UTILIZING P-P PLOTS IN META-ANALYSIS
AS GENERAL MEASURES OF TREATMENT EFFECTS

by

Eric B. Holmgren

Technical Report No. 259
June 1989

Prepared under the Auspices
of
National Science Foundation
DMS 87-08083
Ingram Olkin, Project Director

Department of Statistics
Stanford University
Stanford, California
Abstract

The collection of statistical techniques used to integrate quantitative results from independent studies has come to be known as meta-analysis. Meta-analysis originated with methods of combining p-values to test for significance in a collection of study results which individually may be insignificant. Relatively recently a number of techniques have been developed which aim to explain the variability in results from independent studies each of which compares a treatment group with a control group. An essential aspect of these new meta-analytic techniques that distinguishes them from techniques of combining p-values is that they are based on a summary of the results of a study, the effect size, that quantifies the magnitude of the treatment effect rather than its statistical significance. The effect size is very well suited for analyzing studies that compare a treatment and control population from a simple location family. However in studies where the variances for the treatment and control populations may be unequal and in other more complicated settings the effect size provides only a partial description of the results of a study. There is thus a need for a measure of the treatment effect which is a more complete description of the results of a study than the effect size.

In this thesis we develop the p-p plot as a complete summary of the results of a study that compares a treatment group with a control group. A theoretical framework for interpreting p-p plots in meta-analysis is provided by considering the p-p plot as a maximal invariant and employing utility theory. In this framework the p-p plot can be used to order treatment effects just as a distribution function would be used to order risky alternatives. Maximum likelihood estimates of the p-p plot in the gaussian model where the treatment and control variances may be unequal and in a variety of models involving survival distributions are developed. Further, pooling procedures for estimating a common p-p plot and for facilitating the comparison of p-p plots across studies are presented.
Acknowledgements

I would like to thank my advisor, Ingram Olkin, for his help and support throughout the writing of this dissertation. His boundless energy has always been a source of inspiration. I would also like to thank Byron Brown and Helena Kraemer for reading the dissertation and providing helpful comments.

Special thanks go to my parents for their moral and financial support throughout my years in school and to my sister Laura who always had encouraging words for me.

Finally I would like to thank my wife Stella for making these past two years very special.
# Table of Contents

1 Introduction ............................................. 1
   1.1 P-p plots in meta-analysis ............................ 1
   1.2 A review of meta-analysis ............................ 9
   1.3 A review of probability plots .......................... 17

2 Data Sets .................................................. 24

3 Definitions and Preliminary Results .......................... 33
   3.1 Definitions ........................................... 33
   3.2 The maximal invariance of the p-p plot .................. 35
   3.3 A theoretical basis for interpreting p-p plots ............... 37

4 The P-p Plot for Gaussian Distributions ....................... 50
   4.1 Properties of the p-p plot ............................ 50
   4.2 Estimation of the p-p plot: homogeneous variances .............. 53
      4.2.1 Maximum likelihood estimation of the p-p plot .......... 53
      4.2.2 Simultaneous confidence bands for the p-p plot .......... 54
      4.2.3 Unbiased estimation of the p-p plot .................. 56
   4.3 Estimation of the p-p plot: unequal variances ................. 58
      4.3.1 Maximum likelihood estimation of the p-p plot .......... 58
      4.3.2 Simultaneous confidence bands for the p-p plot .......... 60
   4.4 An example .......................................... 63

5 The Sp-p Plot for Some Survival Distributions ................. 67
   5.1 The exponential model .................................. 68
      5.1.1 Properties of the sp-p plot .......................... 68
      5.1.2 Maximum likelihood estimation of the sp-p plot ............ 69
      5.1.3 Simultaneous confidence bands for the sp-p plot ............ 71
      5.1.4 Unbiased estimation of the sp-p plot .................. 72
   5.2 The gamma model ...................................... 73
      5.2.1 Properties of the sp-p plot .......................... 74
      5.2.2 Maximum likelihood estimation of the sp-p plot ............ 75
      5.2.3 Simultaneous confidence bands for the sp-p plot ............ 78
   5.3 The Weibull model .................................... 80
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>The open education data</td>
<td>26</td>
</tr>
<tr>
<td>2.2</td>
<td>The hypothetical estrogen receptor data</td>
<td>27</td>
</tr>
<tr>
<td>2.3</td>
<td>The hypothetical progesterone receptor data</td>
<td>28</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Two p-p plots that do not cross each other . . . . . . . . . . . . . . . . 20
1.2 Two p-p plots that cross each other . . . . . . . . . . . . . . . . . . . . 21
1.3 Three p-p plots which lie above, below, and on the 45 degree line . . . . 22
1.4 Two p-p plots that cross the 45 degree line . . . . . . . . . . . . . . . . . 23
2.1 The probability of survival in the treatment and control groups . . . . 29

for various concentrations of estrogen receptors
2.2 The probability of survival in the treatment and control groups . . . . 30

for various concentrations of progesterone receptors
2.3 The probability of survival in the treatment and control groups for various 31

concentrations of estrogen receptors for the simulated data
2.4 The probability of survival in the treatment and control groups for various 32

concentrations of progesterone receptors for the simulated data
3.1 P-p plots of $F_T$ versus $F_C$ for Gaussian distributions with $\sigma_T = \sigma_C$ . . . . 48
3.2 P-p plots of $F_T$ versus $F_C$ for Gaussian distributions with $\sigma_T \neq \sigma_C$ . . . . 49
4.1 Examples of p-p plots in the Gaussian model which are not convex, . . . . 65

concave or symmetric
4.2 The ML estimate of the p-p plot for study 6 of . . . . . . . . . . . . . . . . 66

the open education data
5.1 Sp-p plots of $F_T$ versus $F_C$ for the exponential distribution . . . . . . 89
5.2 Sp-p plots of $F_T$ versus $F_C$ for the gamma distribution . . . . . . . . 90
5.3 Examples of sp-p plots in the Weibull model which are not convex, . . . . 91

or concave
5.4 The empirical sp-p plots for the simulated estrogen receptor data . . . . 92

along with the ML estimates
5.5 The empirical sp-p plots for the simulated progesterone receptor data . . . 93

along with the ML estimates
6.1 The pooled estimate of the p-p plot for the open education data . . . . . . 109
6.2 ML estimates of the ordered sp-p plots for . . . . . . . . . . . . . . . . . . . 131

the estrogen receptor data
6.3 The ML estimate of the common sp-p plot for the progesterone receptor data
Chapter 1

Introduction

The collection of statistical techniques used to integrate quantitative results from independent studies has come to be known as meta-analysis. Meta-analysis originated with methods of combining p-values to test for significance in a collection of study results which individually may be insignificant. Relatively recently a number of techniques have been developed which aim to explain the variability in results from independent studies each of which compares a treatment group with a control group. An essential aspect of these new meta-analytic techniques that distinguishes them from techniques of combining p-values is that they are based on a summary of the results of a study, the effect size, that quantifies the magnitude of the treatment effect rather than its statistical significance. The effect size is very well suited for analyzing studies that compare a treatment and control population from a simple location family. However in studies where the variances for the treatment and control populations may be unequal and in other more complicated settings the effect size provides only a partial description of the results of a study. There is thus a need for a measure of the treatment effect which is a more complete description of the results of a study than the effect size. In this thesis we develop the p-p plot as such a measure of study results.

1.1. P-p plots in meta-analysis.

The meta-analysis setting in which we develop the p-p plot involves a collection of $K$ independent studies, each of which compares a treatment and a control group. The $K$ studies may be identical in the way they are conducted or they may differ in a number of experimental conditions which can significantly affect the results. Before introducing
the p-p plot we briefly review methods of meta-analysis in this setting. First we consider methods based on p-values.

The collection of meta-analytic techniques based on p-values provide a more powerful test of the hypothesis of no treatment effect by combining information from independent studies. Let \( p_1, p_2, \ldots, p_K \) denote p-values from \( K \) independent studies and suppose \( p_i \) corresponds to a test of the null hypothesis, \( H_i \), that there is no difference between the treatment and control groups in the \( i \)'th study. Considering these \( K \) studies together it is natural to ask whether the hypothesis of no difference between the treatment and control groups in every study is tenable. This hypothesis is called the combined null hypothesis. When the studies under consideration are exactly the same and test the same hypothesis, the combined null hypothesis is just the null hypothesis under consideration in each study. In the case where the null hypothesis in each study refers to a different set of experimental conditions, the combined null hypothesis may be interpreted as stating there is no treatment effect in any one of the circumstances examined.

Tests of the combined null hypothesis have been proposed by many authors. One of the earliest tests was proposed by Tippett in 1931. If we let \( p_{(1)}, p_{(2)}, \ldots, p_{(K)} \) denote the ordered p-values then Tippett's test rejects the combined null hypothesis at significance level \( \alpha \) if

\[
p_{(1)} \leq 1 - (1 - \alpha)^{1/K}.
\]

The critical value, \( 1 - (1 - \alpha)^{1/K} \), is the \( \alpha \)'th quantile of the first order statistic of a sample of size \( K \) from the uniform distribution on \([0, 1]\). Because this test is based on the first order statistic it is designed to detect alternatives to the combined null hypothesis where the treatment and control distributions are significantly different in just a few studies.

One of the most useful tests of the combined null hypothesis was proposed by Fisher in 1932. Fisher's test rejects the combined null hypothesis at significance level \( \alpha \) if

\[
-2 \log \left( \prod_{i=1}^{K} p_i \right) > \chi^2_{2K}(\alpha).
\]

The chi-square distribution arises from the fact that under the combined null hypothesis
each p-value has a uniform distribution and hence $-2\log p_i$ has a $\chi^2$ distribution with 2 degrees of freedom. Because this test is based on all of the p-values it is designed to detect alternatives to the combined null hypothesis where the treatment and control distributions differ in most of the studies.

The following simple example illustrates the use of Tippett’s and Fisher’s test. Suppose that there are 5 studies with p-values as follows:

<table>
<thead>
<tr>
<th>Study</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>.15</td>
<td>.13</td>
<td>.16</td>
<td>.20</td>
<td>.12</td>
</tr>
</tbody>
</table>

Each of these studies tests the null hypothesis that there is no difference between the treatment and control groups and reports the corresponding p-value. Tippett’s test statistic for this example is .12 and the corresponding p-value is

$$1 - (1 - .12)^5 = .47.$$  

Thus the combined null hypothesis looks reasonable judging by the smallest p-value. The reason for this is that there is not a single p-value much smaller than what would be expected in a random sample from the uniform distribution.

Fisher’s test statistic for this example is

$$-2\log(.15 \times .13 \times .16 \times .20 \times .12) = 19.00.$$  

From the chi-square distribution with 10 degrees of freedom we obtain a p-value of .04. Hence the collection of p-values is more significant than any of the p-values alone. The reason for the smaller p-value in Fisher’s test is that the reported p-values are consistently not quite significant and hence are not uniformly distributed on the unit interval as would be the case if the combined null hypothesis were true. This example shows the usefulness of Fisher’s combined test procedure for accumulating evidence from a number of independent studies.

The other set of statistical techniques that can be applied to a collection of studies that compare a treatment and a control group are based on a measure of the treatment
effect called the effect size. Glass (1976) first proposed the effect size as a measure of treatment effects useful in meta-analysis. The effect size is defined as the difference between the treatment and control group sample means normalized by the sample standard deviation of the control group. The effect size differs from the p-value in that it is not an indication of statistical significance but a measure of the size of the treatment effect. That is, whereas increasing the sample size will reduce the p-value if there is a treatment effect, the sample size has very little impact on the expected effect size. Thus unlike an analysis of p-values, an analysis of effect sizes can provide some insight into the factors affecting the subject of the study.

In addition to being invariant with respect to changes in the sample size the effect size is invariant with respect to linear changes of the scale of measurement. This is readily seen since a linear change in the measurement scale from \( x \) to \( ax + b \) multiplies both the difference between the sample means and the standard deviation of the control group by the constant \( a \). Thus if two studies differ only in the scales of measurement, the effect sizes for the two studies will be the same. This property of the effect size is very desirable in meta-analysis since it is essential to express results from different studies in comparable terms.

There are many techniques available for analyzing effect sizes. If the studies being considered appear to be very similar then the corresponding effect sizes can be combined using a weighted average to obtain a more accurate estimate of the treatment effect. Alternatively if there are important differences between the studies, the between study variability in the effect sizes can be modeled using the characteristics which distinguish the studies. As an example consider the following table.

<table>
<thead>
<tr>
<th>Study</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>25</td>
<td>35</td>
<td>45</td>
<td>55</td>
<td>65</td>
</tr>
<tr>
<td>effect size</td>
<td>1.18</td>
<td>1.02</td>
<td>.83</td>
<td>.57</td>
<td>.42</td>
</tr>
</tbody>
</table>

Each study compares a treatment group and a control group and reports the resulting effect size. There is a clear inverse relationship between age and the effect size in this
collection of studies indicating that the treatment loses its effectiveness as people get older. Thus it might be useful to construct a linear model relating the magnitude of the effect size to the age of the study group. There are other methods for analyzing effect sizes in addition to these. See Hedges and Olkin (1985) for a complete presentation of these techniques.

In some models, the effect size is not only invariant with respect to linear transformations but is a complete summary of the information in a study that is independent of the measurement scale. One example of such a model is that where the treatment and control distributions are Gaussian with identical variances. Indeed, for any model where the treatment and control distributions differ only by a location parameter the effect size is a complete summary of the information that is independent of the measurement scale. On the other hand, there are many models for the treatment and control populations where the effect size is only a partial summary of this information. One example is the model where the treatment and control distributions can differ in both location and scale. Another example is found in the nonparametric setting where the treatment and control distributions are allowed to be any absolutely continuous distribution functions.

In those situations where the effect size is not a complete summary of the information in a study that is independent of the measurement scale, an analysis of effect sizes from a number of studies provides a partial picture of the collection of study results. This suggests a need for representations of the results of a study which are more complete than the effect size and yet remain comparable across studies. In this thesis we consider a representation of the results of a study which is both a complete summary and comparable across studies under very general conditions, ranging from the standard effect size model, to models involving distributions of survival times, to nonparametric models. This representation is called the percentage-percentage or p-p plot.

The p-p plot is defined as the graph of the treatment distribution versus the control distribution. Consequently it can be expressed in functional form as

\[
F_T(F_C^{-1}(p)) \quad \text{or} \quad F_C(F_T^{-1}(p))
\]
depending on which axis is chosen to represent the treatment distribution. In this thesis we choose to study \( F_T(F_C^{-1}(p)) \) because it can be interpreted as a canonical representation of the treatment distribution which is comparable across studies. To see this let \( F_T \) and \( F_C \) denote the respective distribution functions for the treatment and control groups and let \( X_T \) and \( X_C \) be random variables associated with these distributions. Since the p-p plot may be expressed in functional form as

\[
F_T(F_C^{-1}(p)) = P[X_T < F_C^{-1}(p)] = P[F_C(X_T) < p]
\]

the p-p plot is the distribution function of \( F_C(X_T) \). So since \( F_C(\cdot) \) transforms \( X_C \) into a random variable with uniform distribution on the unit interval, the p-p plot is a graph of the treatment distribution with the measurement scale transformed so that the control distribution is uniform on the unit interval.

The formulation of the p-p plot as a canonical representation of the treatment distribution suggests that the p-p plot can be used in the same way as a distribution function as an expression of a treatment effect. Treatment distributions from different studies can be compared directly to gauge the relative magnitudes of treatment effects when the control distributions are homogeneous across studies. Since a single p-p plot is a representation of the treatment distribution corresponding to a uniform control distribution, a collection of p-p plots comprise representations of the treatment distributions corresponding to a common control distribution. Thus p-p plots from different studies can be compared directly as measures of treatment effects.

Another expression similar to \( F_T(F_C^{-1}(p)) \) which may be considered as a representation of the treatment effect is

\[
F_T^{-1}(F_C(x)) \quad \text{or} \quad F_C^{-1}(F_T(x))
\]

These functions are commonly known as q-q plots and have been studied by Doksum (1974), Doksum (1976) and Switzer (1976) as measures of treatment effects. The reason p-p plots and not q-q plots are appropriate as measures of treatment effects in meta-analysis is that p-p plots are invariant with respect to monotone transformations of the
measurement scale whereas q-q plots are not. Thus p-p plots have the advantage over q-q plots that they can be used to compare two studies with substantially different control groups.

To illustrate the comparison of treatment effects using p-p plots, suppose we are comparing two p-p plots, denoted by $P_1$ and $P_2$, and that $P_1$ lies below $P_2$ everywhere as in Figure 1.1. Then the standardized treatment distribution corresponding to $P_1$ is everywhere less than the standardized treatment distribution for $P_2$, which implies that the treatment in Study 1 has a greater impact on the control population than the treatment in Study 2.

It need not be the case that one p-p plot lies everywhere below another. In Figure 1.2, $P_1$ lies below $P_2$ in the interval $[0, .5]$ and lies above $P_2$ in the interval $[.5, 1]$. In this situation we cannot conclude that the treatment in the first study has a greater impact than the treatment in the second. However we can make the following statement. The treatment in the first study has a greater impact than the treatment in the second study over the interval $[0, .5]$ whereas the reverse is true over the interval $[.5, 1]$.

A single p-p plot also provides information about how the treatment and control distributions in the corresponding study compare. To illustrate this we present a few examples in Figure 1.3. The p-p plot marked $P_1$ in Figure 1.3 is a 45 degree line which indicates that $F_T(x) = F_C(x)$ for every $x$ and hence that there is no treatment effect. The p-p plot marked $P_2$ lies everywhere above the 45 degree line which indicates that $F_T(x) > F_C(x)$ for all $x$ and hence that the control is preferred to the treatment. The p-p plot marked $P_3$ lies everywhere below the 45 degree line which implies that the treatment is preferred to the control. Finally the p-p plots in Figure 1.4 lie both above and below the 45 degree line. For these p-p plots we cannot say definitively that the treatment or the control is preferred. However we can assert that over the interval where the p-p plot lies below the 45 degree line the treatment is preferred to the control while the reverse is true over the interval where the p-p plot lies above the 45 degree line. To illustrate this point consider the p-p plot marked $P_1$ in Figure 1.4. Suppose that this is a p-p plot of time to
death under a treatment versus time to death under a control. If an individual facing the choice between the treatment and the control is mainly concerned with staying alive in the short run then he will choose the treatment. However if the individual is solely concerned with improving his chances of being alive in the long run he will choose the control.

At this point we have presented the p-p plot as a transformed treatment distribution and indicated how a single or several p-p plots can be interpreted. The presentation of the results of a study in terms of a p-p plot is especially useful when it is natural to deal with distribution functions as representations of possible outcomes. In survival analysis and reliability theory it is advantageous for reasons of interpretation to deal with survival functions

$$\bar{F}(x) = 1 - F(x)$$

instead of distribution functions $F(x)$. Thus in these areas it is natural to express the treatment effect in terms of survival functions instead of distribution functions as with the p-p plot. This motivates the definition of what we shall call the survival percentage- percentage plot or simply the sp-p plot. The sp-p plot is defined as the graph of the treatment survival function versus the control distribution function and can be interpreted as the graph of the survival function for the treatment group with the scales adjusted so that the control distribution is uniform on the unit interval. This interpretation in terms of survival functions makes the sp-p plot well suited for meta-analysis in the areas of survival analysis and reliability theory.

To develop the use of the sp-p plot as a measure of treatment effects we appeal to the following simple relation between the p-p and sp-p plots. The sp-p plot is one minus the p-p plot. From this relation we can see that there is a 1-1 correspondence between the p-p plot and the sp-p plot. Thus the sp-p plot is a complete summary of the information in an experiment that is independent of the measurement scale whenever the p-p plot is and conversely. Hence the reasons for using the sp-p plot instead of the effect size as a summary of the information in an experiment that is independent of the measurement scale are the same as the reasons for using the p-p plot instead of the effect size.
The relation between the p-p and sp-p plots also allows us to deduce the following guidelines for interpreting sp-p plots from the rules for interpreting p-p plots. If the sp-p plot for the i'th study lies above the sp-p plot for the j'th study then the treatment has a greater impact in the i'th study than the j'th. An sp-p plot which lies on the line segment C from (0,1) to (1,0) indicates no treatment effect. An sp-p plot which lies above C indicates a positive treatment effect while an sp-p plot below C indicates a negative treatment effect. For an sp-p plot that lies both above and below C we cannot say that the treatment is preferred to the control in all circumstances or vice versa.

In the remainder of this chapter we review the areas of meta-analysis and probability plotting. Meta-analysis is reviewed in Section 2 whereas probability plots are reviewed in Section 3. In Chapter 2 we present two data sets which will be used to illustrate the use of p-p and sp-p plots. In Chapter 3 we develop the p-p plot as a summary of experimental results using the concept of maximal invariance. In addition we employ utility theory to show that a meta-analyst should base his comparison of treatment effects across studies on the p-p plot. Chapter 4 deals with estimating the p-p plot in the models where the treatment and control distributions are Gaussian with equal and unequal variances. Chapter 5 studies estimation of the p-p plot when the distributions are survival distributions, namely exponential gamma and Weibull. Finally Chapter 6 deals with pooling information from a number of studies.

1.2. A review of meta-analysis.

In Section 1 we indicated that there are two basic approaches to a meta-analysis of studies that compare a treatment and a control group. One approach is based on p-values and the other is based on effect sizes. In this section we consider both approaches in greater detail. First we present additional methods of combining p-values and some analytical results that help in determining which methods are good. Then we describe methods of meta-analysis based on effect sizes.

