is the number of the null hypotheses rejected and it is an observable random variable, whereas V, S, T and U are unobservable random variables. Unlike the single hypothesis testing situation, there are a variety of generalizations of the concept if a type I error rate. Here we list some of the most of standard ones (Shaffer [1995], Dudoit, Shaffer, and Boldrick [2003]).

- The per-comparison error rate (PCER) is defined as the expected value of the number of type I errors divided by the number of hypotheses, that is, PCER=$E(V)/m$.
- The per-family error rate (PFEW) is defined as the expected number of type I errors, that is, PFER= $E(V)$.
- The familywise error rate (FWER) is defined as the probability of at least one type I error, that is, $Pr(V \geq 1)$.
- The false discovery rate (FDR) of Benjamini and Hochberg [1995] is the expected proportion of type I errors among the rejected hypotheses, that is, FDR= $E(Q)$, where, by

definition, $Q = V/R$ if $R > 0$ and 0 if $R = 0$.

A multiple testing procedure is said to control a particular type I error rate at level α, if this error rate is less than or equal to α. For instance, the familywise error rate being controlled at level α implies that FWER$= Pr(V \geq 1) \leq \alpha$. There are two types of control (e.g., see Dudoit, Shaffer, and Boldrick [2003]): weak control and strong control. Weak control refers to control of the type I error rate only when all the null hypotheses are true, that is, $H_0 = \cap_{i=1}^{m} H_{0i}$ is true. Strong control refers to control of the type I error rate under any combination of true and false null hypotheses, that is, for any subset of true null hypotheses.

In this dissertation, we consider controlling FWER at level α in the weak sense, i.e.,

$$Pr(V \geq 1 | \text{all } H_{01}, \ldots, H_{0m} \quad \text{are true}) \leq \alpha \tag{4.1}$$

holds. When applying the Simes procedure, because $\{V \geq 1\} = \bigcup_{j=1}^{m} \left\{P_{(j)} \leq j\alpha/m\right\}$ and $\{\text{all } H_{01}, \ldots, H_{0m} \quad \text{are true}\}$ is equivalent to $\{H_0 \quad \text{is true}\}$, above inequality (4.1) is the same as the Simes inequality (1.2). The power can be defined as $Pr(V \geq 1 | H_0 \text{is not true})$ and the type II error rate is one minus the power.

4.2 BONFERRONI PROCEDURE

The classical Bonferroni procedure rejects $H_0 = \cap_{i=1}^{m} H_{0i}$ if at least one p-value is less than α/m, so that the FWER is controlled at level α. This is ensured by the Bonferroni inequality (1.1)

$$Pr(\bigcup_{i=1}^{m} \{P_i \leq \alpha/m\} | H_0 \text{is true}) \leq \alpha.$$

The Bonferroni procedure is widely used in many fields because it is easy to perform and requires no distributional assumptions. However, in spite of this (or perhaps because of it), the Bonferroni method has attracted some criticism. Its biggest problem is that it is too conservative: in controlling the family-wise error rate, each individual test is held to an unreasonably high standard, especially when m is large. This increases the probability of a

Type II error, reduces the power and makes it more likely that a false hypothesis H_i will fail to be detected.

Many efforts have been made to overcome this shortcoming. Holm [1979] proposed a sequentially rejective Bonferroni procedure which modifies the criterion in a stagewise manner. Shaffer [1986] modified this again by considering the family of logically related hypotheses. As a simple example of such logically related hypotheses, consider three hypotheses of pairwise equality: $\mu_i = \mu_{i'}$ for $i < i'$ where $i, i' \in \{1, 2, 3\}$ and μ_i is the mean of distribution i. If any one of these is false, at least one other must be false. Thus there cannot be one false and two true hypotheses among these three. If testing all hypotheses of pairwise equality with more than three distributions, there are more such constraints. Other methods were also proposed to make the test less conservative and have large power, such as in the papers of Hochberg [1988], Hommel [1988], and Benjamini and Hochberg [1995].

4.3 SIMES PROCEDURE

Simes [1986] proposed a modified Bonferroni procedure: reject the overall $H_0 = \cap_{i=1}^{m} H_{0i}$ if $p_{(j)} \leq j\alpha/m$ for at least one $j = 1, \ldots, m$, where $p_{(j)}$ is the jth smallest observed p-value. The power of Simes procedure is greater than the classical Bonferroni procedure, because

$$Pr(\bigcup_{i=1}^{m} \{P_i \leq \alpha/m\}) = Pr(\{P_{(1)} \leq \alpha/m\}) \leq Pr(\bigcup_{j=1}^{m} \{P_{(j)} \leq j\alpha/m\}).$$

Simes proved that for independent test statistics, the FWER of the modified procedure equals α:

$$Pr(\bigcup_{j=1}^{m} \{P_{(j)} \leq j\alpha/m\} \,|H_0 \text{ is true}) = \alpha.$$

Theorem 4.3.1 (Simes [1986])**.** *Let $P_{(1)}, \ldots, P_{(m)}$ be the order statistics of m independent uniform $(0, 1)$ random variables and let $A_m(\alpha) = Pr(P_{(j)} > j\alpha/m; j = 1, \ldots, m)$ $(0 \leq \alpha \leq 1)$. Then $A_m(\alpha) = 1 - \alpha$.*

Sarkar [1998] showed that using the Simes procedure to test a set of one-tail null hypotheses with the test statistics being multivariate totally positive of order 2 (see Section 4.4), the FWER is controlled at level α in the weak sense, that is, the Simes inequality (1.2) holds

$$Pr(\bigcup_{j=1}^{m}\{P_{(j)} \leq j\alpha/m\}\,|H_0 \text{ is true}) \leq \alpha.$$