We have already described Tippett’s and Fisher’s procedures for combining p-values. These are by no means the only procedures for combining p-values from independent
studies. Here we outline a general approach to combining p-values which includes the inverse normal method and then go on to consider a few methods of combining p-values that account for sample sizes.

The inverse normal procedure is one of a class of methods that can be used to combine p-values. It was proposed independently by Stouffer, Suchman, De Vinney, Star, and Williams (1949), and by Liptak (1958) and is based on the statistic

\[ Z = \frac{\Phi^{-1}(p_1) + \cdots + \Phi^{-1}(p_K)}{\sqrt{K}}. \]

Under the combined null hypothesis \( \Phi^{-1}(p_i) \) has a standard normal distribution for every \( i \) and hence so does \( Z \). Thus the inverse normal test rejects the combined null hypothesis at significance level \( \alpha \) if \( Z \) exceeds the appropriate quantile of the standard normal distribution. This procedure can be generalized by replacing \( \Phi^{-1} \) with the inverse cumulative distribution function for any distribution function whose convolution is readily calculated. For example one could use the inverse of the exponential cumulative distribution function instead of \( \Phi^{-1} \) and compare \( Z \) with a gamma distribution. In this way the idea behind the inverse normal procedure leads to a whole class of procedures for combining p-values.

All the procedures presented here and in Section 1 have been based solely on the p-values reported by the studies under consideration. When information concerning sample sizes is available these procedures are not ideal. To illustrate, suppose that of 5 studies, 4 have sample sizes of 10 and p-values of .05 and one has a sample size of 10,000 and a p-value of .01. In this situation ignoring the sample sizes is misleading since the 5'th study is based on so many more observations than the other studies. Thus procedures that weight p-values according to the corresponding sample sizes are of interest. Two procedures which account for sample sizes are the weighted chi-square and the weighted inverse Gaussian methods.

The weighted chi-square test is based on the statistic

\[ P = -2 \log(p_1^{w_1} p_2^{w_2} \cdots p_K^{w_K}). \]
$w_1$ through $w_K$ can be chosen in any manner but what is of interest here is that they may be chosen in proportion to the sample sizes of the corresponding p-values. The distribution of $P$ under the combined null hypothesis is a weighted sum of chi-square variables. Simple representations for this distribution are known only for special cases, and in general the critical values for this test have to be computed via a simulation.

The weighted inverse Gaussian method suggested by Mosteller and Bush (1954) is another method that can account for sample sizes and is based on the statistic

$$Z_w = \frac{w_1 \Phi^{-1}(p_1) + \cdots + w_K \Phi^{-1}(p_K)}{\sqrt{w_1^2 + \cdots + w_K^2}},$$

where the weights, $w_1$ through $w_K$, may be chosen in proportion to the corresponding sample sizes. As with $Z$, $Z_w$ has a standard normal distribution under the combined null hypothesis and is rejected if $Z_w$ exceeds the appropriate quantile of a standard normal distribution.

With all of the different procedures for combining p-values that are available there is a need for analytical results that indicate what are good and what are poor procedures. There are two kinds of analytical results available. One set of results is based on decision theory and the other set is based on asymptotic considerations.

The collection of results based on decision theory are concerned with describing admissible tests of the combined null hypothesis. An admissible test is one which has no competitor with uniformly better power and the same type I error. If the model under consideration allows the cumulative distribution function for a single p-value to be any monotone decreasing function then a test is admissible if and only if it is what is called a monotone test. A monotone test is one which whenever it rejects the combined null hypothesis for one collection of p-values $(p_1, p_2, \cdots, p_K)$ rejects for any other collection of p-values $(\tilde{p}_1, \tilde{p}_2, \cdots, \tilde{p}_K)$ such that $\tilde{p}_j \leq p_j$ for all $j$. Thus a test which isn’t monotone can be improved upon and in general one should avoid using such tests. The Fisher and Tippett procedures as well as the inverse normal procedure are monotone. The combined test procedure based on $p_{(2)}$ is an example of a test which isn’t monotone.
Birnbaum (1954) proved another admissibility result for tests of the combined null hypothesis. He showed that for p-values, \( p_i \), computed from statistics, \( T_i \), whose distributions under the null and alternative hypotheses belong to the exponential family, a combined test is admissible if the acceptance region in terms of the \( T_i \) is convex. Fisher’s and Tippett’s procedures as well as the inverse Gaussian procedure all have convex acceptance regions and are monotone. \( p(2) \) is also an example of a test which doesn’t have a convex acceptance region.

A few authors have examined the power of combined test procedures under the assumption that the statistics on which the p-values are based belong to a particular parametric family of distributions. Oosterhof (1969) studies the power of combined test procedures assuming that the underlying statistics are normally distributed. Kozioł and Pearlman (1978) examine the power of combined test procedures assuming that the underlying statistics have non-central chi-square distributions.

Next we consider some asymptotic results. A measure of comparison that turns out to be useful is what is called Exact Bahadur Efficiency. Let \( T_{n_1}^{(1)} \) and \( T_{n_2}^{(2)} \) be two competing test statistics based respectively on sample sizes of \( n_1 \) and \( n_2 \). Let \( K^{(i)}(\epsilon) \) be the smallest sample size such that \( T_{n_i}^{(i)} \) is significant at level \( \epsilon \) for all \( n_i > K^{(i)}(\epsilon) \). Then

\[
\lim_{\epsilon \to 0} \frac{K^{(2)}(\epsilon)}{K^{(1)}(\epsilon)}
\]

is the Exact Bahadur Efficiency of \( T_{n_1}^{(1)} \) relative to \( T_{n_2}^{(2)} \). In typical cases the Exact Bahadur Efficiency of \( T_{n_1}^{(1)} \) relative to \( T_{n_2}^{(2)} \) is computed by first calculating the exact slope for each test. The exact slope of \( T_n \) is defined to be that quantity \( c(\theta) \) such that

\[
-\frac{2}{n} \log(1 - F_n(T_n)) \to_p c(\theta)
\]

where \( 1 - F_n(T_n) \) is the p-value for the statistic \( T_n \) based on a sample size of \( n \). Given the exact slope of \( T_{n_1}^{(1)} \) and \( T_{n_2}^{(2)} \) the Exact Bahadur Efficiency of \( T_{n_1}^{(1)} \) relative to \( T_{n_2}^{(2)} \) is simply \( c_1(\theta)/c_2(\theta) \).

Now we present some exact slopes for a few of the combined test procedures we have
presented. Suppose that we want to combine $K$ independent tests $T_{n_1}^{(1)}, T_{n_2}^{(2)}, \ldots, T_{n_K}^{(K)}$. Let $1 - F_{n_i}(T_{n_i}^{(i)})$ be the p-value associated with the statistic $T_{n_i}^{(i)}$ based on a sample size of $n_i$ and assume that for each $i = 1, \ldots, K$\[ -\frac{2}{n_i} \log(1 - F_{n_i}(T_{n_i}^{(i)})) \to c_i(\theta) \]with probability 1 as $n_i \to \infty$. Assume also that $n$, $K$, and the sample sizes $n_1, \ldots, n_K$ satisfy $n K = n_1 + \cdots + n_K$ and\[ \lim_{n \to \infty} \frac{n_i}{n} = \lambda_i \quad i = 1, \ldots, K. \]

Then we have\[ -\frac{2}{n} \log(1 - F_{n}(T_{n_i}^{(i)})) \to \lambda_i c_i(\theta), \]with probability 1. Thus for Fisher's method of combining p-values the exact slope $c_F(\theta)$ is\[ c_F(\theta) = \sum \lambda_i c_i(\theta). \]

For the inverse Gaussian method of combining p-values the exact slope $c_G(\theta)$ equals\[ c_G(\theta) = \frac{1}{K} \left[ \sum_i (\lambda_i c_i(\theta))^{1/2} \right]^2. \]

For Tippett's procedure which is based on the smallest p-value the exact slope $c_T(\theta)$ is\[ c_T(\theta) = \max_i \lambda_i c_i(\theta). \]

To compare different procedures for combining p-values one thus needs to evaluate the particular situation at hand. However one can make the following statement about Fisher's test which was proved by Littel and Folks (1973). Among all monotone tests, no test is more efficient in terms of Exact Bahadur Efficiency than Fisher's test. To see this let $T_n[T_{n_1}^{(1)}, \ldots, T_{n_K}^{(K)}]$ be a combined test statistic which is monotone increasing in each of its arguments and lets compute its exact slope. By the monotonicity of $T_n$\[ -\frac{2}{n} \log \left( 1 - F_n\left( T_n[T_{n_1}^{(1)}, \ldots, T_{n_K}^{(K)}] \right) \right) \geq -\frac{2}{n} \log \left( \Pi_{i=1}^K \left[ 1 - F_{n_i}(T_{n_i}^{(i)}) \right] \right). \]
Simple algebra now yields

\[
-\frac{2}{n} \log \left( 1 - F_n \left( T_{n1}, \ldots, T_{nK} \right) \right) \geq \sum_{i=1}^{K} -\frac{2}{n} \log(1 - F_{n_i}(T_{n_i})) \rightarrow_P \sum \lambda_i c_i(\theta).
\]

This shows that every monotone test has an exact slope which is less than or equal to the exact slope for Fisher’s test, that is no monotone test is more efficient than Fisher’s test. If one compares \(c_G(\theta)\) and \(c_T(\theta)\) with \(c_F(\theta)\) directly one finds as stated above that \(c_F(\theta)\) is the largest. Clearly \(c_F(\theta)\) is larger than \(c_T(\theta)\) and \(c_F(\theta)\) is larger than \(c_G(\theta)\) since sample variances are always positive.

Now we turn to methods of meta-analysis based on the effect size. Recall that the effect size is defined as the difference between the treatment and control group means normalized by the standard deviation of the control group. In the standard effect size model where the treatment and control groups are Gaussian with means \(\mu_T\) and \(\mu_C\) and variances \(\sigma_T = \sigma_C\), the maximum likelihood estimate of the effect size is

\[
g = \frac{\bar{X}_T - \bar{X}_C}{s_C},
\]

where \(\bar{X}_T\) and \(\bar{X}_C\) are respectively the sample means of the treatment and control groups and \(s_C\) is the sample standard deviation of the control group. Now

\[
\sqrt{\frac{(n_C - 1) n_T}{n_C + n_T}} \frac{\bar{X}_T - \bar{X}_C}{s_C}
\]

has a noncentral t-distribution with non-centrality parameter

\[
\sqrt{\frac{n_T n_C}{n_T + n_C} \delta}.
\]

Since the expected value of a noncentral t-distribution with noncentrality parameter \(\eta\) and degrees of freedom \(v\) is

\[
(v/2)^{1/2} \frac{\Gamma((v-1)/2)}{\Gamma(v/2)} \eta
\]

an unbiased estimate of the effect size can be computed from the maximum likelihood estimate by multiplying the maximum likelihood estimate by an appropriate constant.
In addition to parametric estimates of the effect size a number of nonparametric estimates of the effect size have been proposed. Kraemer and Andrews (1982) proposed a non-parametric estimate based on the proportion of observations in the control group less than the median of the treatment group. If we let \( t \) denote this proportion then the Kraemer and Andrews estimate can be expressed as \( \Phi^{-1}(t) \). While this estimate is nonparametric, it has the drawback that constructing confidence intervals for this estimate is difficult. An alternative procedure for constructing a nonparametric estimate of the effect size is to compute the Mann-Whitney Statistic, \( t_{MW} \), and use \( \sqrt{2} \Phi^{-1}(t_{MW}) \) as an estimate of the effect size. Confidence intervals for this estimate of the effect size are readily obtained via the delta method since the asymptotic distribution of the Mann Whitney Statistic is well known. A drawback of this procedure is that it supposes that the treatment and control distribution can be transformed to Gaussian distributions with identical variances while the Kraemer and Andrews procedure supposes only that the distributions can be transformed to Gaussian distributions.

In the introduction we mentioned that effect size estimates can be combined via a weighted average to obtain a more accurate estimate of the population effect size. A test of whether or not the effect sizes are homogeneous across studies would be useful to see if combining results is reasonable. Hedges (1982) proposed such a test. Suppose that \( K \) studies are under consideration with sample sizes for the treatment and control groups given by \( n_{T,j} \) and \( n_{C,j} \), and maximum likelihood estimates of the effect size, \( g_j \). Let

\[
 n = \sum_{j=1}^{K} n_{T,j} + n_{C,j}, \quad \pi_j^T = n_{T,j}/n, \quad \pi_j^C = n_{C,j}/n. 
\]

The test Hedges proposed is based on the statistic

\[
 H_T = n \sum_{j=1}^{K} \frac{(g_j - g)^2}{\sigma_j^2(g_j)}
\]

where

\[
 g = \frac{\sum_{j=1}^{K} g_j/\sigma_j^2(g_j)}{\sum_{j=1}^{K} 1/\sigma_j^2(g_j)}
\]

and

\[
 \sigma_j^2(\delta_j) = \frac{\pi_j^T \delta_j^2 + \pi_j^C}{\pi_j^T \pi_j^C} + \frac{\delta_j^2}{2(\pi_j^T + \pi_j^C)}. 
\]
Under the hypothesis that the effect sizes are homogeneous across studies $H_T$ has asymptotically a chi-square distribution with $K - 1$ degrees of freedom. Thus if $H_T$ exceeds the $\alpha$'th quantile of a chi-square distribution with $K - 1$ degrees of freedom we reject the hypothesis that the effect sizes are homogeneous across studies at significance level $\alpha$.

If the null hypothesis of homogeneous effect sizes is accepted then one may combine the effect sizes from the various studies by a weighted average. On the other hand if the null hypothesis is rejected then one may consider dividing up the studies into disjoint groups and computing weighted averages of the effect sizes for the different groups. Once the $K$ studies are divided up into $k$ groups with $m_i$ studies in the $i$'th group one can ask if the effect sizes across the groups are homogeneous and if the effect sizes within groups are homogeneous.

Let $n_{i,j}^T$ and $n_{i,j}^C$ be the sample sizes for the treatment and control groups in the $j$'th study of the $i$'th group,

$$ n = \sum_{i=1}^{l} \sum_{j=1}^{m_i} n_{i,j}^T + n_{i,j}^C $$

$$ \pi_{i,j}^T = n_{i,j}^T / n \quad \pi_{i,j}^C = n_{i,j}^C / n. $$

For the problem of testing homogeneity across groups, Hedges proposed looking at the statistic

$$ H_B = n \sum_{i=1}^{l} \sum_{j=1}^{m_i} \frac{(g_{i,j} - g_{..})^2}{\sigma_{i,j}^2(g_{i,j})}, $$

where

$$ g_{i,.} = \frac{\sum_{j=1}^{m_i} g_{i,j} / \sigma_{i,j}^2(g_{i,j})}{\sum_{j=1}^{m_i} 1 / \sigma_{i,j}^2(g_{i,j})}, $$

$$ g_{..} = \frac{\sum_{i=1}^{l} \sum_{j=1}^{m_i} g_{i,j} / \sigma_{i,j}^2(g_{i,j})}{\sum_{i=1}^{l} \sum_{j=1}^{m_i} 1 / \sigma_{i,j}^2(g_{i,j})}, $$

and

$$ \sigma_{i,j}^2(\delta_{i,j}) = \frac{\pi_{i,j}^T + \pi_{i,j}^C}{\pi_{i,j}^T \pi_{i,j}^C} + \delta_{i,j}^2 / 2(\pi_{i,j}^T + \pi_{i,j}^C). $$

$H_B$ has asymptotically a chi-square distribution with $l - 1$ degrees of freedom. If $H_B$ exceeds the $\alpha$'th quantile of a chi-square distribution with $l - 1$ degrees of freedom then
the hypothesis that the effect sizes are homogeneous across classes is rejected at significance level $\alpha$.

For testing homogeneity within classes Hedges proposed using the statistic

$$H_W = \frac{n}{l} \sum_{i=1}^{l} \sum_{j=1}^{m_i} \frac{(g_{ij} - \bar{g}_i)^2}{\sigma_{ij}^2}.$$ 

$H_W$ has asymptotically a chi-square distribution with $n - l$ degrees of freedom and the hypothesis that the effect sizes are homogeneous within classes is rejected at significance level $\alpha$ if $H_W$ exceeds the $\alpha$'th quantile of a chi-square distribution with $n - l$ degrees of freedom.

These tests for the homogeneity of effect sizes are useful in deducing explanations for differences in effect sizes across studies. In addition to computing effect sizes for subgroups of studies one can construct continuous models and perform many other analyses of the effect sizes. See Hedges and Olkin (1985) for a complete presentation of these techniques.

### 1.3. A review of probability plots.

Probability plots, that is p-p and q-q plots, are most well known in statistics as graphical displays which aid in determining whether or not distributional assumptions are met. Wilk and Gnanadesikan (1968) is a standard reference in the literature. The p-p plot is constructed by plotting the percentiles of the empirical distribution versus the percentiles of the hypothetical distribution. A straight line from (0, 0) to (1, 1) indicates that the hypothetical and the empirical distributions agree. The q-q plot on the other hand is constructed by plotting the quantiles of the empirical distribution versus the quantiles of the hypothetical distribution. The straight line $y = x$ indicates that the empirical and hypothetical distributions agree, whereas any other straight line indicates that the treatment and control distributions differ by a scale and a location parameter. A common application of probability plots is the construction of a q-q plot to see if the assumption of normality is satisfied by residuals in a regression model or in an analysis of variance. Probability plots have proved useful in other contexts as well.
Doksum (1974) suggests using the q-q plot as a general description of the difference between two distributions when a simple shift model is not appropriate. Suppose that \( X \) and \( Y \) are random variables with distributions \( F \) and \( G \). In the simple shift model there exists a constant \( \delta \) such that \( X + \delta \) has the same distribution as \( Y \). Doksum showed that the q-q plot, \( G^{-1}(F(x)) \), is a generalization of \( \delta \) in the sense that

\[
X + (G^{-1}(F(x)) - x)
\]

has the same distribution as \( Y \). Switzer (1976) indicates how to use the Kolmogorov Smirnov statistic to construct confidence bands for the shift function

\[
(G^{-1}(F(x)) - x).
\]

In reliability theory Steck, Zimmer and Williams (1974) propose using the p-p plot as a device to decelerate data obtained from accelerated life testing. Let \( G(x) \) be the distribution of the lifetime of a transistor under accelerated conditions and \( F(x) \) be its distribution under normal conditions. Since the p-p plot, \( F(G^{-1}(x)) \), solves the following equation for \( h \),

\[
F(x) = h(G(x))
\]

the p-p plot relates the probabilities of survival obtained under accelerated conditions to the probabilities under normal conditions. Thus if the p-p plot for an accelerated life test is known from previous experience it can be used to translate probabilities computed under accelerated conditions to probabilities under normal conditions.

Finally the p-p plot has found some use in the medical literature under the name of the operating characteristic curve (OCC). Suppose that we have a procedure which yields a measurement between 0 and \( \infty \) and we declare a person to have a disease if the measurement exceeds a critical value \( B \). The operating characteristic curve relates the probability a healthy person is determined to have the disease to the probability a diseased person is determined to have the disease at various values for \( B \). Alternatively we may say that the OCC curve is the p-p plot of the distribution of measurements for the
population of healthy people versus the distribution of measurements for the population of diseased people. The OCC curve has proved useful in describing the properties of a diagnostic test.

As for the estimation of the p-p plot the following two papers are of interest. Pathak, Zimmer and Williams (1979) consider estimation of the p-p plot in the situation where the treatment and control distributions can be transformed to exponential distributions. Aly, Csorgo and Horvath (1988) present the asymptotic distribution of the empirical p-p plot which is defined as the graph of one empirical c.d.f versus the other.
Figure 1.1

Two p-p plots that do not cross each other.
Figure 1.2

Two p-p plots that cross each other.
Figure 1.3

Three p-p plots which lie above, below, and on the 45 degree line.
Figure 1.4

Two p-p plots that cross the 45 degree line.
Chapter 2

Data Sets

In this chapter we present two data sets which will be used throughout this thesis to illustrate the use of p-p and sp-p plots. The first data set is comprised of the results of a number of studies each of which examines the effect of open education on the problem solving ability of students. The second data set is a simulated example which mirrors the results of a clinical trial that was reported in the medical literature. The first data set will be used to illustrate techniques involving the normal distribution while the second data set will be used to illustrate techniques involving the gamma distribution.

Table 2.1 presents the results of a number of studies each of which examines the effect of open education on the problem solving ability of students. Each study reports the means and standard deviations for both the treatment and control groups. This data set has been reported previously by Hedges, Giaconia, and Gage (1981). An analysis of this data using effect sizes indicates that open education has no impact on the problem solving ability of students. In this thesis we reanalyze this data using p-p plots. Since p-p plots form a complete summary of the information in a study more generally than effect sizes, we hope to uncover an impact of open education on problem solving ability that was missed by the analysis of effect sizes.

The second data set is simulated and is motivated by the results of a large study conducted by the National Surgical Adjuvant Breast and Bowel Project (NSABP) and reported by Fisher (1983). The aim of the NSABP study was to examine the effectiveness of various chemo-therapy regimens initiated after surgical removal of primary cancer from the breast. As part of a brief description of the results of the study the graphs in Figures 2.1 and 2.2 were presented.
Figure 2.1 displays four graphs of the probability of survival at time $t$ for a treatment group and a control group. Each graph refers to survival at a specific concentration of estrogen receptors in the cancerous tissue. The graphs indicate that as the concentration of estrogen receptors increases the probability of survival for the control group declines while it increases for the treatment group. Figure 2.2 presents similar information relative to the concentration of progesterone receptors which indicates that as the concentration of progesterone receptors increases the probability of survival for both the treatment and control group improves. While the graphs in Figures 2.1 and 2.2 present a detailed description of the results of the study, they do not provide a representation of the treatment effect that is comparable across the different categories of estrogen and progesterone receptor concentrations. Since the sp-p plot is a representation of the treatment effect that is comparable across categories an examination of the sp-p plots corresponding to these diagrams can aid in assessing how the treatment effect varies as the progesterone and estrogen receptor concentrations change.

Tables 2.2 and 2.3 present the simulated data motivated by the NSABP study. Table 2.2 corresponds to the estrogen receptor data and presents the time until death for each individual in the treatment group and the control group. Table 2.3 corresponds to the progesterone receptor data and presents the same information as in Table 2.2. For the simulated data the probability of survival is graphed as a function of time in Figures 2.3 and 2.4.
Table 2.1

The open education data.

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th></th>
<th></th>
<th></th>
<th>Treatment</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>84</td>
<td>21.4</td>
<td>4.4</td>
<td>84</td>
<td>19.6</td>
<td>3.1</td>
<td>.473</td>
</tr>
<tr>
<td></td>
<td>142</td>
<td>5.7</td>
<td>1.9</td>
<td>142</td>
<td>5.0</td>
<td>1.6</td>
<td>.399</td>
</tr>
<tr>
<td></td>
<td>121</td>
<td>9.47</td>
<td>2.12</td>
<td>154</td>
<td>9.01</td>
<td>1.78</td>
<td>.240</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>59.4</td>
<td>8.8</td>
<td>40</td>
<td>62.9</td>
<td>6.4</td>
<td>.455</td>
</tr>
<tr>
<td></td>
<td>78</td>
<td>62.2</td>
<td>9.2</td>
<td>78</td>
<td>58.7</td>
<td>6.6</td>
<td>.437</td>
</tr>
<tr>
<td></td>
<td>112</td>
<td>10.7</td>
<td>2.1</td>
<td>104</td>
<td>10.6</td>
<td>3.2</td>
<td>.037</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>81.0</td>
<td>11.3</td>
<td>19</td>
<td>86.5</td>
<td>8.22</td>
<td>.564</td>
</tr>
<tr>
<td></td>
<td>73</td>
<td>8.11</td>
<td>2.13</td>
<td>60</td>
<td>10.02</td>
<td>1.79</td>
<td>.963</td>
</tr>
<tr>
<td></td>
<td>174</td>
<td>21.44</td>
<td>7.49</td>
<td>181</td>
<td>24.69</td>
<td>6.50</td>
<td>.464</td>
</tr>
</tbody>
</table>

$\delta_{pooled} = .000779 \quad \sigma_{\delta_{pooled}} = .0490$
Table 2.2

The hypothetical estrogen receptor data.

<table>
<thead>
<tr>
<th>Control Treatment</th>
<th>≥ 10 fmol</th>
<th>≥ 50 fmol</th>
<th>≥ 100 fmol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.85</td>
<td>14.45</td>
<td>10.65</td>
<td>5.94</td>
</tr>
<tr>
<td>11.53</td>
<td>19.65</td>
<td>3.24</td>
<td>1.96</td>
</tr>
<tr>
<td>1.66</td>
<td>11.07</td>
<td>2.82</td>
<td>2.67</td>
</tr>
<tr>
<td>5.78</td>
<td>23.83</td>
<td>2.37</td>
<td>4.60</td>
</tr>
<tr>
<td>3.64</td>
<td>10.77</td>
<td>3.67</td>
<td>2.69</td>
</tr>
<tr>
<td>2.37</td>
<td>21.21</td>
<td>1.16</td>
<td>2.87</td>
</tr>
<tr>
<td>18.11</td>
<td>3.61</td>
<td>3.75</td>
<td>5.37</td>
</tr>
<tr>
<td>8.98</td>
<td>3.53</td>
<td>8.13</td>
<td>11.55</td>
</tr>
<tr>
<td>6.80</td>
<td>22.43</td>
<td>2.86</td>
<td>5.12</td>
</tr>
<tr>
<td>7.37</td>
<td>21.00</td>
<td>1.72</td>
<td>4.97</td>
</tr>
<tr>
<td>14.09</td>
<td>7.93</td>
<td>2.85</td>
<td>4.48</td>
</tr>
<tr>
<td>2.71</td>
<td>10.23</td>
<td>4.35</td>
<td>0.34</td>
</tr>
<tr>
<td>4.66</td>
<td>8.60</td>
<td>8.70</td>
<td>9.18</td>
</tr>
<tr>
<td>16.96</td>
<td>8.77</td>
<td>11.09</td>
<td>2.02</td>
</tr>
<tr>
<td>10.22</td>
<td>25.77</td>
<td>10.88</td>
<td>3.09</td>
</tr>
<tr>
<td>1.49</td>
<td>16.09</td>
<td>0.82</td>
<td>4.48</td>
</tr>
<tr>
<td>19.92</td>
<td>9.58</td>
<td>1.68</td>
<td>12.02</td>
</tr>
<tr>
<td>0.75</td>
<td>20.05</td>
<td>3.43</td>
<td>0.32</td>
</tr>
<tr>
<td>2.38</td>
<td>2.88</td>
<td>10.42</td>
<td>1.03</td>
</tr>
<tr>
<td>7.42</td>
<td>10.85</td>
<td>7.03</td>
<td>2.26</td>
</tr>
<tr>
<td>11.80</td>
<td>13.23</td>
<td>6.48</td>
<td>6.90</td>
</tr>
<tr>
<td>13.44</td>
<td>8.69</td>
<td>3.75</td>
<td>2.89</td>
</tr>
<tr>
<td>10.54</td>
<td>14.29</td>
<td>0.36</td>
<td>0.53</td>
</tr>
<tr>
<td>5.63</td>
<td>40.13</td>
<td>0.85</td>
<td>4.11</td>
</tr>
<tr>
<td>1.73</td>
<td>9.94</td>
<td>2.02</td>
<td>3.07</td>
</tr>
<tr>
<td>12.05</td>
<td>7.85</td>
<td>10.32</td>
<td>1.11</td>
</tr>
<tr>
<td>8.68</td>
<td>5.15</td>
<td>7.11</td>
<td>10.61</td>
</tr>
<tr>
<td>6.20</td>
<td>5.35</td>
<td>5.35</td>
<td>6.24</td>
</tr>
<tr>
<td>10.20</td>
<td>36.49</td>
<td>5.07</td>
<td>0.68</td>
</tr>
<tr>
<td>11.77</td>
<td>7.57</td>
<td>12.08</td>
<td>2.27</td>
</tr>
</tbody>
</table>
Table 2.3

The hypothetical progesterone receptor data.

<table>
<thead>
<tr>
<th></th>
<th>≥ 10 fmol</th>
<th></th>
<th>≥ 50 fmol</th>
<th></th>
<th>≥ 100 fmol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Treatment</td>
<td>Control</td>
<td>Treatment</td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td>2.62</td>
<td>9.01</td>
<td>10.42</td>
<td>22.41</td>
<td>2.48</td>
</tr>
<tr>
<td></td>
<td>3.56</td>
<td>11.64</td>
<td>4.60</td>
<td>5.69</td>
<td>13.28</td>
</tr>
<tr>
<td></td>
<td>6.48</td>
<td>2.00</td>
<td>4.96</td>
<td>30.50</td>
<td>11.50</td>
</tr>
<tr>
<td></td>
<td>7.76</td>
<td>4.17</td>
<td>5.24</td>
<td>21.31</td>
<td>15.54</td>
</tr>
<tr>
<td></td>
<td>2.71</td>
<td>8.99</td>
<td>1.87</td>
<td>3.60</td>
<td>7.49</td>
</tr>
<tr>
<td></td>
<td>0.53</td>
<td>14.92</td>
<td>4.04</td>
<td>7.55</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>8.65</td>
<td>3.73</td>
<td>0.55</td>
<td>4.27</td>
<td>2.36</td>
</tr>
<tr>
<td></td>
<td>1.92</td>
<td>19.19</td>
<td>1.62</td>
<td>12.11</td>
<td>13.46</td>
</tr>
<tr>
<td></td>
<td>1.69</td>
<td>9.13</td>
<td>19.82</td>
<td>2.09</td>
<td>24.88</td>
</tr>
<tr>
<td></td>
<td>2.97</td>
<td>9.70</td>
<td>13.69</td>
<td>2.10</td>
<td>5.04</td>
</tr>
<tr>
<td></td>
<td>0.51</td>
<td>8.45</td>
<td>8.82</td>
<td>16.73</td>
<td>8.86</td>
</tr>
<tr>
<td></td>
<td>2.87</td>
<td>2.29</td>
<td>3.91</td>
<td>4.56</td>
<td>23.02</td>
</tr>
<tr>
<td></td>
<td>2.44</td>
<td>9.37</td>
<td>4.33</td>
<td>15.54</td>
<td>4.19</td>
</tr>
<tr>
<td></td>
<td>5.46</td>
<td>7.82</td>
<td>22.37</td>
<td>15.09</td>
<td>5.05</td>
</tr>
<tr>
<td></td>
<td>4.39</td>
<td>8.76</td>
<td>13.01</td>
<td>5.28</td>
<td>2.47</td>
</tr>
<tr>
<td></td>
<td>5.97</td>
<td>10.54</td>
<td>8.55</td>
<td>20.90</td>
<td>6.71</td>
</tr>
<tr>
<td></td>
<td>1.26</td>
<td>13.13</td>
<td>4.46</td>
<td>18.49</td>
<td>10.73</td>
</tr>
<tr>
<td></td>
<td>3.33</td>
<td>7.00</td>
<td>2.02</td>
<td>8.60</td>
<td>6.60</td>
</tr>
<tr>
<td></td>
<td>0.72</td>
<td>3.28</td>
<td>2.17</td>
<td>29.79</td>
<td>7.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22.15</td>
</tr>
</tbody>
</table>
The probability of survival in the treatment and control groups for various concentrations of estrogen receptors.

Figure 2.1

- **≥ 10 fmol**
  - **C**
  - **T**

- **≥ 30 fmol**
  - **C**
  - **T**

- **≥ 50 fmol**
  - **C**
  - **T**

- **≥ 100 fmol**
  - **C**
  - **T**

*Note: The diagram shows survival probability over time for different concentration levels (10, 30, 50, 100 fmol) with treatment (T) and control (C) groups.*
Figure 2.2

The probability of survival in the treatment and control groups for various concentrations of progesterone receptors.
Figure 2.3

The probability of survival in the treatment and control groups for various concentrations of estrogen receptors, for the simulated data.

- ≥ 10 fmol
- ≥ 50 fmol
- ≥ 100 fmol
Figure 2.4

The probability of survival in the treatment and control groups for various concentrations of progesterone receptors, for the simulated data.

\[ \geq 10 \text{ fmol} \]

\[ \geq 50 \text{ fmol} \]

\[ \geq 100 \text{ fmol} \]
Chapter 3

Definitions and Preliminary Results

In this chapter we present some definitions and preliminary results. In Section 1 the meta-analysis setting under consideration is presented and a formal definition of the p-p plot is given. In Section 2 it is shown, using the concept of a maximal invariant, that the p-p plot summarizes all of the information in a study that is independent of the measurement scale. Finally in Section 3 we provide a theoretical justification of the rules for interpreting p-p plots presented in Chapter 1.

3.1. Definitions.

The meta-analysis setting in which we develop the p-p plot consists of a collection of $K$ independent studies each of which compares a treatment group with a control group. The studies may be identical in the way they are conducted or there may be a number of important differences. We denote the observations from the $K$ studies under consideration as follows.

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_{i1}^T$  $i = 1, n_{T1}$</td>
<td>$X_{i1}^C$ $i = 1, n_{C1}$</td>
</tr>
<tr>
<td>2</td>
<td>$X_{i2}^T$  $i = 1, n_{T2}$</td>
<td>$X_{i2}^C$ $i = 1, n_{C2}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$K$</td>
<td>$X_{iK}^T$  $i = 1, n_{TK}$</td>
<td>$X_{iK}^C$ $i = 1, n_{CK}$</td>
</tr>
</tbody>
</table>

The superscript on $X$ indicates whether the observation is from the treatment or the control group. The second subscript denotes the study that the observation came from while the first subscript denotes the specific observation in the treatment or the
control group of the study. \( n_T \) and \( n_C \) are respectively the number of observations in the treatment and control group of the \( j \)'th study.

The observations in the treatment group for the \( j \)'th study are independent and identically distributed with cumulative distribution function \( F_T \). Likewise the observations in the control group for the \( j \)'th study are independent and identically distributed with distribution function \( F_C \). Further, we make the assumption that observations in different groups or different studies are independent. The following definition presents the p-p plot.

**Definition 3.1:** Let \( F_T(x) \) and \( F_C(x) \) be the respective cumulative distribution functions for the treatment and control group of a study. Then the p-p plot for the study is the graph of \( F_T(x) \) on the vertical axis versus \( F_C(x) \) on the horizontal axis.

Lemma 3.2 presents some equivalent definitions of the p-p plot.

**Lemma 3.2:** The p-p plot for \( F_T(x) \) and \( F_C(x) \) is

1. the function \( F_T(F_C^{-1}(p)) \),
2. the graph of the treatment distribution on a horizontal scale adjusted so that the control distribution is uniform on \([0, 1]\).

**Proof:** To prove (1) note that \( F_T(F_C^{-1}(p)) \) is the functional form of the graph of \( F_T \) versus \( F_C \). The proof of (2) follows from the fact that the p-p plot is the treatment distribution transformed by \( x \rightarrow F_C(x) \). This function transforms the control distribution to the uniform distribution on \([0, 1]\). ||

For the meta-analysis problem considered here we can construct \( K \) p-p plots, one for each study. Each p-p plot is a representation of the treatment effect in the associated study. The fact that a p-p plot can be thought of as a transformed treatment distribution corresponding to a uniform control distribution allows one to interpret the \( K \) p-p plots just as one would interpret \( K \) treatment distributions which all correspond to the same control distribution. This is the key to the usefulness of p-p plots in meta-analysis.
3.2. The maximal invariance of the p-p plot.

In this section we show that the p-p plot summarizes all the information in a study that is independent of the measurement scale. To do this we briefly review the concept of a maximal invariant. A detailed exposition of the concept of invariance in statistics can be found in Lehmann (1983). Let \( \mathcal{F} \) be a family of pairs of distributions defined on \( \mathcal{X} \) and let \( \mathcal{G} \) be a group of 1-1 transformations of \( \mathcal{X} \) onto itself. The concept of a maximal invariant is presented in Definition 3.4. First we need a preliminary definition.

**Definition 3.3:** \( \mathcal{F} \) is defined to be closed with respect to the group of transformations \( \mathcal{G} \) if for any pair of distributions \([F_1(x), F_2(x)] \in \mathcal{F}\), \([F_1(g(x)), F_2(g(x))] \) is also in \( \mathcal{F} \) for all \( g \in \mathcal{G} \).

**Definition 3.4:** Suppose that \( \mathcal{F} \) is closed with respect to \( \mathcal{G} \) and \( \phi \) is defined on \( \mathcal{F} \). Then \( \phi \) is invariant with respect to \( \mathcal{G} \) if

\[
\phi(F_1(g(x)), F_2(g(x))) = \phi(F_1(x), F_2(x))
\]

for all \( g \in \mathcal{G} \) and for all \([F_1, F_2] \in \mathcal{F}\). \( \phi(F_1(x), F_2(x)) \) is a maximal invariant if it is invariant and if whenever \( \phi(F_1(x), F_2(x)) = \phi(F'_1(x), F'_2(x)) \) there exists a transformation \( g \in \mathcal{G} \) such that \( F'_1(x) = F_1(g(x)) \) and \( F'_2(x) = F_2(g(x)) \).

The group of transformations \( \mathcal{G} \) divides \( \mathcal{F} \) into equivalence classes of pairs of distributions which are related by transformations in \( \mathcal{G} \). If we think of the transformations as changing the scale of measurement employed in a study, then each equivalence class represents all possible measurements of a specific study observation. The function \( \phi \) is a maximal invariant if it is constant in each equivalence class and if it assumes distinct values on distinct equivalent classes. Thus a maximal invariant summarizes all the information in an experiment that is independent of the scale of measurement employed. Now we show that the p-p plot is a maximal invariant under very general conditions.

**Lemma 3.5:** Let \( \mathcal{F} \) denote a family of pairs of distributions that is closed under the group of transformations \( \mathcal{G} \). Further, suppose that given any two pairs of distributions, \((F_{T1}(t), F_{C1}(t))\) and \((F_{T2}(t), F_{C2}(t))\), in \( \mathcal{F} \), there exists an element, \( h \), of \( \mathcal{G} \) such that
Chapter 3: Definitions and Preliminary Results

\( F_{C1}(t) = F_{C2}(h(t)) \). Then the p-p plot is a maximal invariant with respect to the group of transformations \( \mathcal{G} \).

**Proof:** Since for all \((F_T(t), F_C(t)) \in \mathcal{F}\) the graphs, \( F_T(t) \) versus \( F_C(t) \) and \( F_T(g(t)) \) versus \( F_C(g(t)) \), are identical for all \( g \in \mathcal{G} \), the p-p plot is invariant. Now suppose that the two p-p plots, \( F_{T1}(t) \) versus \( F_{C1}(t) \) and \( F_{T2}(t) \) versus \( F_{C2}(t) \), are identical. By hypothesis there exists a transformation \( h \) in \( \mathcal{G} \) such that

\[
F_{C1}(t) = F_{C2}(h(t)) \quad \forall \ t.
\]

Since the graphs \( F_{T2}(h(t)) \) versus \( F_{C2}(h(t)) \) and \( F_{T2}(t) \) versus \( F_{C2}(t) \) are identical it follows that

\[
F_{T1}(t) = F_{T2}(h(t)) \quad \forall \ t,
\]

and hence that the pairs of distributions \((F_{T1}, F_{C1})\) and \((F_{T2}, F_{C2})\) differ only by a transformation in \( \mathcal{G} \).||

Now we apply Lemma 3.5 to some specific models to better understand when a p-p plot is a maximal invariant. If \( \mathcal{F} \) is defined as the collection of all pairs of Gaussian distributions with identical variances and \( \mathcal{G} \) is defined as the group of linear transformations, then the lemma shows that the p-p plot is a maximal invariant. The effect size is also a maximal invariant in this model which is the standard model for evaluating effect sizes. Figure 3.1 displays some p-p plots for various pairs of distributions in \( \mathcal{F} \).

A slight variation of the above model in which \( \mathcal{F} \) is defined as the collection of all pairs of Gaussian distributions with variances not necessarily equal and \( \mathcal{G} \) is defined as above strips the effect size of its maximal invariance property. To see this note that the following two pairs of distributions

\[
F_{T1} = N(\mu_T, \sigma^2), \quad F_{C1} = N(\mu_C, \sigma^2),
\]

\[
F_{T2} = N(\mu_T, 2\sigma^2), \quad F_{C2} = N(\mu_C, \sigma^2)
\]

do not differ by a linear transformation yet have the same effect size \((\mu_T - \mu_C)/\sigma\). On the other hand applying Lemma 3.5 we see that the p-p plot continues to be a maximal
invariant. Thus in this model the effect size provides only a portion of the information incorporated in the p-p plot. A few p-p plots corresponding to elements of $\mathcal{F}$ are graphed in Figure 3.2.

Finally consider the nonparametric model where $\mathcal{F}$ is the collection of all pairs of continuous distributions and $\mathcal{G}$ is the group of all continuous monotone increasing transformations. Lemma 3.5 is applicable once again and shows that the p-p plot is a maximal invariant. The p-p plots displayed in Figures 3.1 and 3.2 are also examples of p-p plots in the nonparametric case. Indeed, in this case, any monotone increasing function from $(0,0)$ to $(1,1)$ is an example of a p-p plot.

In summary, Lemma 3.5 and the three examples demonstrate that the p-p plot is a generalization of the effect size which preserves the maximal invariance property. This maximal invariance property can be interpreted as meaning that the p-p plot summarizes all of the information in a study that is independent of the scale of measurement. In models where the effect size isn't a maximal invariant while the p-p plot is, the effect size provides only a portion of the information incorporated in the p-p plot.

### 3.3. A theoretical basis for interpreting p-p plots.

In this section we turn our attention to constructing a theoretical basis for interpreting p-p plots. Recall that we are considering a meta-analysis of $K$ studies each of which compares a treatment group with a control group. Each such study is associated with a pair of distributions that represent the laws governing the treatment and control observations. The task of making comparisons between a number of different studies in this setting can thus be thought of as making a number of comparisons between pairs of distributions. Here we show that if the preferences of a meta analyst among all possible combinations of treatment and control distributions can be described by a preference relation satisfying certain properties then there exists a function $U(\cdot,\cdot)$ of the treatment and control distributions such that the $i$'th study is preferred to the $j$'th study if and only if

$$U(F_{Ti},F_{Ci}) \geq U(F_{Tj},F_{Cj}).$$
$U(\cdot, \cdot)$ can be thought of as a utility function or a preference function. Further we show that if $Z_i$ and $Z_j$ are random variables on the unit interval with distribution functions corresponding to the p-p plots for the $i$'th and $j$'th study, then there exists a function $L$ such that the treatment effect in the $i$'th study is preferred to the treatment effect in the $j$'th study if and only if

$$EL(Z_i) > EL(Z_j).$$

Finally using this result and some results about stochastic orderings we show that the rules for interpreting p-p plots presented in Chapter 1 are valid for any preference relation corresponding to a monotone increasing function $L$. The following definitions allow us to treat these ideas more formally. Definitions 3.6 and 3.7 and Lemma 3.8 come from Blackwell and Girshick (1954).