Theorem 4.3.2 (Sarkar [1998] Theorem 3.1)**.** *Let $X_{(1)} \leq \ldots \leq X_{(m)}$ be the ordered components of an MTP_2 random vector $\mathbf{X} = (X_1, \ldots, X_m)$ and let F_i be the marginal cdf of X_i. Then we have the following:*

1. *For fixed $a_1 \leq \ldots \leq a_m$,*

$$Pr(X_{(1)} \geq a_1, \ldots, X_{(m)} \geq a_m) \geq 1 - \frac{1}{m}\sum_{i=1}^{m} F_i(a_m^-), \tag{4.2}$$

   if $\frac{1}{j}F_i(a_j^-)$ is nondecreasing in $j = 1, \ldots, m$ for all $i = 1, \ldots, m$;

2. *For fixed $b_1 \leq \ldots \leq b_m$,*

$$Pr(X_{(1)} \leq b_1, \ldots, X_{(m)} \leq b_m) \geq \frac{1}{m}\sum_{i=1}^{m} F_i(b_1), \tag{4.3}$$

   if $\frac{1}{j}\bar{F}_i(b_{m-j+1})$ is nondecreasing in $j = 1, \ldots, m$ for all $i = 1, \ldots, m$, where $\bar{F}(x) = 1 - F(x)$.

Proposition 4.3.3 (Sarkar [1998] Proposition 3.1)**.** *The Simes inequality (1.2) holds for MTP_2 random variables with common marginals F.*

Theorem 4.3.2 and Proposition 4.3.3 work well for continuous MTP_2 distributions. However there are some difficulties when applying them for discrete distributions. We will discuss that in Section 5.2.

4.4 DEPENDENT STRUCTURES

In this section, we provide a few of dependence structures.

Definition 1 (Block, Savits, and Shaked [1985]). *A random vector* $\mathbf{X} = (X_1, \ldots, X_m)$ *is said to be positively dependent through stochastic ordering (PDS) if for any* $i = 1, \ldots, m$, *the conditional expectation* $E\left[g(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_m) | X_i = x\right]$ *is nondecreasing in* x *whenever* g *is a nondecreasing Borel measurable function such that the above conditional expectation exists. It is written as*

$$(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_m) \uparrow^{st} X_i = x \tag{4.4}$$

when the vector is PDS.

Definition 2 (Karlin and Rinott [1980a]). *A random vector* $\mathbf{X} = (X_1, \ldots, X_m)$ *is* MTP_2 *if its density* f *satisfies*

$$f(\mathbf{x} \vee \mathbf{y}) f(\mathbf{x} \wedge \mathbf{y}) \geq f(\mathbf{x}) f(\mathbf{y})$$

where

$$\mathbf{x} \vee \mathbf{y} = (\max(x_1, y_1), \ldots, \max(x_m, y_m))$$
$$\mathbf{x} \wedge \mathbf{y} = (\min(x_1, y_1), \ldots, \min(x_m, y_m))$$

When $m = 2$, *the vector is called totally positive of order 2 (*TP_2*).*

Remark. MTP_2 is a stronger condition of positive dependence; it implies PDS.

Definition 3 (Block, Savits, and Shaked [1985]). *A random vector* $\mathbf{X} = (X_1, \ldots, X_m)$ *is said to be negatively dependent through stochastic ordering (NDS) if for any* $i = 1, \ldots, m$,

$$(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_m) \downarrow^{st} X_i = x, \tag{4.5}$$

i.e., the conditional expectation $E\left[g(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_m) | X_i = x\right]$ *is nonincreasing in* x *whenever* g *is a nondecreasing Borel measurable function such that the above conditional expectation exists.*

Next, we introduce a structured type of negative dependence, which implies NDS.

Definition 4 (Block, Savits, and Shaked [1982]). *The random vector* $(X_1, \ldots, X_n)$ *satisfies condition N if there exist* $n+1$ *independent random variables* $S_0, S_1, \ldots, S_n$ *each having a* PF_2 *density (or probability function) and a real number* s *such that*

$$(X_1, \ldots, X_n) \sim ((S_1, \ldots, S_n) | S_0 + S_1 + \ldots + S_n = s)$$

where the "$\sim$*" means the two quantities have the same joint distribution.*

Remarks.

- The function ϕ defined on $(-\infty, \infty)$ is said to be PF_2 if $\phi(\xi - \eta)$ is TP_2 in the variables $-\infty < \xi, \eta < \infty$ (Karlin and Rinott [1980b]).
- Condition N implies NDS.

5.0 APPLICATION OF DEPENDENCE IN SIMES INEQUALITY

5.1 SIMES INEQUALITY FOR POSITIVE DEPENDENCE

We show that when replacing MTP_2 by PDS, the results of Theorem 4.3.2 still hold.

Lemma 5.1.1 (Sarkar [1998] Lemma 2.1)**.** *Let $X_{(1)} \leq \ldots \leq X_{(m)}$ be the ordered components of $\mathbf{X} = (X_1, \ldots, X_m)$ and let F_i be the marginal cdf of X_i. Then,*

$$\begin{aligned} & Pr(X_{(1)} \geq a_1, \ldots, X_{(m)} \geq a_m) \\ = \; & 1 - \frac{1}{m} \sum_{i=1}^{m} F_i(a_n^-) \\ & + \sum_{i=1}^{m} \sum_{j=1}^{m-1} E_{X_i} \left[\left\{ \frac{I(X_i < a_{j+1})}{j+1} - \frac{I(X_i < a_j)}{j} \right\} Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i) \right], \end{aligned}$$

where $X_{(j)}^{(-i)}$ is the jth smallest component of $\mathbf{X}^{(-i)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_m)$ for $j = 1, \ldots, m-1$ and $i = 1, \ldots, m$.