**Definition 3.6:** Let $(R, \beta)$ be a measurable space, and let $\mathcal{P}$ be the set of all probability distributions over $(R, \beta)$. Then a preference relation, $\geq$, on $\mathcal{P}$ is a binary relation such that

1. For every $p_1$ and $p_2$ in $\mathcal{P}$, either $p_2 \geq p_1$ or $p_1 \geq p_2$. (Both may hold.)
2. If $p_3 \geq p_2$ and $p_2 \geq p_1$, then $p_3 \geq p_1$.

The relation $p_2 \geq p_1$ can be interpreted as meaning $p_2$ is preferred or indifferent to $p_1$. If $p_2 \geq p_1$ but not $p_1 \geq p_2$, we say that $p_2$ is preferred to $p_1$ written $p_2 > p_1$. If $p_2 \geq p_1$ and $p_1 \geq p_2$, we say that $p_2$ is indifferent to $p_1$.

**Definition 3.7:** A utility function, $U$, corresponding to a preference relation, $\geq$, is a bounded real valued function defined on $\mathcal{P}$ such that

$$U\left(\sum_{n=1}^{\infty} \lambda_n p_n\right) = \sum_{i=1}^{\infty} \lambda_n U(p_n)$$

for any sequence $\lambda_n \geq 0$ with $\sum_1^{\infty} \lambda_n = 1$ and

$$U(p_2) \geq U(p_1) \quad \text{if and only if} \quad p_2 \geq p_1.$$

Lemma 3.8 presents conditions under which a preference relation has a utility function. A proof is given in Blackwell and Girshick (1954).
Lemma 3.8: Let \((R, \beta)\) be a measurable space and let \(\mathcal{P}\) be the set of all probability distributions on \((R, \beta)\). Then a preference relation, \(\succeq\), on \(\mathcal{P}\) has a utility function if it satisfies the conditions \(H_1\) and \(H_2\).

\(H_1\): If \(p_n \geq q_n\) for all \(n\), then 
\[ \sum \lambda_n p_n \geq \sum \lambda_n q_n \]
for any sequence \(\lambda_n \geq 0\) with \(\sum \lambda_n = 1\). If in addition \(p_n > q_n\) for some \(n\) for which \(\lambda_n > 0\), then
\[ \sum \lambda_n p_n > \sum \lambda_n q_n. \]

\(H_2\): If \(p_1 > p_2 > p_3\), then there are numbers \(\lambda, \mu\), with \(0 < \lambda, \mu < 1\) such that
\[ \lambda p_1 + (1 - \lambda)p_3 < p_2 \]
and
\[ \mu p_1 + (1 - \mu)p_3 > p_2. \]

Now we develop a representation of the preference relation in terms of expected values.

Definition 3.9: A subset \(A\) of \(R\) is a preference interval if \(z \in A\) whenever \(x, y \in A\), \(x \succeq z\) and \(z \succeq y\). In general \(c \succeq d\) means that \(p \succeq q\) where \(p([c]) = q([d]) = 1\).

Lemma 3.10: If in addition to \(H_1\) and \(H_2\) the following conditions, \(H_3\) and \(H_4\), are satisfied then there is a bounded real valued function \(u\) on \(R\) such that for all \(p, q \in \mathcal{P}\),
\[ p \succeq q \quad \text{iff} \quad \int u(x) \, dp(x) \geq \int u(x) \, dq(x). \]

\(H_3\) : \(\beta\) contains all preference intervals in \(R\).

\(H_4\) : If \(p, q \in \mathcal{P}\), \(A \in \beta\) and \(p(A) = 1\) then \(p \succeq q\) if \(c \succeq q\) for all \(c \in A\) and \(q \succeq p\) if \(q \succeq c\) for all \(c \in A\).

Proof: See Fishburn (1982). ||
These definitions and lemmas describe results from utility theory concerning preference relations on the collection of all distributions on \((R, \beta)\). We now extend these results slightly to preference relations on the collection of all pairs of distributions on \((R, \beta)\).

**Definition 3.11:** Let \((R, \beta)\) be a measurable space and let \(\mathcal{P}'\) be the set of all pairs of probability distributions over \((R, \beta)\). Then a preference relation, \(\succeq'\), on \(\mathcal{P}'\) is a binary relation such that

1. For every \((p_{11}, p_{21})\) and \((p_{12}, p_{22})\) in \(\mathcal{P}'\), either \((p_{12}, p_{22}) \succeq' (p_{11}, p_{21})\) or \((p_{11}, p_{21}) \succeq' (p_{12}, p_{22})\). (Both may hold.)
2. If \((p_{13}, p_{23}) \succeq' (p_{12}, p_{22})\) and \((p_{12}, p_{22}) \succeq' (p_{11}, p_{21})\), then \((p_{13}, p_{23}) \succeq' (p_{11}, p_{21})\).

**Definition 3.12** A utility function, \(U'\), corresponding to a preference relation, \(\succeq'\), defined on \(\mathcal{P}'\) is a bounded real valued function defined on \(\mathcal{P}'\) such that

\[
U'\left(\sum_{n=1}^{\infty} \lambda_n (p_{1n}, p_{2n})\right) = \sum_{n=1}^{\infty} \lambda_n U'(p_{1n}, p_{2n})
\]

for any sequence \(\lambda_n \geq 0\) with \(\sum_{n=1}^{\infty} \lambda_n = 1\) and

\[
U'(p_{11}, p_{21}) \geq U'(p_{12}, p_{22})
\]

if and only if

\[
(p_{11}, p_{21}) \succeq' (p_{12}, p_{22}).
\]

The next lemma extends Lemma 3.8 to preference relations on the collection, \(\mathcal{P}'\), of all pairs of distributions on \((R, \beta)\). This extension is needed to deal with a meta-analyst's preferences among pairs of treatment and control distributions.

**Lemma 3.13:** Let \((R, \beta)\) be a measurable space and let \(\mathcal{P}'\) be the collection of all pairs of probability distributions on \((R, \beta)\). Then a preference relation, \(\succeq'\), on \(\mathcal{P}'\) has a utility function, \(U'\), if it satisfies conditions \(H'_1\) and \(H'_2\).

\(H'_1:\) If \((p_{1n}, p_{2n}) \succeq' (q_{1n}, q_{2n})\) for all \(n\), then

\[
\sum \lambda_n (p_{1n}, p_{2n}) \succeq' \sum \lambda_n (q_{1n}, q_{2n})
\]
for any sequence \( \lambda_n \geq 0 \) with \( \sum \lambda_n = 1 \). If in addition \((p_{1n}, p_{2n}) >' (q_{1n}, q_{2n})\) for some \( n \) for which \( \lambda_n > 0 \) then

\[
\sum \lambda_n(p_{1n}, p_{2n}) >' \sum \lambda_n(q_{1n}, q_{2n}).
\]

\(H_2'\) : If

\[(p_{11}, p_{21}) >' (p_{12}, p_{22}) >' (p_{13}, p_{23})\]

then there are numbers \( \lambda, \mu \) with \( 0 < \lambda, \mu < 1 \) such that

\[
\lambda(p_{11}, p_{21}) + (1 - \lambda)(p_{13}, p_{23}) <' (p_{12}, p_{22})
\]

and

\[
\mu(p_{11}, p_{21}) + (1 - \mu)(p_{13}, p_{23}) >' (p_{12}, p_{22}).
\]

**Proof**: Let \( \geq' \) be a preference relation on \( \mathcal{P}' \) which satisfies \( H_1' \) and \( H_2' \). Define a preference relation, \( \geq \), on the collection, \( \mathcal{P} \), of all bivariate distributions on \( R \times R \) from \( \geq' \) as follows. If \( p_1 \) and \( p_2 \in \mathcal{P} \) and \( p^T_i \) and \( p^C_i \) denote the marginal distributions of \( p_i \) then,

\[ p_1 \geq p_2 \quad \text{if and only if} \quad (p^T_1, p^C_1) \geq' (p^T_2, p^C_2). \]

A check of the required conditions in Definition 3.6 shows that \( \geq \) is a preference relation. Furthermore, the relation \( \geq \) satisfies conditions \( H_1 \) and \( H_2 \) of Lemma 3.8. To see this let \( p_n \) and \( q_n \) be two sequences of distributions on \( R \times R \) such that

\[ p_n \geq q_n \quad \forall \quad 0 < n < \infty. \]

By the definition of \( \geq \) we have that

\[(p^T_n, p^C_n) \geq' (q^T_n, q^C_n) \quad \forall \quad 0 < n < \infty. \]

Since \( \geq' \) satisfies \( H_1' \) it follows that

\[
\sum \lambda_n(p^T_n, p^C_n) \geq' \sum \lambda_n(q^T_n, q^C_n) \tag{3.1}
\]
for any sequence of \( \lambda_n \) such that \( \sum \lambda_n = 1 \). (3.1) may be rewritten as
\[
\left[ \left( \sum \lambda_n p_n \right)^T, \left( \sum \lambda_n q_n \right)^T \right] \succeq \left[ \left( \sum \lambda_n p_n \right)^C, \left( \sum \lambda_n q_n \right)^C \right].
\] (3.2)

By the definition of \( \succeq \), (3.2) implies that
\[
\sum \lambda_n p_n \succeq \sum \lambda_n q_n.
\]

This shows that condition \( H_1 \) is satisfied.

Now consider condition \( H_2 \). Let \( p_1, p_2, p_3 \in \mathcal{P} \) be such that \( p_1 > p_2 > p_3 \). By definition of \( \succeq \) we have
\[
(p_1^T, p_1^C) \succeq (p_2^T, p_2^C) \succeq (p_3^T, p_3^C).
\]

By \( H'_2 \) there exists constants \( \lambda, \mu \) with \( 0 < \lambda, \mu < 1 \) such that
\[
\lambda(p_1^T, p_1^C) + (1 - \lambda)(p_3^T, p_3^C) \succeq (p_2^T, p_2^C) \tag{3.3}
\]
\[
\mu(p_1^T, p_1^C) + (1 - \mu)(p_3^T, p_3^C) \succ (p_2^T, p_2^C). \tag{3.4}
\]

(3.3) and (3.4) can be rewritten as
\[
\left[ (\lambda p_1 + (1 - \lambda)p_3)^T, (\lambda p_1 + (1 - \lambda)p_3)^C \right] \succeq (p_2^T, p_2^C)
\]
\[
\left[ (\mu p_1 + (1 - \mu)p_3)^T, (\mu p_1 + (1 - \mu)p_3)^C \right] \succ (p_2^T, p_2^C)
\]

which implies that
\[
\lambda p_1 + (1 - \lambda)p_3 < p_2
\]
\[
\mu p_1 + (1 - \mu)p_3 > p_2.
\]

This shows that condition \( H_2 \) is satisfied.

Thus by Lemma 3.8 there exists a utility function \( U \) corresponding to \( \succeq \). If we define
\[
U'(p_{11}, p_{12}) = U(p_{11} \times p_{12})
\]
where \( p_{11} \times p_{12} \) is the bivariate distribution constructed from \( p_{11} \) and \( p_{12} \) assuming independence then
\[
U'(p_{11}, p_{12}) \succeq U'(p_{21}, p_{22})
\]
if and only if

\[(p_{11}, p_{12}) \succeq (p_{21}, p_{22}).\]

Moreover

\[
\sum \lambda_n U'(p_{n1}, p_{n2}) = \sum \lambda_n U(p_{n1} \times p_{n2}) = U\left( \sum \lambda_n p_{n1} \times p_{n2} \right) = U'\left( \sum \lambda_n p_{n1}, \sum \lambda_n p_{n2} \right)
\]

\[= U'\left( \sum \lambda_n (p_{n1}, p_{n2}) \right)\]

Therefore \(U'\) so defined is a utility function on \(P'\) corresponding to \(\succeq'\).

**Definition 3.14:** A subset \(A\) of \(R \times R\) is a preference interval relative to \(\succeq'\) if \((z_1, z_2) \in A\) whenever \((x_1, x_2), (y_1, y_2) \in A\) \((x_1, x_2) \succeq' (z_1, z_2)\) and \((z_1, z_2) \succeq' (y_1, y_2)\). In general \((c_1, c_2) \succeq' (d_1, d_2)\) means \((p_{11}, p_{12}) \succeq' (p_{21}, p_{22})\) where \(p_{11}(c_1) = 1, p_{12}(c_2) = 1, p_{21}(d_1) = 1,\) and \(p_{22}(d_2) = 1\).

**Lemma 3.15:** If in addition to \(H'_1\) and \(H'_2\) the following conditions, \(H'_3\) and \(H'_4\), are satisfied then there is a bounded real valued function \(u(x_1, x_2)\) on \(R \times R\) such that for all \((p_1, p_2), (q_1, q_2) \in P'\),

\[ (p_1, p_2) \succeq' (q_1, q_2) \]

if and only if

\[
\int \int u(x_1, x_2) \, dp_1(x_1) \, dp_2(x_2) \geq \int \int u(x_1, x_2) \, dq_1(x_1) \, dq_2(x_2).
\]

\(H'_3\): The sigma-field in \((R, R)\) generated by \((\beta, \beta)\) contains all preference intervals in \((R, R)\) relative to \(\succeq'\).

\(H'_4\): If \((p_1, p_2), (q_1, q_2) \in P'\) and \(p_1(A) = p_2(A) = 1\) then \((p_1, p_2) \succeq' (q_1, q_2)\) if \((c_1, c_2) \succeq' (q_1, q_2)\) for all \((c_1, c_2) \in A\) and \((q_1, q_2) \succeq' (p_1, p_2)\) if \((q_1, q_2) \succeq' (c_1, c_2)\) for all \((c_1, c_2) \in A\).

**Proof:** It follows from \(H'_1, H'_2, H'_3\) and \(H'_4\) that \(\succeq'\) defined on \(P\) satisfies \(H_1, H_2, H_3\) and \(H_4\). Thus by Lemma 3.10 there exists a bounded real valued function, \(u(x_1, x_2)\), defined on \(R \times R\) which is such that \(p \succeq q\) if and only if

\[
\int u(x_1, x_2) \, dp(x_1, x_2) \geq \int u(x_1, x_2) \, dq(x_1, x_2).
\]
Chapter 3: Definitions and Preliminary Results

Now

\[(p_1, p_2) \succeq' (q_1, q_2)\]

if and only if

\[p_1 \times p_2 \geq q_1 \times q_2\]

if and only if

\[\int u(x_1, x_2) \, dp_1(x_1) \, dp_2(x_2) \geq \int u(x_1, x_2) \, dq_1(x_1) \, dq_2(x_2).\]

We now develop a representation of the preference relation, \(\succeq'\), in terms of the p-p plot. The development utilizes the fact that a meta-analyst is indifferent to transformations of the measurement scale.

**Theorem 3.16**: Let \((R, \beta)\) be a measurable space defined on the real line and let \(\succeq'\) be a preference relation on \(\mathcal{P}'\) that is invariant with respect to a group of transformations \(\mathcal{G}\) and satisfies \(H_1', H_2', H_3', H_4'\). Let \(\mathcal{F}\) be a family of pairs of distributions which is closed with respect to \(\mathcal{G}\) and suppose that given any two pairs of distribution functions, \((p_1(t), p_2(t))\) and \((q_1(t), q_2(t))\) \(\in \mathcal{F}\), there exists an element \(h \in \mathcal{G}\) such that \(F_{p_2}(t) = F_{q_2}(h(t))\) where \(F_p\) denotes the cumulative distribution function associated with \(p\). Then there exists a bounded real valued function \(L\) on \((0, 1)\) such that for all \((p_1, p_2), (q_1, q_2) \in \mathcal{F}\),

\[(p_1, p_2) \succeq' (q_1, q_2) \iff \int L(x) \, d(F_{p_1}F_{p_2}^{-1}(x)) \geq \int L(x) \, d(F_{q_1}F_{q_2}^{-1}(x))\]

**Proof**: Fix \(p_0\) such that \((p_0, p) \in \mathcal{F}\) for some \(p\). Since \(H_1', H_2', H_3', H_4'\) are all satisfied, Lemma 3.15 implies that there exists a bounded real valued function, \(L(x, y)\), on \(R \times R\) such that

\[(p_1, p_2) \succeq' (q_1, q_2)\]

if and only if

\[\int \int L(x, y) \, dF_{p_2}(x) \, dF_{p_1}(y) \geq \int \int L(x, y) \, dF_{q_2}(x) \, dF_{q_1}(y).\]

Denote

\[\int \int L(x, y) \, dF_{p_2}(x) \, dF_{p_1}(y)\]
by $U_L(F_{p_1}, F_{p_2})$. Since $\succeq'$ is invariant with respect to transformations in $G$,

$$U_L(F_{p_1}, F_{p_2}) = \int \int L(x, y) dF_{p_2}(g(x)) dF_{p_1}(g(y)),$$

for all $g \in G$. If we let $g$ be such that $F_{p_2}(g(x)) = F_{p_0}(x)$ then

$$U_L(F_{p_1}, F_{p_2}) = \int \int L(x, y) dF_{p_0}(x) dF_{p_1}(g(y)) = \int L_1(y) dF_{p_1}(g(y)),$$

where

$$L_1(y) = \int L(x, y) dF_{p_0}(x).$$

By making the substitution $y = F_{p_0}^{-1}(s)$ and noting that $g F_{p_0}^{-1}(s) = F_{p_2}^{-1}(s)$ we obtain

$$U_L(F_{p_1}, F_{p_2}) = \int L_1(F_{p_0}^{-1}(s)) dF_{p_1}(g(F_{p_0}^{-1}(s)))$$

$$= \int L_2(s) dF_{p_1}(F_{p_2}^{-1}(s)).$$

That is there is a function, $L_2(s) = L_1(F_{p_0}^{-1}(s))$, corresponding to the preference relation, $\succeq'$, which is such that

$$(p_1, p_2) \succeq' (q_1, q_2)$$

if and only if

$$\int L_2(s) dF_{p_1}(F_{p_2}^{-1}(s)) \geq \int L_2(s) dF_{q_1}(F_{q_2}^{-1}(s)).$$

Since $L(x, y)$ is bounded and real valued so is $L_2(s)$. $||$

We have just shown that a meta-analyst’s preferences can be represented in terms of the expected value of a function with respect to the distribution whose c.d.f. is the p-p plot. Now we show that the rules for interpreting p-p plots presented in Chapter 1 are valid for any preference relation corresponding to a monotone increasing function $L$. This involves applying the first order stochastic dominance criteria to the representation of the preference relation in terms of the p-p plot. For an introduction to stochastic dominance see Whitmore and Findlay (1978). Under the first order stochastic dominance criteria the distribution function $F(\cdot)$ is preferred to the distribution function $G(\cdot)$ if and only if $F(x)$ is less than or equal to $G(x)$ for all $x$. An interesting property of this ordering is that if
\( F \) is preferred to \( G \) then for any monotone increasing function \( L(x) \),

\[
E_F(L(x)) \geq E_G(L(x)).
\]

For distribution functions on the unit interval such as the p-p plot this follows from an integration by parts argument.

\[
\int_0^1 L(x) \, dF(x) = F(1)L(1) - F(0)L(0) - \int_0^1 F(x) \, dL(x)
\]

\[
= L(1) - \int_0^1 F(x) \, dL(x),
\]

since \( F(0) = 0 \). So since \( G(x) \geq F(x) \) for all \( x \) and \( G(0) = 0 \) we have

\[
\int_0^1 L(x) \, dF(x) \geq L(1) - \int_0^1 G(x) \, dL(x)
\]

\[
= \int_0^1 L(x) \, dG(x).
\]

To utilize the stochastic dominance criteria to interpret p-p plots note that we have described the meta-analysts preferences as a comparison of the quantity

\[
\int L_2(p) \, dF_T(F_C^{-1}(p))
\]

across the studies under consideration. The larger this integral is the more desirable the corresponding treatment effect. Using the stochastic dominance criterion just described, it follows that if p-p plot 1 lies below p-p plot 2 then the treatment in study 1 is preferred to the treatment in study 2 for all preference relations which correspond to a monotone increasing function \( L_2 \). Many different preference relations correspond to such a function \( L_2 \), so this is a strong statement about the desirability of the treatment in study 1 compared to the treatment in study 2.

Further, if p-p plot 1 lies below p-p plot 2 in the interval \([0, a]\) and above p-p plot 2 in the interval \([a, 1]\), then we can dichotomize the description of the treatment effect by employing two classes of utility functions, \( U_1 \), which consists of monotone increasing functions that are constant over the interval \([a, 1]\), and \( U_2 \), which consists of monotone increasing functions that are constant over the interval \([0, a]\). Using an integration by
Section 3.3: A theoretical basis for interpreting p-p plots

parts argument it can be shown that the treatment in study 1 is preferred to the treatment in study 2 for all preference relations in $U_1$ and that the treatment in study 2 is preferred to the treatment in study 1 for all preference relations in $U_2$. $U_1$ corresponds to all preference relations concerned only about the outcomes in the interval $[0, a]$ while $U_2$ corresponds to all preference relations concerned only about outcomes in the interval $[a, 0]$.

The concept of stochastic dominance can also be employed to interpret a single p-p plot. To see this note that if a p-p plot lies below the 45 degree line, then

$$F_T(F_C^{-1}(p)) \leq p \quad \forall \ 0 < p < 1.$$  

Substituting $F_C(x)$ for $p$ we obtain

$$F_T(x) \leq F_C(x) \quad \forall \ x,$$

which says that $F_T$ stochastically dominates $F_C$. Thus the treatment is preferred to the control for any preference relation on the set of all possible distributions that can be expressed as the expected value of a monotone increasing function.
Figure 3.1

P-p plots of $F_T$ versus $F_C$ for Gaussian distributions with $\sigma_T = \sigma_C$ and $\delta = (\mu_T - \mu_C)/\sigma = -2, -1, 0, 1, 2$. 
Section 3.3: A theoretical basis for interpreting p-p plots  49

Figure 3.2

P-p plots of $F_T$ versus $F_C$ for Gaussian distributions with $\tau = \sigma_C/\sigma_T = 2, .5$ and $\delta = (\mu_T - \mu_C)/\sigma_C = -2, -.5, 0, .5, 2$. 

\[
\begin{align*}
\tau &= .5, \quad \delta = -.5 \\
\tau &= 2, \quad \delta = 2 \\
\tau &= 2, \quad \delta = 0 \\
\tau &= .5, \quad \delta = 0 \\
\tau &= .5, \quad \delta = .5
\end{align*}
\]
Chapter 4

The P-p Plot for Gaussian Distributions

In the previous chapter it is shown that the p-p plot corresponding to a treatment and a control distribution is a complete description of the treatment effect that is comparable across studies. Thus given the data associated with an individual study,

\[ X_1^T, \ldots, X_{n_T}^T \quad \text{and} \quad X_1^C, \ldots, X_{n_C}^C, \]

it is useful to construct estimates of and confidence bands for the p-p plot so that the treatment effect can be compared with treatment effects in other similar studies. In this chapter we focus on these statistical problems for the model where \( X^T \) and \( X^C \) are normally distributed random variables with means \( \mu_T \) and \( \mu_C \) and variances \( \sigma_T^2 \) and \( \sigma_C^2 \), respectively. First the case where \( \sigma_T \) and \( \sigma_C \) are assumed equal is investigated. This case corresponds to the standard effect size model. Then the case where the variances for the treatment and control groups may be different is considered. In both of these cases, unbiased and maximum likelihood estimation of the p-p plot as well as the construction of confidence bands are explored. First we present some properties of the p-p plot for Gaussian distributions.

4.1. Properties of the p-p plot.

The p-p plot is symmetric and concave or convex in the case where the treatment and the control distributions are Gaussian with a common variance. This is revealed by the next two lemmas. The first lemma states that in the standard effect size model the p-p plot is symmetric about the line \( 1 - p \) while the second lemma states that the p-p plot is either convex or concave depending on whether the effect size is positive or negative.
In the case where the treatment and control variances are not equal the p-p plot is not necessarily symmetric, convex or concave as we demonstrate by some examples.

To obtain these results we first need to determine the functional form of the p-p plot in the Gaussian model. As has already been shown the p-p plot can be represented as the function \( F_T \left( F_C^{-1}(p) \right) \) on the interval \((0, 1)\). Upon substituting the Gaussian distributions the functional form becomes

\[
F_T(F_C^{-1}(p)) = \Phi \left( \frac{\mu_C + \sigma_C \Phi^{-1}(p) - \mu_T}{\sigma_T} \right) \\
= \Phi \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right),
\]

for \(0 < p < 1\) where \(k_p \equiv \Phi^{-1}(p)\). When the treatment and control variances are equal to a common value, \(\sigma^2\), (4.1) simplifies to

\[
F_T(F_C^{-1}(p)) = \Phi(k_p - \delta),
\]

where \(\delta = (\mu_T - \mu_C)/\sigma\) is the effect size. Now we are ready to prove the lemmas.

**Lemma 4.1** The p-p plot in the standard effect size model, \(\Phi(k_p - \delta)\), is symmetric about the line \(1 - p\).

**Proof:** The p-p plot is symmetric about the line \(1 - p\) if and only if the mirror image of every point on the p-p plot with respect to the line \(1 - p\) lies on the p-p plot. Now the mirror image of the point \(a = (a_1, a_2)\) with respect to the line \(1 - p\) is the point \(b = (1 - a_2, 1 - a_1)\). Thus, to prove the lemma it is enough to show that for all \(0 < p < 1\),

\[
(1 - \Phi(k_p - \delta), 1 - p) = (1 - \Phi(k_p - \delta), \Phi(k_1 - \Phi(k_p - \delta) - \delta)),
\]

or equivalently,

\[
\Phi(k_1 - \Phi(k_p - \delta) - \delta) = 1 - p.
\]
Chapter 4: The P-p Plot for Gaussian Distributions  52

Since \(1 - \Phi(x) = \Phi(-x)\) and \(\Phi(k_x) = k_\Phi(x) = x\)

\[
\Phi \left( k_{1-\Phi(k_{p-\delta})} - \delta \right) = \Phi \left( k_{\Phi(-k_p+\delta)} - \delta \right)
\]

\[
= \Phi (-k_p + \delta - \delta)
\]

\[
= 1 - \Phi (k_p)
\]

\[
= 1 - p
\]

**Lemma 4.2** The p-p plot in the standard effect size model, \(\Phi(k_p - \delta)\), is convex for \(\delta > 0\) and concave for \(\delta < 0\).

**Proof:** To prove this result it suffices to compute the second derivative of \(\Phi(k_p - \delta)\) and show that it is positive when \(\delta > 0\) and negative when \(\delta < 0\). Let \(\phi(x)\) denote the density for a standard normal distribution. Then

\[
\frac{d}{dp} \Phi(k_p - \delta) = \phi(k_p - \delta) \frac{dk_p}{dp},
\]

\[
\frac{d^2}{dp^2} \Phi(k_p - \delta) = \frac{d\phi(k_p - \delta)}{dk_p} \left( \frac{dk_p}{dp} \right)^2 + \phi(k_p - \delta) \frac{d^2 k_p}{dp^2}. \tag{4.2}
\]

Also

\[
\frac{dk_p}{dp} = \left( \frac{d\Phi(x)}{dx} \bigg|_{x=k_p} \right)^{-1} = \frac{1}{\phi(k_p)}, \tag{4.3}
\]

\[
\frac{d^2 k_p}{dp^2} = - \frac{1}{[\phi(k_p)]^2} \frac{d\phi(k_p)}{dk_p} \frac{dk_p}{dp} = - \frac{1}{[\phi(k_p)]^2} \frac{d\phi(k_p)}{dk_p}. \tag{4.4}
\]

Substitution of (4.3) and (4.4) in (4.2) yields

\[
\frac{d^2}{dp^2} \Phi(k_p - \delta) = \frac{d\phi(k_p - \delta)}{dk_p} \frac{1}{\phi(k_p)^2} - \phi(k_p - \delta) \frac{1}{[\phi(k_p)]^3} \frac{d\phi(k_p)}{dk_p}. \tag{4.5}
\]

Since \(\frac{d}{dx} \phi(x) = -x\phi(x)\), (4.5) becomes

\[
\frac{d^2}{dp^2} \Phi(k_p - \delta) = -(k_p - \delta) \phi(k_p - \delta) \frac{1}{\phi(k_p)^2} + \phi(k_p - \delta) \frac{1}{\phi(k_p)^3} k_p \phi(k_p)
\]

\[
= \frac{\delta \phi(k_p - \delta)}{\phi(k_p)^2}
\]

which has the sign of \(\delta\), and completes the proof. ||
Section 4.2: Estimation of the p-p plot: homogeneous variances

Lemma 4.2 implies that the p-p plot is less than p for \( \delta \) greater than zero and greater than p for \( \delta \) less than zero. This follows from the fact that p represents the line that connects the points \((0,0)\) and \((1,1)\) on the p-p plot and any line segment connecting two points on a convex curve lies above the curve. When \( \delta \) equals zero the p-p plot is the line segment connecting \((0,0)\) and \((1,1)\).

The curves in Figure 4.1 are graphs of the p-p plot when the treatment and control variances are not equal. These graphs are not convex, concave or symmetric with respect to the line \(1 - p\). This shows that when the variances are not equal in the Gaussian model the results in Lemma 4.1 and 4.2 do not hold.

4.2. Estimation of the p-p plot: homogeneous variances.

In this section we consider estimating the p-p plot in the Gaussian model when \(\sigma_T\) and \(\sigma_C\) are equal. This model is appropriate when the treatment doesn’t affect the variability of the observations. We present the maximum likelihood estimate of the p-p plot, compute its asymptotic variance and obtain an approximate simultaneous confidence band. Also we consider the question of the existence of an unbiased estimate of the p-p plot.

4.2.1. Maximum likelihood estimation of the p-p plot

Suppose that \(X^T_1, \ldots, X^T_{n_T}\) and \(X^C_1, \ldots, X^C_{n_C}\) are random samples from \(\mathcal{N}(\mu_T, \sigma^2)\) and \(\mathcal{N}(\mu_C, \sigma^2)\) distributions respectively, and that \(\mu_T, \mu_C\) and \(\sigma^2\) are all unknown. Let \(\hat{\mu}\) and \(\hat{\sigma}^2\) be the maximum likelihood estimates of \(\mu_T - \mu_C\) and \(\sigma^2\). That is

\[
\hat{\mu} = \bar{X}^T - \bar{X}^C
\]

\[
\hat{\sigma}^2 = \frac{1}{n_T + n_C} \left[ \sum_{i=1}^{n_T} (X^T_i - \bar{X}^T)^2 + \sum_{i=1}^{n_C} (X^C_i - \bar{X}^C)^2 \right].
\]

The maximum likelihood estimate of the p-p plot is obtained immediately.

**Lemma 4.3** The maximum likelihood estimate of the p-p plot is

\[
\Phi \left( k_p - \frac{\hat{\mu}}{\hat{\sigma}} \right).
\]
Chapter 4: The P-p Plot for Gaussian Distributions  54

To gain an idea of how accurate an estimate of the p-p plot is, it is necessary to have some notion of its variability. In Lemma 4.4 we present a formula for the asymptotic distribution of any estimate of the p-p plot based on estimates of $\mu$ and $\sigma$. We immediately utilize this lemma in Theorem 4.5 to compute the asymptotic distribution of the maximum likelihood estimate of the p-p plot presented in Lemma 4.3.

**Lemma 4.4:** Suppose that $\hat{\mu}$ and $\hat{\sigma}^2$ are estimates of $\mu$ and $\sigma^2$ satisfying

$$\sqrt{n_T + n_C} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma}^2 - \sigma^2 \end{pmatrix} \rightarrow_L \mathcal{N}(0, \Sigma)$$

as $n_T \to \infty$ and $n_C \to \infty$ such that $n_T/(n_T + n_C) \to r$. Then

$$\sqrt{n_T + n_C} \left[ \Phi(k_p - \hat{\mu}/\hat{\sigma}) - \Phi(k_p - \mu/\sigma) \right] \rightarrow_L \mathcal{N}(0, (a,b) \Sigma (a,b)'),$$

where

$$a = -\frac{1}{\sigma} \phi(k_p - \mu/\sigma), \quad b = \frac{\mu}{2\sigma^3} \phi(k_p - \mu/\sigma).$$

**Proof:** The first partial derivatives of the p-p plot with respect to $\mu$ and $\sigma^2$ are $a$ and $b$ respectively. Since $a$ and $b$ are continuous the p-p plot is totally differentiable and we have

$$\Phi(k_p - \hat{\mu}/\hat{\sigma}) - \Phi(k_p - \mu/\sigma) = (\hat{\mu} - \mu) a + \left(\hat{\sigma}^2 - \sigma^2\right) b + o_p(1/\sqrt{n_T + n_C}).$$

The theorem now follows from standard results on asymptotic distributions. See for example Rao (1973). ||

Now we compute the asymptotic distribution of $\Phi(k_p - \hat{\mu}/\hat{\sigma})$.

**Theorem 4.5:** Suppose that $\hat{\mu}$ and $\hat{\sigma}$ are as in Lemma 4.3 and that $n_T \to \infty$, $n_C \to \infty$ such that $n_T/(n_T + n_C) \to r$. Then

$$\sqrt{n_T + n_C} \left[ \Phi(k_p - \hat{\mu}/\hat{\sigma}) - \Phi(k_p - \mu/\sigma) \right] \rightarrow_L \mathcal{N} \left[ 0, (1/r + 1/(1 - r) + \delta^2/2) \phi^2(k_p - \delta) \right].$$

**Proof:** The proof follows by a direct application of Lemma 4.4 with

$$\Sigma = \begin{pmatrix} \frac{1}{r} & \frac{1}{1-r} \\ \frac{1}{1-r} & 2\sigma^4 \end{pmatrix}.$$
4.2.2. Simultaneous confidence bands for the p-p plot

Since it is of interest to compare a p-p plot from a single study with the 45 degree line and to p-p plots from other studies a simultaneous confidence band for the p-p plot over the whole range of control percentiles can be useful. Here we present an approximate simultaneous confidence band for the p-p plot in the Gaussian model when the treatment and control variances are equal.

Suppose that $X_{1T}^T, \ldots, X_{n_T}^T$ and $X_{1C}^C, \ldots, X_{n_C}^C$ are random samples from normal populations $\mathcal{N}(\mu_T, \sigma^2)$ and $\mathcal{N}(\mu_C, \sigma^2)$, respectively, where the parameters $\mu_T, \mu_C$ and $\sigma$ are unknown. Let $\hat{\delta} = \hat{\mu}/\hat{\sigma}, z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, and

$$v = \frac{1}{n_T} + \frac{1}{n_C} + \frac{\hat{\delta}^2}{2(n_T + n_C - 2)}.$$

Then we may state the following lemma.

**Lemma 4.6** The band

$$\left[ \Phi \left( k_p - \hat{\delta} - z_{\alpha/2} \sqrt{v} \right), \Phi \left( k_p - \hat{\delta} + z_{\alpha/2} \sqrt{v} \right) \right]$$

is an approximate 1-\(\alpha\) percent simultaneous confidence band for $\Phi \left( k_p - \delta \right)$ for $0 < p < 1$.

**Proof:** First note that $\hat{\delta}$ is approximately normally distributed with mean

$$\delta = \frac{(\mu_T - \mu_C)}{\sigma},$$

and variance

$$\frac{1}{n_T} + \frac{1}{n_C} + \frac{\hat{\delta}^2}{2(n_T + n_C - 2)}.$$

Since $\hat{\delta}$ converges in probability to $\delta$, we can substitute $\hat{\delta}$ for $\delta$ in the expression for the asymptotic variance of $\hat{\delta}$. Thus

$$P \left( \hat{\delta} - z_{\alpha/2} \sqrt{v} < \delta < \hat{\delta} + z_{\alpha/2} \sqrt{v} \right) = 1 - \alpha,$$

and hence

$$P \left[ \Phi \left( k_p - \hat{\delta} - z_{\alpha/2} \sqrt{v} \right) < \Phi \left( k_p - \delta \right) < \Phi \left( k_p - \hat{\delta} + z_{\alpha/2} \sqrt{v} \right) \right] = 1 - \alpha.$$
It is readily seen that in the Gaussian model the boundaries of this simultaneous confidence band correspond to two distinct p-p plots. Note also that this confidence band is a $1 - \alpha$ percent confidence interval for $\Phi(k_p - \delta)$ for any specific $p$ between zero and one.

4.2.3. Unbiased estimation of the p-p plot

In this section we show that a bounded unbiased estimate of $\Phi(k_p - \delta)$ doesn’t exist. Suppose there exists an unbiased estimate of $\Phi(k_p - \delta)$ in the case where $\mu_T$, $\mu_C$, and $\sigma^2$ are all unknown. Then there also exists an unbiased estimate of $\Phi(k_p - \delta)$ when just $\sigma^2$ is unknown since we can pretend not to know $\mu_T$ and $\mu_C$. Thus if we can show that there doesn’t exist an unbiased estimate of $\Phi(k_p - \delta)$ when just $\sigma^2$ is unknown it follows that there doesn’t exist an unbiased estimate when $\mu_T$, $\mu_C$, and $\sigma^2$ are all unknown.

So suppose that just $\sigma^2$ is unknown and let $X^T_1, \ldots, X^T_{n_T}$ and $X^C_1, \ldots, X^C_{n_C}$ be random samples from $\mathcal{N}(\mu_T, \sigma^2)$ and $\mathcal{N}(\mu_C, \sigma^2)$ distributions respectively. Since the statistic

$$v = \frac{1}{n_T + n_C} \left[ \sum_{i=1}^{n_T} (X^T_i - \mu_T)^2 + \sum_{i=1}^{n_C} (X^C_i - \mu_C)^2 \right]$$

is complete and sufficient we restrict our attention to estimates of the p-p plot that are functions of $v$. We further restrict our search to estimates that are bounded. Let $l(k_p, v)$ be an unbiased estimate of $\Phi(k_p - \delta)$ that is bounded and let $n = n_T + n_C$. Since $l(k_p, v)$ is bounded there exists a constant $C$ such that $l(k_p, v) + C$ is a bounded nonnegative and unbiased estimate of $\Phi(k_p - \delta) + C$. Unbiasedness implies

$$\int_0^\infty [l(k_p, v) + C] \frac{1}{\Gamma(n/2) 2^{n/2}} v^{n/2-1} \exp \left(-\frac{v}{2\sigma^2}\right) dv = (\sigma^2)^{n/2} [\Phi(k_p - \delta) + C].$$

The substitution $t = \frac{v}{2\sigma^2}$ produces the following equation.

$$\int_0^\infty [l(k_p, v) + C] \frac{1}{\Gamma(n/2) 2^{n/2}} v^{n/2-1} \exp \left(-tv\right) dv$$

$$= \frac{1}{(2t)^{n/2}} \Phi \left(k_p - (\mu_T - \mu_C) \sqrt{2t}\right) + \frac{C}{(2t)^{n/2}}.$$

Now since
Section 4.2: Estimation of the p-p plot: homogeneous variances

\[ \frac{d}{dt} \left[ (l(k_p, v) + C) v^{n/2-1} e^{-tv} \right] = \left[ (l(k_p, v) + C) v^{n/2} e^{-tv} \right], \]

is bounded by the integrable function

\[ \left[ (l(k_p, v) + C) v^{n/2} e^{-(t_0-\epsilon)v} \right], \]

for all \( t \) in \( (t_0 - \epsilon, t_0 + \epsilon) \), we have that

\[ \frac{d}{dt} \int_0^\infty \left[ (l(k_p, v) + C) v^{n/2-1} e^{-tv} \right] dv = -\int_0^\infty \left[ (l(k_p, v) + C) v^{n/2} e^{-tv} \right] dv, \]

for all \( t \in (0, \infty) \), by the dominated convergence theorem. Thus the existence of an unbiased estimate of the p-p plot implies that

\[ \frac{d}{dt} \left[ \frac{1}{(2t)^{n/2}} \Phi \left( k_p - (\mu_T - \mu_C) \sqrt{2t} \right) + \frac{C}{(2t)^{n/2}} \right] < 0, \quad (4.6) \]

for all \( t \in (0, \infty) \) and for all \( p \in (0, 1) \). The derivative in (4.6) is equal to

\[ \frac{1}{(2t)^{n+1/2}} \left[ \frac{-n}{(2t)^{1/2}} \Phi(k_p - (\mu_T - \mu_C) \sqrt{2t} + C) - (\mu_T - \mu_C) \phi(k_p - (\mu_T - \mu_C) \sqrt{2t}) \right]. \quad (4.7) \]

When \( \mu_T - \mu_C \) is less than zero, (4.7) may be positive. To see this fix \( \mu_T - \mu_C < 0 \) and choose \( t \) so that

\[ \frac{n(1 + C)}{(2t)^{1/2}} < -\mu_T - \mu_C \max_x \phi(x). \]

Now choose \( p \) so that

\[ -(\mu_T - \mu_C) \phi(k_p - (\mu_T - \mu_C) \sqrt{2t}) = -(\mu_T - \mu_C) \max_x \phi(x). \]

For this \( t \) and \( p \) the derivative is positive. Thus we have obtained a contradiction and we can assert that there does not exist an unbiased estimate of the p-p plot that is bounded.

Further we can assert that there does not exist an unbiased estimate of the p-p plot when the treatment and control variances may be unequal since whenever there exists an unbiased estimate of the p-p plot in this situation there exists an unbiased estimate in the setting where the treatment and control variances are identical.

These results concerning the existence of unbiased estimates eliminate the possibility of averaging unbiased estimates of p-p plots to obtain greater accuracy in the Gaussian
model. A different approach is needed to obtain more accurate estimates of the p-p plot from the results of a number of studies.


In this section estimation of the p-p plot in the model where $\sigma_T$ and $\sigma_C$ may be unequal is considered. This model is appropriate when the treatment not only affects the average value of an observation but also its variability. We present the maximum likelihood estimate of the p-p plot and compute its asymptotic distribution. In addition confidence intervals for and unbiased estimation of the p-p plot are discussed.