Theorem 5.1.2. *Let $X_{(1)} \leq \ldots \leq X_{(m)}$ be the ordered components of a PDS random vector $\mathbf{X} = (X_1, \ldots, X_m)$, and let F_i be the marginal cdf of X_i. Then we have the following.*

1. *For fixed $a_1 \leq \ldots \leq a_m$,*

$$Pr(X_{(1)} \geq a_1, \ldots, X_{(m)} \geq a_m) \geq 1 - \frac{1}{m} \sum_{i=1}^{m} F_i(a_m^-), \tag{5.1}$$

if $\frac{1}{j} F_i(a_j^-)$ is nondecreasing in $j = 1, \ldots, m$ for all $i = 1, \ldots, m$;

2. For fixed $b_1 \leq \ldots \leq b_m$,

$$Pr(X_{(1)} \leq b_1, \ldots, X_{(m)} \leq b_m) \geq \frac{1}{m} \sum_{i=1}^{m} F_i(b_1), \tag{5.2}$$

if $\frac{1}{j}\bar{F}_i(b_{m-j+1})$ is nondecreasing in $j = 1, \ldots, m$ for all $j = 1, \ldots, m$, where $\bar{F}(x) = 1 - F(x)$.

Proof. By Lemma 5.1.1, if suffices to show that

$$\sum_{i=1}^{m} \sum_{j=1}^{m-1} E_{X_i} \left[\left\{ \frac{I(X_i < a_{j+1})}{j+1} - \frac{I(X_i < a_j)}{j} \right\} Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i) \right] \geq 0$$

for the first part of the theorem.

Define

$$h_j(X_i) = \frac{I(X_i < a_{j+1})}{j+1} - \frac{I(X_i < a_j)}{j},$$

and

$$g_j(X_i) = Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i).$$

Notice that

$$h_j(X_i) = \begin{cases} -\frac{1}{j(j+1)} & \text{if} \quad X_i < a_j, \\ \frac{1}{j+1} & \text{if} \quad a_j \leq X_i < a_{j+1}, \\ 0 & \text{otherwise} \end{cases}$$

and

$$g_j(X_i) = E\left[I(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m) | X_i \right]$$

where the expectation is calculated with respect to $X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_m$.

Since $X_1, \ldots, X_m$ is PDS and $I(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m)$ is a nondecreasing Borel measurable function, $g_j(X_i)$ is nondecreasing in X_i. Thus,

$$\begin{aligned}
& E_{X_i}(h_j(X_i)g_j(X_i)) \\
= \; & E_{X_i}(I(X_i < a_j)h_j(X_i)g_j(X_i)) + E_{X_i}(I(a_j \leq X_i < a_{j+1})h_j(X_i)g_j(X_i)) \\
& + E_{X_i}(I(X_i \geq a_{j+1})h_j(X_i)g_j(X_i)) \\
\geq \; & E_{X_i}(I(X_i < a_j)h_j(X_i))g_j(a_j) + E_{X_i}(I(a_j \leq X_i < a_{j+1})h_j(X_i))g_j(a_j) + 0 \\
= \; & E_{X_i}(h_j(X_i))g_j(a_j) \\
= \; & E_{X_i}\left\{\frac{I(X_i < a_{j+1})}{j+1} - \frac{I(X_i < a_j)}{j}\right\} Pr(X_j^{(-i)} \geq a_{j+1}, \ldots, X_{m-1}^{(-i)} \geq a_m | X_i = a_j) \\
= \; & \left\{\frac{F_i(a_{j+1}^-)}{j+1} - \frac{F_i(a_j^-)}{j}\right\} Pr(X_j^{(-i)} \geq a_{j+1}, \ldots, X_{m-1}^{(-i)} \geq a_m | X_i = a_j)
\end{aligned}$$

By the monotonicity assumption on the marginals that $\frac{1}{j}F_i(a_j^-)$ is nondecreasing in $j = 1, \ldots, m$ for all $i = 1, \ldots, m$,

$$\frac{F_i(a_{j+1}^-)}{j+1} - \frac{F_i(a_j^-)}{j} \geq 0.$$

We conclude that $E_{X_i}(h_j(X_i)g_j(X_i)) \geq 0$ and hence

$$\sum_{i=1}^{m}\sum_{j=1}^{m-1} E_{X_i}\left[\left\{\frac{I(X_i < a_{j+1})}{j+1} - \frac{I(X_i < a_j)}{j}\right\} Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i)\right] \geq 0.$$

The second part then can be proved by applying part 1 to $-\mathbf{X} = (-X_1, \ldots, -X_m)$ which is also PDS.

Proposition 5.1.3. *Consider m null hypotheses $H_{01}, \ldots, H_{0m}$ and the corresponding alternative hypotheses $H_{\alpha 1}, \ldots, H_{\alpha m}$. If*

(a) $H_{\alpha 1}, \ldots, H_{\alpha m}$ are all left-tailed or all right-tailed,

(b) the marginal distribution of the test statistics are identical with cdf F and F is continuous,

(c) the test statistics $X_1, \ldots, X_m$ are PDS random variables,

then the Simes inequality (1.2) holds.

Proof. When $H_{\alpha 1}, \ldots, H_{\alpha m}$ are all left-tailed, let a_j be the value so that $F(a_j) = \frac{j\alpha}{m}$. Therefore, for any $j = 1, \ldots, m$,

$$\frac{F(a_j)}{j} = \frac{F(a_{j+1})}{j+1}$$

By (5.1)

$$\begin{aligned} & Pr(\bigcup_{j=1}^{m} \{P_{(j)} \leq j\alpha/m\}) \\ = \;& Pr(\bigcup_{j=1}^{m} \{X_{(j)} \leq a_j\}) \\ = \;& 1 - Pr(\bigcap_{j=1}^{m} \{X_{(j)} \geq a_j\}) \\ \leq \;& 1 - (1 - F(a_m)) = \alpha. \end{aligned}$$

An analogous proof holds for the right-tailed case.