4.3.1. Maximum likelihood estimation of the p-p plot

Suppose that

$$X^T_1, \ldots, X^T_{n_T} \quad \text{and} \quad X^C_1, \ldots, X^C_{n_C}$$

are random samples from normal populations $\mathcal{N}(\mu_T, \sigma^2_T)$ and $\mathcal{N}(\mu_C, \sigma^2_C)$, respectively, and that the parameters $\mu_T, \mu_C, \sigma_T$ and $\sigma_C$ are unknown. Since the MLE's of these parameters are

$$\hat{\mu}_T = \bar{X}^T$$

$$\hat{\mu}_C = \bar{X}^C$$

$$\hat{\sigma}_T^2 = \frac{1}{n_T} \sum_{i=1}^{n_T} (X^T_i - \hat{\mu}_T)^2$$

$$\hat{\sigma}_C^2 = \frac{1}{n_C} \sum_{i=1}^{n_C} (X^C_i - \hat{\mu}_C)^2.$$ 

we have the following lemma.

**Lemma 4.7** The maximum likelihood estimate of the p-p plot is

$$\Phi \left( k_p \frac{\hat{\sigma}_C}{\hat{\sigma}_T} - \frac{\hat{\mu}_T - \hat{\mu}_C}{\hat{\sigma}_T} \right).$$

In Lemma 4.8 we show how to compute the asymptotic distribution of any estimate of the p-p plot that is based on estimates of $\mu_T, \mu_C, \sigma_T, \sigma_C$. This result allows us to deduce the accuracy of various estimates of the p-p plot. We use Lemma 4.8 in Theorem 4.9 to compute the asymptotic distribution of the maximum likelihood estimate of the p-p
plot presented in Lemma 4.7. We also use Lemma 4.8 in later chapters when we consider estimates of the p-p plot based on pooled data from a number of different studies.

**Lemma 4.8:** Suppose that $\hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}_T$ and $\hat{\sigma}_C$ are estimates of $\mu_T, \mu_C, \sigma_T$, and $\sigma_C$ satisfying

$$
\sqrt{n_T + n_C} \begin{pmatrix}
\hat{\mu}_T - \mu_T \\
\hat{\mu}_C - \mu_C \\
\hat{\sigma}_T^2 - \sigma_T^2 \\
\hat{\sigma}_C^2 - \sigma_C^2
\end{pmatrix} \to_L \mathcal{N}(0, \Sigma)
$$

as $n_T \to \infty$ and $n_C \to \infty$ such that $n_T/(n_T + n_C) \to r$.

Then

$$
\sqrt{n_T + n_C} \left[ \Phi \left( k_p \frac{\hat{\sigma}_C}{\sigma_T} - \frac{\hat{\mu}_T - \hat{\mu}_C}{\sigma_T} \right) - \Phi \left( k_p \frac{\hat{\sigma}_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right) \right] \\
\to_L \mathcal{N} \left( 0, \phi^2 \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right) (a, b, c, d) \Sigma (a, b, c, d)' \right),
$$

where

$$
a = b = \frac{1}{\sigma_T} \\
c = -\frac{k_p \sigma_C - \mu_T + \mu_C}{2 \sigma_T^3} \\
d = \frac{k_p}{2 \sigma_T \sigma_C}.
$$

**Proof:** The first partial derivatives of the p-p plot with respect to $\mu_T, \mu_C, \sigma_T^2$ and $\sigma_C^2$ are $\phi(k_p \sigma_C - \mu_T - \mu_C)$ multiplied by $a, b, c,$ and $d$ respectively. Since $a, b, c$ and $d$ are continuous the first partial derivatives of the p-p plot are continuous which implies that the p-p plot is totally differentiable. Thus

$$
\Phi \left( k_p \frac{\hat{\sigma}_C}{\sigma_T} - \frac{\hat{\mu}_T - \hat{\mu}_C}{\sigma_T} \right) - \Phi \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right) \\
= [(\hat{\mu}_T - \mu_T) a + (\hat{\mu}_C - \mu_C) b + (\hat{\sigma}_T^2 - \sigma_T^2) c + (\hat{\sigma}_C^2 - \sigma_C^2) d] \phi \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma} \right) \\
+ o_p(1/\sqrt{n_T + n_C}).
$$

The theorem now follows from standard results on asymptotic distributions. ||
Theorem 4.9: Suppose that \( \hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}_T^2 \) and \( \hat{\sigma}_C^2 \) are as in Lemma 4.7, and that \( n_T \to \infty, n_C \to \infty \) such that \( n_T/(n_T + n_C) \to r \). Then

\[
\sqrt{n_T + n_C} \left[ \Phi \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\hat{\mu}_T - \hat{\mu}_C}{\sigma_T} \right) - \Phi \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right) \right] 
\to L N \left( 0, \phi^2 \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right) \left[ a^2 \frac{\sigma_T^2}{r} + b^2 \frac{\sigma_C^2}{1-r} + c^2 \frac{2\sigma_T^2}{r} + d^2 \frac{2\sigma_C^2}{1-r} \right] \right)
\]

Proof: The result follows by a direct application of Lemma 4.8 with

\[
\Sigma = \begin{pmatrix}
\frac{\sigma_T^2}{r} & 0 & 0 & 0 \\
0 & \frac{\sigma_C^2}{1-r} & 0 & 0 \\
0 & 0 & 2\frac{\sigma_T^2}{r} & 0 \\
0 & 0 & 0 & 2\frac{\sigma_C^2}{1-r}
\end{pmatrix}
\]

4.3.2. Simultaneous confidence bands for the p-p plot

In Section 4.3 we present the asymptotic distribution of the maximum likelihood estimate of the p-p plot at a particular value of \( p \). This result enables us to compute approximate confidence intervals for the p-p plot at any specific percentile of the control distribution. The confidence interval takes the form

\[
\left[ \hat{F}_T(\hat{F}_C^{-1}(p)) - Z_{\alpha/2}\hat{\sigma}(p), \hat{F}_T(\hat{F}_C^{-1}(p)) + Z_{\alpha/2}\hat{\sigma}(p) \right],
\]

where \( \hat{F}_T(\hat{F}_C^{-1}(p)) \) and \( \hat{\sigma}(p) \) are respectively the maximum likelihood estimates of the p-p plot and the standard deviation of the p-p plot at \( p \), and \( Z_{\alpha/2} \) is the \( 1 - \alpha/2 \) quantile of the standard normal distribution. In meta-analysis the relation of a p-p plot to the 45 degree line and to other p-p plots is of interest which suggests the need for a simultaneous confidence band for the p-p plot. In this section we determine the constant \( C_{\alpha/2} \) to substitute for \( Z_{\alpha/2} \) which makes the collection of confidence intervals for all \( 0 < p < 1 \) a \( 1 - \alpha \) percent simultaneous confidence band for the p-p plot. First we present some preliminary results.
Definition 4.10: The collection of random variables \((Z(t), 0 < t < 1)\) is a finite dimensional Gaussian random function if it can be expressed as

\[
\sum_{i=1}^{k} u_i(t)X_i,
\]

where the \(X_i\) are independent Gaussian random variables with zero means and variances equal to 1 and the \(u_i\) are real valued differentiable functions.

The following lemma shows that asymptotically the maximum likelihood estimate of the p-p plot behaves as a finite dimensional Gaussian random function.

Lemma 4.11: Suppose that \(\hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}_T^2\) and \(\hat{\sigma}_C^2\) are estimates of \(\mu_T, \mu_C, \sigma_T^2,\) and \(\sigma_C^2\) satisfying

\[
\sqrt{n_T + n_C} \begin{pmatrix} \hat{\mu}_T - \mu_T \\ \hat{\mu}_C - \mu_C \\ \hat{\sigma}_T^2 - \sigma_T^2 \\ \hat{\sigma}_C^2 - \sigma_C^2 \end{pmatrix} \rightarrow_L \mathcal{N}(0, \Sigma)
\]

as \(n_T \to \infty\) and \(n_C \to \infty\) such that \(n_T/(n_T + n_C) \to r\).

Then

\[
\sqrt{n_T + n_C} \left[ \Phi \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\hat{\mu}_T - \hat{\mu}_C}{\sigma_T} \right) - \Phi \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right) \right] \quad (4.8)
\]

converges in distribution to a finite dimensional Gaussian random function with covariance function given by

\[
\phi \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right) \phi \left( k_p \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right) [a, b, c(p_1), d(p_1)] \sum [a, b, c(p_2), d(p_2)]',
\]

where

\[
-a = b = \frac{1}{\sigma_T},
\]

\[
c(p) = -\frac{k_p\sigma_C - \mu_T + \mu_C}{2\sigma_T^3}
\]

\[
d(p) = \frac{k_p}{2\sigma_T\sigma_C}.
\]

Proof: The finite dimensional distributions of (4.8) converge to those of a finite dimensional Gaussian random function with (4.9) as a covariance function. Thus to show
that (4.8) converges in distribution to a finite dimensional random function it needs to be shown that the sequence of distributions corresponding to (4.8) is tight. A sequence of probability measures, \( P_n \), on the space of continuous functions on the unit interval, \( C[0,1] \), is tight if and only if (Billingsley 1968)

1. For each positive \( \eta \), there exists an \( a \) such that

\[
P_n(x \in C[0,1] : |x(0)| > a) \leq \eta, \quad n \geq 1,
\]

2. For each positive \( \epsilon \) and \( \eta \), there exists a \( \delta \) with \( 0 < \delta < 1 \), and an \( n_0 \) such that

\[
P_n(x \in C[0,1] : w_x(\delta) \geq \epsilon) \leq \eta, \quad n \geq n_0,
\]

where \( w_x(\delta) = \sup_{|s-t|<\delta}|x(s) - x(t)| \).

Condition 1 is satisfied since

\[
P_n(x \in C[0,1] : |x(0)| = 0) = 1 \quad \forall n.
\]

To see that Condition 2 is satisfied note that there exists a \( \delta \) and a \( \psi \) such that \( w_x(\delta) \leq \epsilon \) for all estimates of the p-p plot whose parameters satisfy

\[
|\mu_T, \mu_C, \sigma_T, \sigma_C - \tilde{\mu}_T, \tilde{\mu}_C, \tilde{\sigma}_T, \tilde{\sigma}_C| < \psi.
\]

Thus since

\[
(\tilde{\mu}_T, \tilde{\mu}_C, \tilde{\sigma}_T, \tilde{\sigma}_C) \to_P (\mu_T, \mu_C, \sigma_T, \sigma_C)
\]

it follows that condition 2 is satisfied.

To construct a simultaneous confidence band for the p-p plot we now appeal to the following theorem a proof of which is contained in Knowles (1988).

**Theorem 4.12:** Let \( (Z(t), 0 \leq t \leq 1) \) be a finite dimensional Gaussian random function with mean zero and variance \( \sigma^2(t) \) and suppose that its correlation function can be expressed as

\[
r(s, t) = \sum_{i=1}^{k} u_i(s)u_i(t),
\]
where \( u_1, u_2, \ldots u_k \) are differentiable functions on the unit interval. Then

\[
P(\sup[Z(t)/\sigma(t)] > z) \leq \frac{e^{-z^2/2}}{2\pi} \int_0^1 \left[ \sum_{i=1}^k (u_i(t))^2 \right]^{1/2} dt + 1 - \Phi(z).
\]

To implement this procedure for computing a simultaneous confidence band we need to compute

\[
\int_0^1 \left[ \sum_{i=1}^k (u_i^2(t)) \right]^{1/2} dt.
\]

For fixed \( \mu_T, \mu_C, \sigma_T, \) and \( \sigma_C \) the correlation between two points on the p-p plot is

\[
\text{corr} \left[ \Phi \left( k_{p_1} \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right), \Phi \left( k_{p_2} \frac{\sigma_C}{\sigma_T} - \frac{\mu_T - \mu_C}{\sigma_T} \right) \right] = \frac{v(p_1)\Sigma v(p_2)'}{\sqrt{v(p_1)\Sigma v(p_1)'v(p_2)\Sigma v(p_2)'}}.
\]

where

\[
v(p) = \left( -\frac{1}{\sigma_T}, -\frac{1}{\sigma_T}, -\frac{k_p\sigma_C - \mu_T + \mu_C}{2\sigma_T^3}, \frac{k_p}{2\sigma_T\sigma_C} \right).
\]

By inspection we see that

\[
u(p) = \frac{v(p)\Sigma^{1/2}}{\sqrt{v(p)\Sigma v(p)'}}.
\]

Upon differentiation of \( u(p) \) we obtain

\[
\frac{du(p)}{dp} = -\frac{1}{2} (v\Sigma v')^{-3/2} \left( \frac{dv}{dp} \Sigma v' + v\Sigma \frac{dv'}{dp} \right) v\Sigma^{1/2} + (v\Sigma v')^{-1/2} \frac{dv}{dp} \Sigma^{1/2},
\]

where

\[
\frac{dv(p)}{dp} = \left( 0, 0, \frac{\sigma_C'}{2\sigma_T^2 \phi(\Phi^{-1}(p))}, \frac{1}{2\sigma_T \sigma_C \phi(\Phi^{-1}(p))} \right).
\]

Thus

\[
\sum_{i=1}^k u_i(t)^2 = \frac{du(p)}{dp} \frac{du(p)'}{dp},
\]

and the evaluation of

\[
\int_0^1 \left[ \sum_{i=1}^k (u_i^2(t)) \right]^{1/2} dt
\]

may now be completed by numerical integration.

In the context of an estimation problem we do not know the values for \( \mu_T, \mu_C, \sigma_T \) and \( \sigma_C \). We are thus forced to use the maximum likelihood estimates as though they are the true values for the parameters.
4.4. An example.

Now we consider study 6 of the open education data set and describe the results of the study using p-p plots. The maximum likelihood estimate of the p-p plot and a 90 percent simultaneous confidence band are presented in Figure 4.2. These are computed as indicated in Section 3. Since the confidence band does not contain the 45 degree line the p-p plot is significantly different at the 90 percent level of significance from the situation where the treatment and control distributions are identical. Indeed since the estimate of the p-p plot lies below the 45 degree line for $p > .5$ and above the 45 degree line for $p < .5$ the treatment is beneficial to performance in the range [.5, 1] and is detrimental to performance in the range [0, .5]. Thus it appears that open education improves the problem solving ability of good students and hinders the performance of poor students.
Examples of p-p plots in the Gaussian model which are not convex, concave or symmetric.
Figure 4.2

The maximum likelihood estimate of the p-p plot for study 6 of the open education data set and a 90 percent simultaneous confidence band.
Chapter 5

The Sp-p Plot for Some Survival Distributions

Survival distributions play an important role in the medical literature and in reliability theory and are often used to compare different groups of individuals. For example when a number of different treatments are compared the survival distributions for the various treatments are frequently plotted together on the same graph and interpreted. This is a reasonable procedure for comparing treatment effects when the control groups corresponding to the various treatment groups are homogeneous. However the usefulness of comparing survival distributions as measures of treatment effects when the control distributions are not homogeneous is questionable since the variability in the control groups isn’t accounted for.

A comparison of sp-p plots when the control groups are not homogeneous is one way to account for variability among the control groups. The sp-p plots can be interpreted as standardized survival distributions where the standardization is accomplished by transforming the measurement scale so that the control distribution is uniform on the unit interval. Thus the transformed control distributions corresponding to sp-p plots are homogeneous and the sp-p plots may be compared directly as measures of treatment effects.

In this chapter we consider estimation of the sp-p plot in several models involving survival distributions. Specifically, we consider models for the treatment and control distributions that involve the exponential, gamma and Weibull distributions. An example utilizing the techniques developed for the gamma model is presented at the end of this chapter.
5.1. The exponential model.

In this section we consider the sp-p plot in the model where the treatment and the control distributions are exponential with parameters $\lambda_T$ and $\lambda_C$ respectively. By Lemma 3.2 the functional form for the sp-p plot in this model is given by

$$\tilde{F}_T(F_C^{-1}(p)), \quad (5.1)$$

where

$$\tilde{F}_T(x) = \exp(-\lambda_T x), \quad (5.2)$$
$$F_C(x) = 1 - \exp(-\lambda_C x). \quad (5.3)$$

Substituting (5.2) and (5.3) into (5.1) and simplifying we obtain,

$$\tilde{F}_T(F_C^{-1}(p)) = \exp[(\lambda_T/\lambda_C) \log(1 - p)]$$
$$= (1 - p)^{\lambda},$$

where $\lambda = \lambda_T/\lambda_C$.

We first present some properties of the sp-p plot in the exponential model and then consider maximum likelihood estimation of the sp-p plot and the construction of simultaneous confidence bands. It is also shown that in the exponential model a bounded unbiased estimate of the sp-p plot doesn’t exist.

5.1.1. Properties of the sp-p plot

The sp-p plot has some nice properties in the exponential case. Here it is demonstrated that when the treatment and control distributions are exponential, the sp-p plot is either convex or concave. Further we show that the sp-p plot is a maximal invariant in the exponential model.

Lemma 5.1: Suppose that the treatment and control distributions are exponential with parameters $\lambda_T$ and $\lambda_C$, respectively and $\lambda = \lambda_T/\lambda_C$. Then the sp-p plot is convex for $\lambda > 1$ and concave for $\lambda < 1$. 
Proof: It suffices to show that the second derivative is positive when \( \lambda > 1 \) and negative when \( \lambda < 1 \). Thus,
\[
\frac{d}{dp} (1 - p)^\lambda = -\lambda (1 - p)^{\lambda-1},
\]
\[
\frac{d^2}{dp^2} (1 - p)^\lambda = \lambda (\lambda - 1)(1 - p)^{\lambda-2}.
\]
The lemma implies that the sp-p plot is greater than \( 1 - p \) for \( \lambda < 1 \) and less than \( 1 - p \) for \( \lambda > 1 \), since \( 1 - p \) represents the line that connects the points \((0, 1)\) and \((1, 0)\) on the sp-p plot. Figure 5.1 displays some p-p plots for the exponential model.

Now we consider the maximal invariance of the sp-p plot in this model. One exponential distribution can be obtained from any other by a transformation of the form \( x \rightarrow ax \) where \( a > 0 \). The collection of all such transformations, \( A \), is a group. Thus by Lemma 3.5 the p-p plot in the exponential model is a maximal invariant with respect to \( A \). Since there is a one to one correspondence between p-p plots and sp-p plots, the sp-p plot is also a maximal invariant in the exponential model. Therefore the sp-p plot summarizes all of the information in the exponential model that is independent of multiplicative changes in the measurement scale.

5.1.2. Maximum likelihood estimation of the sp-p plot

In this subsection we obtain the maximum likelihood estimate of the sp-p plot. Suppose that
\[
X_1^T, \ldots, X_{n_T}^T \quad \text{and} \quad X_1^C, \ldots, X_{n_C}^C
\]
are random samples from exponential distributions with parameters \( \lambda_T \) and \( \lambda_C \) respectively. It is well known that the maximum likelihood estimates for \( \lambda_T \) and \( \lambda_C \) are \( 1/\bar{X}^T \) and \( 1/\bar{X}^C \) respectively, where \( \bar{X}^T \) and \( \bar{X}^C \) are the sample means. Thus the maximum likelihood estimate for \( \lambda \) is
\[
\hat{\lambda} = \frac{\bar{X}^C}{\bar{X}^T}
\]
and the maximum likelihood estimate for the sp-p plot is
\[
(1 - p)^{\hat{\lambda}}.
\]
We have proved the following lemma.

**Lemma 5.2:** The maximum likelihood estimate of the sp-p plot is \((1 - p)^\hat{\lambda}\).

In order to gauge the accuracy of the maximum likelihood estimate of the sp-p plot we need some indication of its variability. Lemma 5.3 presents the asymptotic distribution of estimates of the form of the maximum likelihood estimate. Using Lemma 5.3 we present in Theorem 5.4 the asymptotic distribution of the maximum likelihood estimate of the sp-p plot. Lemma 5.3 can be used in other contexts as well, for example in computing the asymptotic distribution of pooled estimates of the sp-p plot.

**Lemma 5.3:** Suppose that \(\hat{\lambda}_T\) and \(\hat{\lambda}_C\) are estimates of \(\lambda_T\) and \(\lambda_C\) satisfying

\[
\sqrt{n_T + n_C} \left( \frac{\hat{\lambda}_T - \lambda_T}{\hat{\lambda}_C - \lambda_C} \right) \rightarrow_L \mathcal{N}(0, \Sigma)
\]

as \(n_T \rightarrow \infty\) and \(n_C \rightarrow \infty\) such that \(n_T/(n_T + n_C) \rightarrow \tau\). Let \(\lambda = \lambda_T/\lambda_C\) and \(\hat{\lambda} = \hat{\lambda}_T/\hat{\lambda}_C\).