Remark. Some problems arise when F is a discrete distribution, we only illustrate in the left-tailed case:

$$\begin{aligned} & Pr(\bigcup_{j=1}^{m} \{P_{(j)} \leq j\alpha/m\}) \\ = \;& Pr(\bigcup_{j=1}^{m} \{X_{(j)} < \inf\{x : F(x) > j\alpha/m\}\}) \\ = \;& 1 - Pr(\bigcap_{j=1}^{m} \{X_{(j)} \geq \inf\{x : F(x) > j\alpha/m\}\}). \end{aligned}$$

Let $a_j = \inf\{x : F(x) > j\alpha/m\}$. It is clear that $a_1 \leq a_2 \leq \ldots \leq a_m$ and $F(a_m^-) \leq j\alpha/m$. In order to apply Theorem 5.1.2, the marginal cdf F needs to satisfy that $\frac{1}{j}F(a_j^-)$ is nondecreasing in j. Then

$$Pr(\bigcup_{j=1}^{m} \{P_{(j)} \leq j\alpha/m\}) \leq 1 - F(a_m^-) \leq \alpha.$$

However, this monotonicity condition is not always satisfied as in the continuous case. As a counter example, consider a discrete distribution with the cdf

$$F(t) = \begin{cases} 0 & \text{for } 0 \leq t < t_1, \\ \alpha/3 & \text{for } t_1 \leq t < t_2, \\ 1 & \text{for } t \geq t_2. \end{cases}$$

where $0 < \alpha < 1$. Suppose $m = 2$, then $a_1 = a_2 = t_2$ and

$$F(a_1^-) = F(t_2^-) \geq \frac{1}{2}F(t_2^-) = \frac{1}{2}F(a_2^-).$$

A sample application of Proposition 5.1.3 to a multivariate normal distribution follows.

Example 1. (Multivariate normal test statistics) Consider a vector of test statistics $\mathbf{Y} = (Y_1, \ldots, Y_m)$ which has a multivariate normal distribution with means 0 and variances 1. If the correlations $\rho_{ij} \geq 0$ for all $i \neq j$, then $\mathbf{Y}$ is PDS and the Simes inequality follows.

Another standard distribution often used is the multivariate t distribution. Unfortunately, unlike the multivariate normal distribution, the PDS property does not follow by examining the correlations. Consequently, the Simes inequality does not readily follow from the above Theorem. Sarkar [1998] recognized this and developed a corollary (see Sarkar [1998] Corollary 3.1). However, this result appears not to be correct. We give an alternative demonstration for the multivariate-t distribution in the following theorem.

Theorem 5.1.4. *Let $\mathbf{Y} = (Y_1, \ldots, Y_m)$ be the random vector described in Example 2. Let νS^2 have a χ^2 distribution with degrees of freedom ν, and assume that $\mathbf{Y}$ and S are independent. Then $\mathbf{X} = (X_1, \ldots, X_m) = (\frac{Y_1}{S}, \ldots, \frac{Y_m}{S})$ is the standard multivariate-t distribution. Let F_i denote the marginal distribution of X_i for $i = 1, \ldots, m$. Then*

1. *For fixed $a_1 \leq \ldots \leq a_m < 0$,*

$$Pr(X_{(1)} \geq a_1, \ldots, X_{(m)} \geq a_m) \geq 1 - \frac{1}{m}\sum_{i=1}^{m} F_i(a_m), \tag{5.3}$$

if $\frac{1}{j}F_i(a_j)$ is nondecreasing in $j = 1, \ldots, m$ for all $j = 1, \ldots, m$;

2. *For fixed $0 < b_1 \leq \ldots \leq b_m$,*

$$Pr(X_{(1)} \leq b_1, \ldots, X_{(m)} \leq b_m) \geq \frac{1}{m}\sum_{i=1}^{m} F_i(b_1), \tag{5.4}$$

if $\frac{1}{j}\bar{F}_i(b_{m-j+1})$ is nondecreasing in $j = 1, \ldots, m$ for all $j = 1, \ldots, m$, where $\bar{F}(x) = 1 - F(x)$.

The following lemma is needed in the proof of the above theorem.

Lemma 5.1.5. *If $X = \frac{Y}{S}$ where $Y \sim N(0,1)$ and $nS^2 \sim \chi_n^2$ are independent, it is well known that X has a t-distribution with degrees of freedom n. Then for $x < x' < 0$ and $y < 0$*

$$Pr(Y \leq y|X = x) > Pr(Y \leq y|X = x') \tag{5.5}$$

$$Pr(S \leq s|X = x) > Pr(S \leq s|X = x') \tag{5.6}$$

where $s > 0$;

Proof. The density functions of Y and S are

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\}, \text{for } -\infty < y < \infty.$$
$$f_S(s) = \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})} s^{n-1} \exp\left\{-\frac{ns^2}{2}\right\}, \text{for } s > 0.$$

Using the transformation $X = \frac{Y}{S}$ and $Y = Y$, we derive the joint density of X and Y as

$$f_{X,Y}(x,y) = \begin{cases} \frac{n^{\frac{n}{2}}}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})\sqrt{2\pi}} \frac{(y^2)^{\frac{n}{2}}}{(x^2)^{\frac{(n+1)}{2}}} \exp\left\{-\frac{ny^2}{2}(1+\frac{n}{x^2})\right\} & \text{if } x<0, y<0 \text{ or } x>0, y>0 \\ 0 & \text{otherwise.} \end{cases}$$

Similarly the joint density of X and S is

$$f_{X,S}(x,s) = \begin{cases} \frac{n^{\frac{n}{2}} s^n}{2^{\frac{n}{2}-1}\Gamma(\frac{n}{2})\sqrt{2\pi}} \exp\left\{-\frac{ns^2}{2}(1+\frac{x^2}{n})\right\} & \text{if } s>0 \\ 0 & \text{otherwise.} \end{cases}$$