Then

\[
\sqrt{n_T + n_C} \left[ (1 - p)^{\hat{\lambda}} - (1 - p)^{\lambda} \right] \rightarrow_L \mathcal{N}(0, (a, b) \Sigma (a, b)',
\]

where

\[
a = \frac{1}{\lambda_C} \log(1 - p) (1 - p)^\lambda, \quad \text{and} \quad b = -\lambda \cdot a.
\]

**Proof:** The first partial derivatives of the sp-p plot with respect to \(\lambda_T\) and \(\lambda_C\) are \(a\) and \(b\) respectively. Since \(a\) and \(b\) are continuous, the p-p plot is totally differentiable, and we have

\[
(1 - p)^{\hat{\lambda}} - (1 - p)^{\lambda} = (\hat{\lambda}_T - \lambda_T)a + (\hat{\lambda}_C - \lambda_C)b + o_p(1 / \sqrt{n_T + n_C}).
\]

The theorem now follows from standard results on asymptotic distributions. ||

**Theorem 5.4:** Let \(\hat{\lambda} = \hat{X}_C / \hat{X}_T\) and suppose that \(n_T \rightarrow \infty\) and \(n_C \rightarrow \infty\) such that \(n_T/(n_T + n_C) \rightarrow \tau\). Then

\[
\sqrt{n_T + n_C} \left[ (1 - p)^{\hat{\lambda}} - (1 - p)^{\lambda} \right] \rightarrow_L \mathcal{N} \left( 0, \frac{\lambda^2}{r(1 - r)} \left[ \log(1 - p) (1 - p)^{\lambda} \right]^2 \right).
\]
Section 5.1: The exponential model

Proof: The result follows by a direct application of Theorem 5.3 with

\[ \Sigma = \begin{pmatrix} \lambda_T^2/r & 0 \\ 0 & \lambda_C^2/(1-r) \end{pmatrix}. \]

5.1.3. Simultaneous confidence bands for the sp-p plot

Because of the importance of the general shape of the sp-p plot in interpreting the treatment effect a simultaneous confidence band for the whole sp-p plot is of interest. Here we present a simultaneous confidence band for the sp-p plot in the model where the treatment and control distributions are exponential.

Suppose that

\[ X_1^T, \ldots, X_{n_T}^T, X_1^C, \ldots, X_{n_C}^C \]

are random samples from exponential distributions with parameters \( \lambda_T \) and \( \lambda_C \) respectively. Suppose also that \((a_1, a_2)\) is a \( 1-\alpha \) percent confidence interval for a beta distributed random variable with parameters \( n_T \) and \( n_C \). Then we have the following lemma.

Lemma 5.6: The interval

\[ \left[ (1-p)^{\frac{n_C}{n_T} \frac{a_2}{1-a_2}}, (1-p)^{\frac{n_C}{n_T} \frac{a_1}{1-a_1}} \right] \]

is a \( 1-\alpha \) percent confidence interval for the sp-p plot.

Proof: First note that

\[ \lambda_T \sum_{i=1}^{n_T} X_i^T \quad \text{and} \quad \lambda_C \sum_{i=1}^{n_C} X_i^C \]

have gamma distributions with parameters \( \lambda \) and \( r \) equal to 1, \( n_T \) and 1, \( n_C \) respectively. Thus

\[ 1/ \left( 1 + \frac{\hat{\lambda} n_C}{\lambda n_T} \right) \]

has a beta distribution with parameters \( n_T \) and \( n_C \) which implies that

\[ P \left( a_1 < 1/ \left( 1 + \frac{\hat{\lambda} n_C}{\lambda n_T} \right) < a_2 \right) = 1-\alpha \]
and hence
\[
P \left( \frac{\lambda_{nC}}{nT} \frac{a_2}{1 - a_2} < \lambda < \frac{\lambda_{nC}}{nT} \frac{a_1}{1 - a_1} \right) = 1 - \alpha.
\]
The lemma now follows since \((1 - p)\lambda\) is monotone in \(\lambda\) for all \(0 < p < 1\).

5.1.4. Unbiased estimation of the sp-p plot

In this section we consider unbiased estimation of the sp-p plot in the model where the treatment and the control distributions are exponential. That is we consider unbiased estimation of
\[
(1 - p)\lambda = \exp(\lambda \log(1 - p)). \tag{5.4}
\]
Suppose that \(X^T_1, \ldots, X^T_{nT}\) and \(X^C_1, \ldots, X^C_{nC}\) are random samples from exponential distributions with parameters \(\lambda_T\) and \(\lambda_C\) respectively. Since \(\sum X^T_i\) and \(\sum X^C_i\) are jointly complete and sufficient we restrict our attention to estimates that are functions of these statistics. We further restrict our attention to estimates that are bounded.

So suppose \(g(u, v, p)\), where \(u = \sum X^T_i\) and \(v = \sum X^C_i\), is an unbiased estimate of the sp-p plot, (5.4), that is bounded. Since \(g(u, v, p)\) is bounded there exists a constant \(C\) such that \(g(u, v, p) + C\) is an unbiased estimate of \(\exp(\lambda \log(1 - p)) + C\) that is bounded and non-negative. Unbiasedness implies that,
\[
\int_0^\infty \int_0^\infty [g(u, v, p) + C] \frac{\lambda_T^{nT}}{\Gamma(nT)} \frac{\lambda_C^{nC}}{\Gamma(nC)} u^{nT-1}v^{nC-1} \exp(-\lambda_T u - \lambda_C v) \, du \, dv
\]
\[
= \exp(\lambda \log(1 - p)) + C.
\]
Incorporating the gamma functions into \(g\) we can write
\[
\int_0^\infty \int_0^\infty [g(u, v, p) + C] u^{nT-1}v^{nC-1} \exp(-\lambda_T u - \lambda_C v) \, du \, dv
\]
\[
= \frac{1}{\lambda_T^{nT} \lambda_C^{nC}} \exp(\lambda \log(1 - p)) + C. \tag{5.5}
\]
Because \(g(u, v, p) + C\) is bounded and nonnegative, the derivative with respect to \(\lambda_C\) of the integrand in equation (5.5) is integrable with respect to product measure on \(u\) and \(v\).
Moreover the derivative of the integrand is less than or equal to
\[
[g(u, v, p) + C] u^{nT-1}v^{nC-1} \exp(-\lambda_T u - (\lambda_C - \epsilon) v)
\]
for all \( \lambda_C \in (\tilde{\lambda}_C - \epsilon, \tilde{\lambda}_C + \epsilon) \). Thus by the Lebesgue dominated convergence theorem

\[
\frac{d}{d\lambda_C} \int_0^\infty \int_0^\infty [g(u,v,p) + C] u^{n_T-1}v^{n_C-1} \exp(-\lambda_T u - \lambda_C v) \, du \, dv = -\int_0^\infty \int_0^\infty [g(u,v,p) + C] u^{n_T-1}v^{n_C-1} \exp(-\lambda_T u - \lambda_C v) \, du \, dv \leq 0
\]

for all \( \lambda_T \) and \( \lambda_C > 0 \). Thus if there exists an unbiased estimate of the sp-p plot that is bounded between 0 and 1, then

\[
\frac{d}{d\lambda_C} \frac{1}{\lambda_T^{n_T} \lambda_C^{n_C}} \exp(\lambda_T \log(1-p)) + C \leq 0
\]

(5.6)

for all \( \lambda_T \) and \( \lambda_C > 0 \). But the derivative in (5.6) is equal to

\[
\frac{1}{\lambda_T^{n_T-1} \lambda_C^{n_C+1}} \left[ \left( -\log(1-p) - \frac{n_C}{\lambda_T} \right) \exp(\lambda_T \log(1-p)) - \frac{C \, n_C}{\lambda_T} \right]
\]

which is greater than or less than zero depending on \( \lambda_T, \lambda_C \) and \( p \). Thus we have obtained a contradiction and must conclude that there doesn’t exist an unbiased estimate of the sp-p plot that is bounded.

Note that this result rules out the existence of an unbiased estimate of the sp-p plot in the gamma and Weibull model as well since these models are generalizations of the the exponential model.

If an unbiased estimate of the sp-p plot did exist then data from a number of different studies could be pooled by simply taking a weighted average of the unbiased sp-p plots from each study. Since there doesn’t exist an unbiased estimate of the sp-p plot a different procedure needs to be developed to pool results from independent studies.

### 5.2. The gamma model.

In this section we consider the sp-p plot in the model where the treatment and control observations have gamma distributions. Specifically we suppose that the observations from the treatment and the control groups have densities

\[
\frac{\lambda_T^{r_T} x^{r_T-1}}{\Gamma(r_T)} \exp(-\lambda_T x),
\]
\[ \frac{\lambda^c}{\Gamma(r)} x^{r-1} \exp(-\lambda x), \]
respectively, with unknown parameters \( \lambda_T, \lambda_C, r_T, \) and \( r_C. \) Let
\[ G_r(t) = \int_0^t \frac{1}{\Gamma(r)} x^{r-1} e^{-x} \, dx \]
denote the c.d.f., so that
\[ \bar{F}_T(x) = 1 - G_{r_T}(\lambda_T x), \]
\[ F^{-1}_C(p) = \frac{1}{\lambda_C} G^{-1}_{r_C}(p). \]
Then the sp-p plot becomes
\[ \bar{F}_T(F^{-1}_C(p)) = 1 - G_{r_T} \left( \lambda \frac{G^{-1}_C(p)}{r_C} \right) \]
\[ = \int_0^\infty \frac{1}{\Gamma(r)} x^{r-1} \exp(-x) \, dx, \]
where \( \lambda = \lambda_T/\lambda_C \) as in the previous section.

In this section maximum likelihood estimation of the sp-p plot in the gamma model is considered. In addition the asymptotic distribution of the derived estimate and the construction of confidence bands for the sp-p plot are investigated. First we present some properties of the sp-p plot in this case.

5.2.1. Properties of the sp-p plot

The following theorem shows that in the special case where \( r_T = r_C, \) the sp-p plot is either convex or concave.

**Theorem 5.7:** When \( r_T = r_C \) in the gamma model, the sp-p plot is convex if \( \lambda > 1 \) and concave if \( \lambda < 1. \)

**Proof:** It suffices to show that the second derivative of the sp-p plot is positive when \( \lambda > 1 \) and negative when \( \lambda < 1. \) Thus
\[ \frac{d}{dp} \int_0^\infty \frac{1}{\Gamma(r)} x^{r-1} e^{-x} \, dx = \]
\[ = \frac{1}{\Gamma(r)} \left( \lambda \frac{G^{-1}_C(p)}{r_C} \right)^{r-1} \exp \left( -\lambda \frac{G^{-1}_C(p)}{r_C} \right) \frac{d}{dp} \left( \lambda \frac{G^{-1}_C(p)}{r_C} \right). \]
Further,
\[
\frac{d}{dp} G^{-1}_r (p) = \left( \frac{d}{dx} G_r (x)|_{x=G^{-1}_r (p)} \right)^{-1} = \frac{\Gamma(r)}{G^{-1}_r (p)^{r-1} \exp \left(-G^{-1}_r (p) \right)}.
\]

So the right hand side of (5.7) becomes
\[
-\lambda^r \exp \left(-[\lambda - 1] G^{-1}_r (p) \right).
\]

Finally
\[
\frac{d^2}{dp^2} \int_{\lambda G^{-1}_r (p)}^\infty \frac{1}{\Gamma(r)} s^{r-1} e^{-s} \, ds = \lambda^r (\lambda - 1) \frac{\Gamma(r)}{G^{-1}_r (p)^{r-1} \exp \left(-[\lambda - 2] G^{-1}_r (p) \right)}
\]

which completes the proof. ||

When \( r_T \) and \( r_C \) are not equal the sp-p plot is not necessarily convex. This is evident from Figure 5.2 which presents some graphs of the sp-p plot when \( r_T \) and \( r_C \) are not equal.

Now we turn our attention to determining the maximal invariance of the sp-p plot. For a fixed shape parameter, \( r \), one gamma distribution can be obtained from any other by a transformation of the form, \( x \rightarrow ax \) where \( a > 0 \). Denote the collection of all such transformations by \( G \). Since \( G \) is a group of transformations Lemma 3.5 implies that the p-p plot is a maximal invariant in the gamma model where the shape parameters \( r_T \) and \( r_C \) for the treatment and control distributions are fixed. By the 1-1 correspondence between p-p plots and sp-p plots the sp-p plot is also a maximal invariant in this model. Thus for fixed \( r_T \) and \( r_C \) the sp-p plot summarizes all of the information in the experiment that is independent of multiplicative changes in the measurement scale.

If we do not want to suppose that \( r_T \) and \( r_C \) are fixed then we may imbed the family of pairs of gamma distributions in the collection of all pairs of absolutely continuous distributions and treat the sp-p plot in the gamma model as a maximal invariant with respect to the group of continuous monotone transformations.
5.2.2. Maximum likelihood estimation of the sp-p plot

In this subsection we obtain the maximum likelihood estimate of the sp-p plot and compute its asymptotic distribution. Suppose that we have random samples,

\[ X_1^T, \ldots, X_{n_T}^T \quad \text{and} \quad X_1^C, \ldots, X_{n_C}^C \]

from the treatment and the control distributions which are gamma with parameters \( r_T, \lambda_T \) and \( r_C, \lambda_C \) respectively. In addition suppose that \( r_T, \lambda_T, r_C, \) and \( \lambda_C \) are all unknown.

**Lemma 5.8:** The maximum likelihood estimate of the sp-p plot is

\[ \tilde{G}_{r_T} \left( \tilde{\lambda} G_{r_C}^{-1} (p) \right), \]

where \( \tilde{\lambda} = \tilde{\lambda}_T / \tilde{\lambda}_C \) and \( \tilde{r}_T, \tilde{\lambda}_T, \tilde{r}_C, \) and \( \tilde{\lambda}_C \) are defined by

\[ \tilde{X}^T = \frac{r_T}{\tilde{\lambda}_T}, \quad \tilde{X}^C = \frac{r_C}{\tilde{\lambda}_C}, \quad (5.8) \]

\[ \log \left[ \frac{\tilde{X}^T / \left( \prod_{i=1}^{n_T} X_i^T \right)^{1/n_T}}{\left[ \prod_{i=1}^{n_C} X_i^C \right]^{1/n_C}} \right] = \log(\tilde{r}_T) - \psi(\tilde{r}_T), \quad (5.9) \]

\[ \log \left[ \frac{\tilde{X}^C / \left( \prod_{i=1}^{n_C} X_i^C \right)^{1/n_C}}{\left[ \prod_{i=1}^{n_T} X_i^T \right]^{1/n_T}} \right] = \log(\tilde{r}_C) - \psi(\tilde{r}_C). \quad (5.10) \]

The function \( \psi(x) = -\frac{d}{dx} \log \Gamma(x) \) is the digamma function.

**Proof:** Equations (5.8), (5.9), and (5.10) yield the maximum likelihood estimates for \( \tilde{r}_T, \tilde{\lambda}_T, \tilde{r}_C, \) and \( \tilde{\lambda}_C \). The lemma now follows from the correspondence between the sp-p plot and \( (r_T, \lambda_T, r_C, \lambda_C) || \)

Before we present the asymptotic distribution of the maximum likelihood estimate of the sp-p plot we consider Lemma 5.9 which derives the asymptotic distribution of estimates of the sp-p plot based on estimates of \( (r_T, \lambda_T, r_C, \lambda_C) \). We use Lemma 5.9 in Theorem 5.10 to derive the asymptotic distribution of the estimate presented in Lemma 5.8. In addition we use Lemma 5.9 in later chapters when we consider estimates of the sp-p plot based on pooled data from a number different studies.
**Lemma 5.9:** Suppose that $\tilde{\lambda}_T, \tilde{\lambda}_C, \tilde{r}_T$ and $\tilde{r}_C$ are estimates of $\lambda_T, \lambda_C, r_T$ and $r_C$ satisfying

$$
\sqrt{n_T + n_C} \begin{pmatrix}
\tilde{\lambda}_T - \lambda_T \\
\tilde{\lambda}_C - \lambda_C \\
\tilde{r}_T - r_T \\
\tilde{r}_C - r_C
\end{pmatrix} \rightarrow_L N(0, \Sigma)
$$

as $n_T \to \infty$ and $n_C \to \infty$ such that $n_T/(n_T + n_C) \to r$. Let $\lambda = \lambda_T/\lambda_C$ and $\tilde{\lambda} = \tilde{\lambda}_T/\tilde{\lambda}_C$.

Then

$$
\sqrt{n_T + n_C} \left[ \tilde{G}_{T'} \left( \frac{\tilde{\lambda}}{\tilde{\lambda}_C} G_{r_C}^{-1}(p) \right) - G_{r_T} \left( \lambda G_{r_C}^{-1}(p) \right) \right]
\rightarrow_L N \left( 0, (a, b, c, d) \Sigma (a, b, c, d)' \right)
$$

where

$$
a = -g_{r_T} \left( \lambda G_{r_C}^{-1}(p) \right) \frac{1}{\lambda_C} G_{r_C}^{-1}(p),
$$

$$
b = -\lambda a,
$$

$$
c = \frac{\partial}{\partial r_T} G_{r_T} \left( \lambda G_{r_C}^{-1}(p) \right),
$$

$$
d = \frac{\partial}{\partial r_C} G_{r_T} \left( \lambda G_{r_C}^{-1}(p) \right).
$$

**Proof:** The first partial derivatives of the sp-p plot with respect to $\lambda_T, \lambda_C, r_T$ and $r_C$ are $a, b, c$ and $d$ respectively. Since $a, b, c$ and $d$ are continuous, the sp-p plot is totally differentiable, and we have that

$$
\sqrt{n_T + n_C} \left[ \tilde{G}_{T'} \left( \frac{\tilde{\lambda}}{\tilde{\lambda}_C} G_{r_C}^{-1}(p) \right) - G_{r_T} \left( \lambda G_{r_C}^{-1}(p) \right) \right] =
$$

$$(\tilde{\lambda}_T - \lambda_T)a + (\tilde{\lambda}_C - \lambda_C)b + (\tilde{r}_T - r_T)c + (\tilde{r}_C - r_C)d + o_p(1/\sqrt{n_T + n_C}).$$

The theorem now follows from standard results on asymptotic distributions. ||

Note that $c$ and $d$ are computed numerically because of the difficulty of representing these derivatives analytically.

**Theorem 5.10:** Let $\tilde{\lambda}_T, \tilde{\lambda}_C, \tilde{r}_T$ and $\tilde{r}_C$ be defined as in (5.8) - (5.10), let $\tilde{\lambda} = \tilde{\lambda}_T/\tilde{\lambda}_C$ and suppose that $n_T \to \infty$ and $n_C \to \infty$ such that $n_T/(n_T + n_C) \to r$. Then
\[
\sqrt{n_T + n_C} \left[ G_{r_T} \left( \lambda G_{r_C}^{-1}(p) \right) - G_{r_T} \left( \lambda G_{r_C}^{-1}(p) \right) \right] \\
\sim_{L} \mathcal{N} \left( 0, \frac{1}{r} (a, c) \Sigma_T (a, c)' + \frac{1}{1 - r} (b, d) \Sigma_C (b, d)' \right)
\]

where
\[
\Sigma_T = \frac{1}{\left( r_T \psi'(r_T) - 1 \right)} \begin{pmatrix} \lambda_T^2 \psi'(r_T) & \lambda_T \\ \lambda_T & r_T \end{pmatrix},
\]
\[
\Sigma_C = \frac{1}{\left( r_C \psi'(r_C) - 1 \right)} \begin{pmatrix} \lambda_C^2 \psi'(r_C) & \lambda_C \\ \lambda_C & r_C \end{pmatrix},
\]
and \(a, b, c\) and \(d\) are as in Lemma 5.9.

**Proof:** The result follows by a direct application of Theorem 5.9 with \(\Sigma\) defined as
\[
\Sigma = \begin{pmatrix} \Sigma_T / r & 0 \\ 0 & \Sigma_C / (1 - r) \end{pmatrix}.
\]

### 5.2.3. Simultaneous confidence bands for the sp-p plot

In Section 5.2.2 we present the asymptotic distribution of the maximum likelihood estimate of the sp-p plot. This result enables us to compute approximate confidence intervals for the sp-p plot at any specific percentile of the control distribution. The confidence interval takes the form
\[
\left[ 1 - \hat{F}_T(\hat{F}_C^{-1}(p)) - z_{\alpha/2} \hat{\sigma}(p), 1 - \hat{F}_T(\hat{F}_C^{-1}(p)) + z_{\alpha/2} \hat{\sigma}(p) \right]
\]
where \(1 - \hat{F}_T(\hat{F}_C^{-1}(p))\) and \(\hat{\sigma}(p)\) are respectively the maximum likelihood estimates of the sp-p plot and the standard deviation of the sp-p plot at \(p\), and \(z_{\alpha/2}\) is the \(1 - \alpha/2\) quantile of the standard normal distribution. In meta-analysis the relation of the sp-p plot to the 45 degree line and to other sp-p plots is of interest which suggests the need for a simultaneous confidence band for the sp-p plot. In this section we determine the constant \(C_{\alpha/2}\) to substitute for \(z_{\alpha/2}\) which makes the collection of confidence intervals for \(0 < p < 1\) a \(1 - \alpha\) percent simultaneous confidence band for the sp-p plot. We start by showing that asymptotically the maximum likelihood estimate of the sp-p plot behaves as a finite dimensional Gaussian random function.
Lemma 5.11: Suppose that $\tilde{\lambda}_T, \tilde{\lambda}_C, \tilde{r}_T$ and $\tilde{r}_C$ are estimates of $\lambda_T, \lambda_C, r_T$ and $r_C$ satisfying

$$
\sqrt{n_T + n_C} \begin{pmatrix}
\tilde{\lambda}_T - \lambda_T \\
\tilde{\lambda}_C - \lambda_C \\
\tilde{r}_T - r_T \\
\tilde{r}_C - r_C
\end{pmatrix} \rightarrow_L N(0, \Sigma)
$$

as $n_T \to \infty$ and $n_C \to \infty$ such that $n_T/(n_T + n_C) \to r$. Let $\lambda = \lambda_T/\lambda_C$ and $\bar{\lambda} = \bar{\lambda}_T/\bar{\lambda}_C$. Then

$$
\sqrt{n_T + n_C} \left[ \tilde{G}_{rrT} \left( \lambda G_{rrC}^{-1}(p) \right) - \bar{G}_{rrT} \left( \lambda G_{rrC}^{-1}(p) \right) \right]
$$

(5.11)

converges in distribution to a finite dimensional Gaussian random function with covariance function given by

$$
[a(p_1), b(p_1), c(p_1), d(p_1)] \sum [a(p_2), b(p_2), c(p_2), d(p_2)],
$$

(5.12)

where

$$
a(p) = -g_{rrT} \left( \lambda G_{rrC}^{-1}(p) \right) \frac{1}{\lambda_C} G_{rrC}^{-1}(p),
b(p) = -\lambda a,
c(p) = \frac{\partial}{\partial r_T} \tilde{G}_{rrT} \left( \lambda G_{rrC}^{-1}(p) \right),
d(p) = \frac{\partial}{\partial r_C} \tilde{G}_{rrT} \left( \lambda G_{rrC}^{-1}(p) \right).
$$

Proof: The finite dimensional distributions of (5.11) converge to those of a finite dimensional Gaussian random function with (5.12) as a covariance function. Thus to show that (5.11) converges in distribution to a finite dimensional random function it needs to be shown that the sequence of distributions corresponding to (5.11) is tight. A sequence of probability measures, $P_n$, on the space of continuous functions on the unit interval, $C[0,1]$, is tight if and only if (Billingsley 1968)

1. For each positive $\eta$, there exists an $a$ such that

$$
P_n(x \in C[0,1] : \|x(0)\| > a) \leq \eta, \quad n \geq 1.
$$
2. For each positive $\epsilon$ and $\eta$, there exists a $\delta$ with $0 < \delta < 1$, and an $n_0$ such that

$$P_n(x \in C[0,1] : w_x(\delta) \geq \epsilon) \leq \eta \quad n \geq n_0,$$

where $w_x(\delta) = \sup_{|s-t|<\delta}|x(s) - x(t)|$.