Dividing these joint density functions by the marginal density of X,

$$f_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}, \text{for } -\infty < x < \infty,$$

The conditional density function of Y given $X = x$ is

$$\begin{aligned} f_{Y|X=x}(y) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\ &= \begin{cases} \frac{(y^2(1+\frac{n}{x^2}))^{\frac{n}{2}}}{2^{\frac{n-1}{2}}\Gamma(\frac{n+1}{2})}(1+\frac{n}{x^2})^{\frac{1}{2}} \exp\left\{-\frac{y^2(1+\frac{n}{x^2})}{2}\right\} & \text{for } x<0, y<0 \text{ or } x>0, y>0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

and the conditional density function of S given $X = x$ as

$$f_{S|X=x}(s) = \frac{f_{X,S}(x,s)}{f_X(x)} = \begin{cases} \frac{(s^2(n+x^2))^{\frac{n}{2}}}{2^{\frac{n-1}{2}}\Gamma(\frac{n+1}{2})}(n+x^2)^{\frac{1}{2}} \exp\left\{-\frac{s^2(n+x^2)}{2}\right\} & \text{for } s > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore if $x < 0$ and $y < 0$,

$$\begin{aligned} & Pr(Y \leq y|X = x) \\ = & \int_{-\infty}^{y} \frac{(u^2(1+\frac{n}{x^2}))^{\frac{n}{2}}}{2^{\frac{n-1}{2}}\Gamma(\frac{n+1}{2})}(1+\frac{n}{x^2})^{\frac{1}{2}} \exp\left\{-\frac{u^2(1+\frac{n}{x^2})}{2}\right\} du \\ = & \int_{-\infty}^{(1+\frac{n}{x^2})^{\frac{1}{2}}y} \frac{(w^2)^{\frac{n}{2}}}{2^{\frac{n-1}{2}}\Gamma(\frac{n+1}{2})} \exp\left\{-\frac{w^2}{2}\right\} dw, \end{aligned}$$

which is strictly decreasing in x; and

$$\begin{aligned} & Pr(S \leq s|X = x) \\ = & \int_{0}^{s} \frac{(u^2(n+x^2))^{\frac{n}{2}}}{2^{\frac{n-1}{2}}\Gamma(\frac{n+1}{2})}(n+x^2)^{\frac{1}{2}} \exp\left\{-\frac{u^2(n+x^2)}{2}\right\} du \\ = & \int_{0}^{(n+x^2)^{\frac{1}{2}}s} \frac{(w^2)^{\frac{n}{2}}}{2^{\frac{n-1}{2}}\Gamma(\frac{n+1}{2})} \exp\left\{-\frac{w^2}{2}\right\} dw \end{aligned}$$

strictly decreases in x. These prove (5.5) and (5.6).

Proof of Theorem 5.1.4. Refer to the proof of Theorem 5.1.2, for the first part, it suffices to prove for $j = 1, \ldots, m-1$, that

$$Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i = x)$$

is nondecreasing in x when $x < 0$.

The proof is very similar that of Lemma 3.1 in Benjamini and Yekutieli [2001]. We prove the theorem in three steps.

Step 1: For each pair of $x < x' < 0$ we construct a new random variable S' whose marginal distribution is stochastically smaller than the marginal distribution of S, but whose conditional distribution given $X_i = x'$ is identical to the conditional distribution of S given $X_i = x$.

By Lemma 5.1.5, we know that the cdf of $S|X_i = x'$ is less than cdf of $S|X_i = x$, i.e.,

$$F_{S|X_i=x'}(s) < F_{S|X_i=x}(s). \tag{5.7}$$

Since S and X_i are continuous random variables, the cdf's of $S|X_i = x$ and $S|X_i = x'$ are continuous and strictly increasing. The inverses of these cdf's exist and are well defined. They are also continuous and strictly increasing. Define

$$h_{x,x'}(\tau) = F^{-1}_{S|X_i=x}(F_{S|X_i=x'}(\tau))$$

and

$$h_{x',x}(\tau) = F^{-1}_{S|X_i=x'}(F_{S|X_i=x}(\tau)).$$

Then by (5.7),

$$h_{x,x'}(s) = F^{-1}_{S|X_i=x}(F_{S|X_i=x'}(s)) < F^{-1}_{S|X_i=x}(F_{S|X_i=x}(s)) = s.$$

Define the new random variable S' as

$$S' = h_{x,x'}(S)$$

and let

$$s' = h_{x',x}(s).$$

Then

(a) $s = h_{x,x'}(h_{x',x}(s)) = h_{x,x'}(s') < s'$.

(b) $F_{S|X=x}(s) = F_{S|X=x}(h_{x,x'}(s')) = F_{S|X=x'}(s')$.

(c) $\{S \leq s'\} = \{S \leq h_{x',x}(s)\} = \{S' \leq h_{x,x'}(h_{x',x}(s))\} = \{S' \leq s\}$.

Hence

$$Pr(S \le s|X_i = x) = Pr(S \le s'|X_i = x') = Pr(S' \le s|X = x'),$$

that is $S|X_i = x$ and $S'|X_i = x'$ are identically distributed.

Step 2: We show that the newly defined random variable S' satisfies

$$\begin{aligned} & Pr(X_{(j)}^{(-i)} \ge a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \ge a_m | X_i = x, S = s) \\ \le \; & Pr(X_{(j)}^{(-i)} \ge a_{k+1}, \ldots, X_{(m-1)}^{(-i)} \ge a_m | X_i = x', S' = s). \end{aligned} \tag{5.8}$$

First, for $x < x' < 0$, $s > 0$ and $s' = h_{x',x}(s)$, we show that $sx < s'x' < 0$. When fixing $X_i = x$,

$$Pr(Y_i \le sx|X_i = x) = Pr(\frac{Y_i}{x} \le s|X_i = x) = Pr(S \ge s|X_i = x) = 1 - F_{S|X_i=x}(s).$$

When fixing $X_i = x'$, similarly,

$$Pr(Y_i \le s'x'|X_i = x') = Pr(\frac{Y_i}{x'} \le s'|X_i = x') = Pr(S \ge s'|X_i = x') = 1 - F_{S|X_i=x'}(s').$$

By the fact (b) in step 1 that $F_{S|X=x}(s) = F_{S|X=x'}(s')$,

$$Pr(Y_i \le sx|X_i = x) = Pr(Y_i \le s'x'|X_i = x'). \tag{5.9}$$

due to the inequality (5.5) in Lemma 5.1.5, $sx < s'x' < 0$ must to be true.

Next, we prove the inequality (5.8). Since $(Y_1, \ldots, Y_m)$ are PDS and $sx < s'x' < 0$,

$$\begin{aligned} & Pr(X_{(j)}^{(-i)} \ge a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \ge a_m | X_i = x, S = s) \\ = \; & Pr(\frac{Y_{(j)}^{(-i)}}{S} \ge a_{j+1}, \ldots, \frac{Y_{(m-1)}^{(-i)}}{S} \ge a_m | \frac{Y_i = x}{S}, S = s) \\ = \; & Pr(Y_{(j)}^{(-i)} \ge sa_{k+1}, \ldots, Y_{(m-1)}^{(-i)} \ge sa_m | Y_i = sx, S = s) \\ \le \; & Pr(Y_{(j)}^{(-i)} \ge sa_{j+1}, \ldots, Y_{(m-1)}^{(-i)} \ge sa_m | Y_i = s'x', S = s). \end{aligned}$$

Then because $(Y_1, \ldots, Y_m)$ and S are independent and $0 < s < s'$,

$$
\begin{aligned}
& Pr(Y_{(j)}^{(-i)} \geq sa_{j+1}, \ldots, Y_{(m-1)}^{(-i)} \geq sa_m | Y_i = s'x', S = s) \\
= \; & Pr(Y_{(j)}^{(-i)} \geq sa_{j+1}, \ldots, Y_{(m-1)}^{(-i)} \geq sa_m | Y_i = s'x', S = s') \\
\leq \; & Pr(Y_{(j)}^{(-i)} \geq s'a_{j+1}, \ldots, Y_{(m-1)}^{(-i)} \geq s'a_m | Y_i = s'x', S = s') \\
= \; & Pr(\frac{Y_{(j)}^{(-i)}}{S} \geq a_{j+1}, \ldots, \frac{Y_{(m-1)}^{(-i)}}{S} \geq a_m | \frac{Y_i = x}{S}, S = s') \\
= \; & Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i = x', S' = s) \\
= \; & Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i = x', S' = s)
\end{aligned}
$$

Step 3: The above result leads to our goal in that

$$
\begin{aligned}
& Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i = x) \\
= \; & E_{S|X=x} \left[Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i = x, S) \right] \\
\leq \; & E_{S'|X=x'} \left[Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i = x', S') \right] \\
= \; & Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i = x').
\end{aligned}
$$

The second part of the theorem follows by applying the first part to $-\mathbf{X}$.

5.2 SIMES INEQUALITY REVERSES FOR NEGATIVE DEPENDENCE

The Simes procedure cannot be used for Condition N or NDS random vectors since the Simes inequality (1.2) reverses in this case.

Theorem 5.2.1. *Let $X_{(1)} \leq \ldots \leq X_{(m)}$ be the ordered components of a random vector $\mathbf{X} = (X_1, \ldots, X_m)$ whose distribution satisfies condition N.*

1. *For fixed $a_1 \leq \ldots \leq a_m$,*

$$Pr(X_{(1)} \geq a_1, \ldots, X_{(m)} \geq a_m) \leq 1 - \frac{1}{m} \sum_{i=1}^{m} F_i(a_m^-), \tag{5.10}$$

if $\frac{1}{j} F_i(a_j^-)$ is nonincreasing in $j = 1, \ldots, m$ for all $i = 1, \ldots, m$;

2. For fixed $b_1 \leq \ldots \leq b_m$*,*

$$Pr(X_{(1)} \leq b_1, \ldots, X_{(m)} \leq b_m) \leq \frac{1}{m} \sum_{i=1}^{m} F_i(b_1), \tag{5.11}$$

if $\frac{1}{j}\bar{F}_i(b_{m-j+1})$ *is nonincreasing in* $j = 1, \ldots, m$ *for all* $i = 1, \ldots, m$*, where* $\bar{F}(x) = 1 - F(x)$*.*

Proof. As in the proof of Sarkar's Theorem 3.1 in Sarkar [1998], through the use of his Lemma 2.1

$$\begin{aligned} & Pr(X_{(1)} \geq a_1, \ldots, X_{(m)} \geq a_m) - \left\{1 - \frac{1}{n}\sum_{i=1}^{m} F_i(a_n^-)\right\} \\ = & \sum_{i=1}^{m}\sum_{j=1}^{m-1} E\left[\left\{\frac{I(X_i < a_{j+1})}{j+1} - \frac{I(X_i < a_j)}{j}\right\} Pr(X_{(j)}^{(-i)} \geq a_{j+1}, \ldots, X_{(m-1)}^{(-i)} \geq a_m | X_i)\right], \end{aligned}$$

where $X_{(1)}^{(-i)} \leq \ldots \leq X_{(m-1)}^{(-i)}$ are the ordered test statistics without X_i.

Now for a fixed i and j let $h(X_i)$ be the difference of the indicator functions and $g(X_i)$ be the conditional probability. By the assumption of condition N it follows from Block, Savits, and Shaked [1985] that the distribution is NDS. Then as in Sarkar [1998] the conditional probability of the set $g(X_i)$ is nonincreasing in X_i. We also notice that

$$h(X_i) = \begin{cases} -\frac{1}{j(j+1)} & \text{if} \quad X_i < a_j, \\ \frac{1}{j+1} & \text{if} \quad a_j \leq X_i < a_{j+1}, \\ 0 & \text{otherwise} \end{cases}$$

Therefore,

$$\begin{aligned} & E(h(X_i)g(X_i)) \\ = & E(I(X_i < a_j)h(X_i)g(X_i)) + E(I(a_j \leq X_i < a_{j+1})h(X_i)g(X_i)) + E(I(a_{j+1} \leq X_i)h(X_i)g(X_i)) \\ \leq & E(I(X_i < a_j)h(X_i))g(a_j) + E(I(a_j \leq X_i < a_{j+1})h(X_i))g(a_j) + 0 \\ = & E(f(X_i))g(a_j) \\ = & E\left\{\frac{I(X_i < a_{j+1})}{j+1} - \frac{I(X_i < a_j)}{j}\right\} Pr(X_j^{(-i)} \geq a_{j+1}, \ldots, X_{m-1}^{(-i)} \geq a_m | X_i = a_j) \\ = & \left\{\frac{F_i(a_{j+1}^-)}{j+1} - \frac{F_i(a_j^-)}{j}\right\} Pr(X_j^{(-i)} \geq a_{j+1}, \ldots, X_{m-1}^{(-i)} \geq a_m | X_i = a_j) \end{aligned}$$

By the monotonicity assumption,

$$\frac{F_i(a_{j+1}^-)}{j+1} - \frac{F_i(a_j^-)}{j} \leq 0.$$

Each (i, j)th term is thus bounded above by a nonpositive value. Hence we have proven (5.10).

The second part then can be proved by applying part 1 to $-\mathbf{X} = (-X_1, \ldots, -X_m)$.

Proposition 5.2.2. *Consider m null hypotheses $H_{01}, \ldots, H_{0m}$ and the corresponding alternative hypotheses $H_{\alpha 1}, \ldots, H_{\alpha m}$. If*

(a) $H_{\alpha 1}, \ldots, H_{\alpha m}$ are all left-tailed or all right-tailed,

(b) the marginal distribution of the test statistics are identical with cdf F and F is continuous,

(c) the test statistics $X_1, \ldots, X_m$ are condition N random variables.

Then the Simes inequality (1.2) reverses.

Proof. When $H_{\alpha 1}, \ldots, H_{\alpha m}$ are all left-tailed, let a_j be the value so that $F(a_j) = \frac{j\alpha}{m}$. Therefore, for any $j = 1, \ldots, m$,

$$\frac{F(a_j)}{j} = \frac{F(a_{j+1})}{j+1}$$

By (5.10)

$$\begin{aligned}
& Pr(\bigcup_{j=1}^{m} \{P_{(j)} \leq j\alpha/m\}) \\
= \; & Pr(\bigcup_{j=1}^{m} \{X_{(j)} \leq a_j\}) \\
= \; & 1 - Pr(\bigcap_{j=1}^{m} \{X_{(j)} \geq a_j\}) \\
\geq \; & 1 - (1 - F(a_m)) = \alpha.
\end{aligned}$$

An analogous proof holds for the right-tailed case.

Remark. It is still a mystery in the discrete case. Even when $\frac{1}{j}F(inf\,\{x : F(x) > j\alpha/m\}^{-})$ is nonincreasing in $j = 1, \ldots, m$ for all $i = 1, \ldots, m$ in the case of left-tailed tests,

$$\begin{aligned} & Pr(\bigcup_{j=1}^{m}\{P_{(j)} \leq j\alpha/m\}) \\ = & \; Pr(\bigcup_{j=1}^{m}\{X_{(j)} < inf\,\{x : F(x) > j\alpha/m\}\}) \\ = & \; 1 - Pr(\bigcap_{j=1}^{m}\{X_{(j)} \geq inf\,\{x : F(x) > j\alpha/m\}\}) \\ \geq & \; F(inf\,\{x : F(x) > \alpha\}^{-}). \end{aligned}$$

However, the lower bound of $F(inf\,\{x : F(x) > \alpha\}^{-})$ is unknown, we don't know if the inequality reverses or not.

In Block, Savits, and Shaked [1982] it is shown that many standard multivariate distributions thought to be negatively dependent satisfy the condition N. Some of these distributions are: 1) the equicorrelated multivariate normal with nonpositive correlation; 2) the Dirichlet; and 3) the Dirichlet compound multinomial. The verifications that these distributions satisfy condition N are, in general, well known properties that the distributions have structures as specified in the definition. For example, in Bayesian analysis the Dirichlet distribution can be expressed as the conditional distribution of independent gammas given that the sum is fixed.

The multivariate normal with arbitrary nonpositive correlation does not satisfy condition N; however it does satisfy the NDS condition given in Block, Savits, and Shaked [1985] as observed in Example 4.1 of that paper. Thus it satisfies the above theorem. We give the example below for emphasis.

Example 2. Let $\mathbf{X} = (X_1, \ldots, X_m)$ have a multivariate normal distribution with means 0, variances 1 and correlations $\rho_{ij} \leq 0$, and let $0 < \alpha < 1$. Let Φ be the marginal cumulative distribution function and z_b satisfy $\Phi(z_a) = 1 - a$. Since the distribution is NDS, Theorem 5.2.1 applies. Proposition 5.2.2 thus gives the reverse Simes inequality.

6.0 FUTURE WORK

6.1 THE STUDY OF FAILURE RATES OF MIXTURES

We examined the mixture (3.4) for the case of a mixture of exponentials and an IFR gamma distribution with shape parameter $\alpha \geq 2$. A question arises immediately: what happens when the mixture (3.4) consists of a mixture of exponentials and an IFR gamma with $1 < \alpha < 2$? Furthermore, what happens when the mixture consists of a mixture of exponentials and a mixture of IFR gammas with $1 < \alpha < 2$?

These questions remain unresolved except in special cases. We will attempt to obtain a general result by examining those special cases and generalizing the current results. Meanwhile, we will continue searching for more relaxed conditions for mixtures to have a bathtub-shaped failure rate.

6.2 CONTROLLING FDR UNDER DEPENDENCY

In Chapter 4 and Chapter 5, we discussed the Simes procedure controlling the FWER in the weak sense. The assumption is that all the null hypotheses are true. However, in general, it is not realistic to assume that all the null hypotheses are actually true. More often, only a subset of the null hypotheses are true, and the others are not. In such a case, we would like to control the type I error rate in the strong sense. In our future work, we will try to extend our results in this direction. In particular, we will study the the BH procedure (Benjamini and Hochberg [1995]).

The BH procedure rejects only a subset of hypotheses, $H_{(1)}, \dots, H_{(k)}$, where k is the largest j for which $P_{(j)} \leq \frac{j}{m}\alpha$. This procedure intends to control the false discovery rate (see Chapter 4) and assumes that only a subset of null hypotheses are true. Since FDR$\leq$FWER (when all the null hypotheses are true, FER = FWER), it is argued by Benjamini and Hochberg [1995] that since any procedure that controls the FWER also controls the FDR, it can be less stringent if a procedure controls the FDR only and a gain in power may be expected.

According to Benjamini and Hochberg [1995], their approach to multiple significance testing is philosophically different from many classical approaches which requires the control of the FWER in the strong sense. They proved that The BH procedure controls the FDR instead, and thereby also controls the FWER in the weak sense for the independent test statistics and Benjamini and Yekutieli [2001] showed that under a certain type of positive dependence, the procedure controls the FDR too. Thus, as we can see, similar to the Simes procedure, the dependence structure affects the use of the BH procedure. In the future, we will try to study how the BH procedure is related to dependence structure of the test statistics. Meanwhile we will continue to search multiple testing procedures which are potentially less conservative and more powerful.

BIBLIOGRAPHY

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1):289–300, 1995.

Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 26:1165–1188, 2001.

H.W. Block and H. Joe. Tail behavior of the failure rate functions of mixtures. *Lifetime Data Analysis*, 3(269-288), 1997.

H.W. Block, T.H. Savits, and M. Shaked. Some concepts of negative dependence. *Annals of Probability*, 10:765–772, 1982.

H.W. Block, T.H. Savits, and M. Shaked. A concept of negative dependence using stochastic ordering. *Statistics and Probability Letters*, 3:81–86, 1985.

H.W. Block, J. Mi, and T.H. Savits. Burn-in and mixed populations. *Journal of Applied Probability*, 30:692–702, 1993.

H.W. Block, Y.L. Li, and T.H. Savits. Initial and final behavior of failure rate functions for mixtures and systems. *Journal of Applied Probability*, 40:721–740, 2003a.

H.W. Block, T.H. Savits, and E.T. Wondmagegnehu. Mixtures of distributions with linear failure rates. *Journal of Applied Probability*, 40:485–504, 2003b.

C.A. Clarotti and F. Spizzichino. Bayes burn-in decision procedures. *Probability in the Engineering and Informational Sciences*, 4:437–445, 1991.

G.M. Constantine and T.H. Savits. A multivariate Faa Di Bruno formula with applications. *Transactions of the American Mathematical Society*, 348:503–520, 1996.

S Dudoit, J.P. Shaffer, and J.C. Boldrick. Multiple hypothesis testing in microarray experiments. *Statistical Science*, 18(1):71–103, 2003.

W. Feller. *An introduction to probability theory and its applications, Vol II.* John Wiley & Sons, New York, 1966.

R.E. Glaser. Bathtub and related failure rate characterizations. *Journal of the American Statistical Association*, 75:667–672, 1980.

L.J. Gleser. The gamma distribution as a mixture of exponential distributions. *The American Statistician*, 43:115–117, 1989.

R.C. Gupta and R. Warren. Determination of change points of non-monotomic failure rates. *Communications in Statistics: Theory and Methods*, 30:1903–1920, 2001.

J. Gurland and J. Sethuraman. Reversal of increasing failure rates when pooling failure data. *Technometrics*, 36:416–418, 1994.

Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75:800–803, 1988.

S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6:65–70, 1979.

G. Hommel. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75(2):383–386, 1988.

N.P. Jewell. Mixtures of exponential distributions. *Annals of Statistics*, 10:479–484, 1982.

R. Jiang and D.N.P. Murthy. Mixture of Weibull distributions-parametric characterization of failure rate functions. *Applied Stochastic Models and Data Analysis*, 14:47–65, 1998.

S. Karlin and Y. Rinott. Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis*, 10:467–498, 1980a.

S. Karlin and Y. Rinott. Classes of orderings of measures and related correlation inequalities. II. Multivariate reverse rules distributions. *Journal of Multivariate Analysis*, 10:499–516, 1980b.

C.D. Lai and M. Xie. *Stochastic ageing and dependence for reliability.* Springer, New York, 2006.

F. Proschan. Theoretical explanation of observed decreasing failure rate. *Tecknometrics*, 5: 375–383, 1963.

E. Samuel-Cahn. Is the Simes improved Bonferroni procedure conservative? *Biometrika*, 83 (4):928–933, 1996.

S.K. Sarkar. Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *Annals of Statistics*, 26:494–504, 1998.

J.P. Shaffer. Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, 81(395):826–831, 1986.

J.P. Shaffer. Multiple hypothesis testing: a review. *Annual Review of Psychology*, 46(561-584), 1995.

R.J. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.

E.T. Wondmagegnehu, J. Navarro, and P.J. Hernandez. Bathtub shaped failure rates from mixtures: a practical point of view. *IEEE Transactions on Reliability*, 54:270–275, 2005.