Condition 1 is satisfied since

$$P_n(x \in C[0,1] : |x(0)| = 1) = 1 \quad \forall n.$$

To show that Condition 2 is satisfied note that there exists a $\delta$ and a $\psi$ such that $w_x(\delta) \leq \epsilon$ for all estimates of the sp-p plot with

$$|(\lambda_T, \lambda_C, r_T, r_C) - (\hat{\lambda}_T, \hat{\lambda}_C, \hat{r}_T, \hat{r}_C)| < \psi.$$

Since

$$(\hat{\lambda}_T, \hat{\lambda}_C, \hat{r}_T, \hat{r}_C) \to_P (\lambda_T, \lambda_C, r_T, r_C),$$

it follows that Condition 2 is satisfied.

Theorem 4.12 in Chapter 4 can be used to compute a conservative value for $C_\alpha$. To implement this procedure for computing a simultaneous confidence band it is necessary to compute

$$\int_0^1 \left[ \sum u(t)^2 \right]^{1/2} dt$$

numerically. One can start by computing $u(p)$ as in the gaussian case. The remaining differentiation and integration can be performed numerically.

### 5.3. The Weibull model.

In this section we investigate the sp-p plot in the model where the treatment and control observations are assumed to have Weibull distributions. That is we assume

$$\bar{F}_T (x) = \exp \left( -\frac{x}{a_T} \right)^b,$$

$$F_C (x) = 1 - \exp \left( -\frac{x}{a_C} \right)^b.$$
The functional form for the sp-p plot is thus

\[ F_T(F_C^{-1}(p)) = \exp\left(-a_T^{b_T}[-\log(1 - p)]^b\right), \]

where \( a = a_C/a_T \) and \( b = b_T/b_C \). We consider maximum likelihood estimation of the sp-p plot in this model and compute the asymptotic distribution of the derived estimate. First we examine some of the properties of the sp-p plot in this case.

5.3.1. Properties of the sp-p plot

Here we present some properties of the sp-p plot in the Weibull model. First we show that under certain circumstances the sp-p plot is either convex or concave. Then we consider the maximal invariance of the sp-p plot in the Weibull model.

For the convexity of the sp-p plot in the Weibull model note that when \( b_T \) and \( b_C \) are equal the sp-p plots for the Weibull model correspond to the sp-p plots for the exponential model. Thus they are convex when \( a < 1 \) and concave when \( a > 1 \). When \( b_T \) doesn't equal \( b_C \) the sp-p plots are not necessarily convex. This is evident from Figure 5.3 which displays some graphs of the sp-p plot in this model.

Now consider the maximal invariance of the sp-p plot in the Weibull model. For a fixed shape parameter, one Weibull distribution can be obtained from any other by a multiplicative transformation of the measurement scale. Thus Lemma 3.5 implies that the sp-p plot is a maximal invariant with respect to multiplicative transformations in the Weibull model with fixed shape parameters \( b_T \) and \( b_C \). If we do not want to consider \( b_T \) and \( b_C \) as fixed then we may imbed the collection of all pairs of Weibull distributions in the collection of all pairs of absolutely continuous distributions and consider the sp-p plot as a maximal invariant with respect to the the group of monotone increasing transformations.

5.3.2. Maximum likelihood estimation of the sp-p plot

In this section we present the maximum likelihood estimate of the sp-p plot in the Weibull model and compute its asymptotic distribution. Suppose that

\[ X_1^T, \ldots, X_{n_T}^T \quad \text{and} \quad X_1^C, \ldots, X_{n_C}^C \]
are random samples from Weibull distributions with parameters $a_T$, $b_T$ and $a_C$, $b_C$ respectively and that all of the parameters are unknown. Then we have the following lemma.

**Lemma 5.12:** The maximum likelihood estimate of the sp-p plot in the Weibull model is given by

$$F_T(F_C^{-1}(p)) = \exp\left(-\hat{a}^b \left[-\log(1 - p)\right]^b\right)$$

where $\hat{a} = \hat{a}_C/\hat{a}_T$, $\hat{b} = \hat{b}_T/\hat{b}_C$ and $\hat{a}_T$, $\hat{b}_T$, $\hat{a}_C$ and $\hat{b}_C$ are determined by

$$\hat{a}_T^b = \frac{1}{n_T} \sum_{i=1}^{n_T} (X_i^T)^b, \quad \frac{1}{n_T} \sum_{i=1}^{n_T} \log(X_i^T),$$

$$\hat{a}_C^b = \frac{1}{n_C} \sum_{i=1}^{n_C} (X_i^C)^b, \quad \frac{1}{n_C} \sum_{i=1}^{n_C} \log(X_i^C),$$

$$1/\hat{b}_T = \left(\sum_{i=1}^{n_T} (X_i^T)^b \log X_i^T\right) \left(\sum_{i=1}^{n_T} (X_i^T)^b\right)^{-1} - \frac{1}{n_T} \sum_{i=1}^{n_T} \log X_i^T,$$

$$1/\hat{b}_C = \left(\sum_{i=1}^{n_C} (X_i^C)^b \log X_i^C\right) \left(\sum_{i=1}^{n_C} (X_i^C)^b\right)^{-1} - \frac{1}{n_C} \sum_{i=1}^{n_C} \log X_i^C.$$

**Proof:** The maximum likelihood estimates of $a_T$, $b_T$, $a_C$ and $b_C$ are the solutions to the above equations. The lemma follows from the correspondence between the parameters $a_T$, $b_T$, $a_C$ and $b_C$ and the sp-p plot. ||

In order to determine the accuracy of an estimate of the sp-p plot we need some idea of its variability. Lemma 5.13 presents the asymptotic distribution of any estimate of the sp-p plot that is based on estimates of $a_T$, $b_T$, $a_C$ and $b_C$. With the aid of Lemma 5.13 we present in Theorem 5.14 the asymptotic distribution of the maximum likelihood estimate presented in Lemma 5.12. Moreover we use Lemma 5.13 in later chapters when we consider estimation of the sp-p plot based on pooled data from a number of different studies.

**Lemma 5.13:** Suppose that $\tilde{a}_T$, $\tilde{b}_T$, $\tilde{a}_C$ and $\tilde{b}_C$ are estimates of $a_T$, $b_T$, $a_C$ and $b_C$
satisfying
\[
\sqrt{n_T + n_C} \begin{pmatrix}
\frac{\hat{a}_T - a_T}{b_T - b_T} \\
\frac{\hat{b}_T - b_T}{a_C - a_C} \\
\frac{\hat{a}_C - a_C}{b_C - b_C}
\end{pmatrix} \rightarrow_L \mathcal{N}(0, \Sigma)
\]
as \(n_T \rightarrow \infty\) and \(n_C \rightarrow \infty\) such that \(n_T/(n_T + n_C) \rightarrow r\). If \(\hat{a} = \hat{a}_C / \hat{a}_T\) and \(\hat{b} = \hat{b}_T / \hat{b}_C\) then
\[
\sqrt{n_T + n_C} \left[ \exp \left( -\hat{a}^{b_T} \left( -\log(1 - p) \right)^{\hat{b}} \right) - \exp \left( -a^{b_T} \left( -\log(1 - p) \right)^{b} \right) \right] \\
\rightarrow_L \mathcal{N}(0, E^2 (e, f, g, h) \Sigma (e, f, g, h)')
\]
where
\[
E = \exp \left( -a^{b_T} \left( -\log(1 - p) \right)^{b} \right) \left[ -\log(1 - p) \right]^b,
\]
\[
e = a_C^{b_T} \frac{b_T}{a_T^{b_T+1}},
\]
\[
f = -\log \left( a \left[ -\log(1 - p) \right]^{1/b_C} \right) a^{b_T},
\]
\[
g = -\frac{1}{a_T^{b_T}} b_T a_C^{b_T-1},
\]
\[
h = a^{b_T} \frac{b}{b_C} \log \left[ -\log(1 - p) \right].
\]

**Proof:** The first partial derivatives of the sp-p plot with respect to \(a_T, b_T, a_C, b_C\) are \(E\) multiplied by \(e, f, g\) and \(h\) respectively. Since \(e, f, g\) and \(h\) are continuous the sp-p plot is totally differentiable and we obtain
\[
\exp \left( -\hat{a}^{b_T} \left( -\log(1 - p) \right)^{\hat{b}} \right) - \exp \left( -a^{b_T} \left( -\log(1 - p) \right)^{b} \right) =
\]
\[
\left[ (\hat{a}_T - a_T)e + (\hat{b}_T - b_T)f + (\hat{a}_C - a_C)g + (\hat{b}_C - b_C)h \right] \cdot E + o_p(1/\sqrt{n_T + n_C}).
\]
The result now follows from standard results on asymptotic distributions. ||

**Theorem 5.14:** Let \(\hat{a}_T, \hat{b}_T, \hat{a}_C\) and \(\hat{b}_C\) be as in Lemma 5.12, let \(\hat{a} = \hat{a}_C / \hat{a}_T\), and suppose that \(n_T \rightarrow \infty\) and \(n_C \rightarrow \infty\) such that \(n_T/(n_T + n_C) \rightarrow r\). Then
\[
\sqrt{n_T + n_C} \left[ \exp \left( -\hat{a}^{b_T} \left( -\log(1 - p) \right)^{\hat{b}} \right) - \exp \left( -a^{b_T} \left( -\log(1 - p) \right)^{b} \right) \right]
\]
5.3.3. Simultaneous confidence bands for the sp-p plot

In Section 5.3.2 we present the asymptotic distribution of the maximum likelihood estimate of the sp-p plot at a particular value of $p$. This result enables us to compute approximate confidence intervals for the sp-p plot at any specific percentile of the control distribution. The confidence interval takes the form

$$
\left[ \hat{F}_T(\hat{F}_{C}^{-1}(p)) - z_{\alpha/2} \hat{\sigma}(p), \hat{F}_T(\hat{F}_{C}^{-1}(p)) + z_{\alpha/2} \hat{\sigma}(p) \right]
$$

where $\hat{F}_T(\hat{F}_{C}^{-1}(p))$ and $\hat{\sigma}(p)$ are respectively the maximum likelihood estimates of the sp-p plot and the standard deviation of the sp-p plot at $p$, and $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution. In meta-analysis the relation of a sp-p plot to the 45 degree line and to other sp-p plots is of interest which suggests the need for a simultaneous confidence band for the sp-p plot. In this section we determine the constant $C_{\alpha/2}$ to substitute for $z_{\alpha/2}$ which makes the collection of confidence intervals for all $0 < p < 1$ a simultaneous confidence band for the sp-p plot. We start by showing in the following lemma that asymptotically the sp-p plot behaves as a finite dimensional Gaussian random function.
Lemma 5.15: Suppose that $\hat{a}_T$, $\hat{b}_T$, $\hat{a}_C$ and $\hat{b}_C$ are estimates of $a_T$, $b_T$, $a_C$ and $b_C$ satisfying

$$\sqrt{n_T + n_C} \begin{pmatrix} \hat{a}_T - a_T \\ \hat{b}_T - b_T \\ \hat{a}_C - a_C \\ \hat{b}_C - b_C \end{pmatrix} \to L \mathcal{N}(0, \Sigma)$$

as $n_T \to \infty$ and $n_C \to \infty$ such that $n_T/(n_T + n_C) \to r$. If $\tilde{a} = \hat{a}_C / \hat{a}_T$ and $\tilde{b} = \hat{b}_T / \hat{b}_C$ then

$$\sqrt{n_T + n_C} \left[ \exp \left( - \tilde{a}^b T \left[ -\log (1 - p) \right]^b \right) - \exp \left( -a_T^{b_T} \left[ -\log (1 - p) \right]^{b_T} \right) \right], \quad (5.13)$$

converges in distribution to a finite dimensional Gaussian random function with covariance function given by

$$E(p_1) E(p_2) [e, f(p_1), g, h(p_1)] \Sigma [e, f(p_2), g, h(p_2)]' \quad (5.14)$$

where

$$E(p) = \exp \left( -a_T^{b_T} \left[ -\log (1 - p) \right]^{b_T} \right) \left[ -\log (1 - p) \right]^{b_T},$$

$$e = a_C^{b_T} \frac{b_T}{a_T^{b_T+1}},$$

$$f(p) = -\log \left(a \left[ -\log (1 - p) \right]^{1/b_C} \right) a_T^{b_T},$$

$$g = -\frac{1}{a_T^{b_T}} b_T a_C^{b_T-1},$$

$$h(p) = b_T \frac{1}{b_C} \log \left[ -\log (1 - p) \right].$$

Proof: The finite dimensional distributions of (5.13) converge to those of a finite dimensional Gaussian random function with (5.14) as a covariance function. Thus to show that (5.13) converges in distribution to a finite dimensional random function it needs to be shown that the sequence of distributions corresponding to (5.13) is tight. A sequence of probability measures, $P_n$, on the continuous functions on the unit interval, $C[0,1]$, is tight if and only if (Billingsley 1968)
1. For each positive $\eta$, there exists an $a$ such that

$$P_n(x \in C[0,1]: |x(0)| > a) \leq \eta, \quad n \geq 1.$$ 

2. For each positive $\epsilon$ and $\eta$, there exists a $\delta$ with $0 < \delta < 1$, and an $n_0$ such that

$$P_n(x \in C[0,1]: w_x(\delta) \geq \epsilon) \leq \eta, \quad n \geq n_0,$$

where $w_x(\delta) = \sup_{|s-t|<\delta} |x(s) - x(t)|$. Condition 1 is satisfied since

$$P_n(x \in C[0,1]: |x(0)| = 1) = 1 \quad \forall n.$$ 

To see that Condition 2 is satisfied note that there exists a $\delta$ and a $\psi$ such that $w_x(\delta) \leq \epsilon$ for all estimates of the sp-p plot such that

$$|(a_T, b_T, a_C, b_C) - (\tilde{a}_T, \tilde{b}_T, \tilde{a}_C, \tilde{b}_C)| < \psi.$$ 

Thus since

$$(\tilde{a}_T, \tilde{b}_T, \tilde{a}_C, \tilde{b}_C) \rightarrow_P (a_T, b_T, a_C, b_C)$$

Condition 2 is satisfied.

To implement this procedure it is necessary to evaluate

$$\int_0^1 \left[ \sum u'(t)^2 \right]^{1/2} dt$$

numerically as in the Gamma model.

5.4. An example.

In this section we construct sp-p plots for the estrogen and progesterone receptor data using the techniques developed for the gamma model. First consider the estrogen receptor data. The treatment and control survival distributions for each estrogen category are displayed in Figure 2.3. Note that the control distributions are not homogeneous across categories and hence that a comparison of sp-p plots as measures of treatment
effects is warranted. The estimates of the sp-p plot derived from the gamma model are presented in Figure 5.4. From the general shape of these estimates we can see that there is a positive treatment effect in each case since the 90 percent simultaneous confidence bands lie above the line $1 - p$. In addition it appears that the treatment effect increases as the progesterone receptor concentration increases since the sp-p plots move to the right. The estimates and confidence bands also allow us to make statements like the following.

1. When the concentration of estrogen receptors is greater than 10 fmole we are 90 percent confident that between 25 and 70 percent of the treatment group will be alive when 80 percent of the control group is dead.

2. When the concentration of estrogen receptors is greater than 50 fmole, we are 90 percent confident that between 40 and 80 percent of the treatment group will be alive when 80 percent of the control group is dead.

3. When the concentration of estrogen receptors is greater than 100 fmole, we are 90 percent confident that between 75 and 99 percent of the treatment group will be alive when 80 percent of the control group is dead.

Statements 1, 2, and 3 are comparable since they all refer to survival in the treatment group when 80 percent of the control group is dead.

Next we construct the sp-p plots for the hypothetical progesterone receptor data. The treatment and control survival distributions for each category are displayed in Figure 2.4. Since the control distributions are not homogeneous looking at the sp-p plots is useful. The estimates for the sp-p plots derived from the gamma model are presented in Figure 5.5. The similarity of these estimates suggest that the treatment effect doesn't vary as the progesterone receptor concentrations change. That is as the progesterone receptor concentrations increase the relationship between the percentiles of the treatment and control groups doesn't vary. As with the estrogen data these estimates also allow us to make statements about the mortality of the treatment group which are comparable across the different progesterone categories. For example we may state that:
1. When the concentration of progesterone receptors is greater than 10 fmole we are 90 percent confident that between 40 and 95 percent of the treatment group will be alive when 80 percent of the control group is dead.

2. When the concentration of progesterone receptors is greater than 50 fmole, we are 90 percent confident that between 20 and 75 percent of the treatment group will be alive when 80 percent of the control group is dead.

3. When the concentration of progesterone receptors is greater than 100 fmole, we are 90 percent confident that between 50 and 95 percent of the treatment group will be dead when 80 percent of the control group is dead.
Figure 5.1

Sp-p plots of $F_T$ versus $F_C$ for the exponential distribution with scale parameters $\lambda_T$ and $\lambda_C$ satisfying $\lambda = \lambda_T/\lambda_C = 4, 2, 1, .5, .25$. 
Figure 5.2

Sp-p plots of $F_T$ versus $F_C$ for the gamma distribution with parameters $(r_T, \lambda_T, r_C, \lambda_C)$. 

![Diagram showing Sp-p plots for gamma distribution parameters](image-url)

- $(8, 5, 2, 1)$
- $(8, 3, 2, 1)$
- $(8, 2, 2, 1)$
- $(2, 2, 8, 1)$
- $(2, 3, 8, 1)$
- $(2, 4, 8, 1)$

$F_T$ vs. $F_C$
Examples of sp-p plots in the Weibull model, with parameters \((a, b_T, b_C)\), which are not convex or concave.
Figure 5.4

The empirical Sp-p plots for the estrogen receptor data along with the maximum likelihood estimates and 90 percent simultaneous confidence bands.
Figure 5.5

The empirical Sp-p plots for the progesterone receptor data along with the maximum likelihood estimates and 90 percent simultaneous confidence bands.
Chapter 6

Combined Estimates of P-p and Sp-p

Plots.

In Chapters 4 and 5 we consider the problem of estimating a single p-p or sp-p plot based on the information in a single study. The estimates derived therein are useful when it is desired to present the results of a study in a way that is readily comparable with results from other studies. In a meta-analysis, these estimates are just the starting point since when dealing with a number of studies we are concerned with more than just the presentation of results. We are also interested in improving the accuracy of estimates and the comparisons of results across studies. These concerns call for different estimation procedures.

In this chapter we develop two types of estimates of p-p and sp-p plots that may be usefully employed in a meta-analysis. One estimate improves the accuracy of the estimate of a treatment effect common to a number of studies. The other estimate aids in the comparison of results across studies. We develop both of these estimates using the method of maximum likelihood. In addition to studying these estimation problems we construct tests of the assumptions underlying these approaches. One test is for the hypothesis that the treatment effects are homogeneous and the other is for the hypothesis that the treatment effects can be ordered. We develop both of these tests using the likelihood ratio statistic. All of these estimates and tests are developed in three different models, the Gaussian, the gamma, and the Weibull. We start with the Gaussian model.
6.2. The Gaussian model.

Suppose that $K$ studies are under consideration and that for $j$'th study, the treatment and control populations are Gaussian with means $\mu_{Tj}$, $\mu_{Cj}$ and standard deviations $\sigma_{Tj}$, $\sigma_{Cj}$ respectively. That is for $j = 1, \ldots, K$,

\[ F_{Tj}(x) = \mathcal{N}(\mu_{Tj}, \sigma_{Tj}) , \]
\[ F_{Cj}(x) = \mathcal{N}(\mu_{Cj}, \sigma_{Cj}) . \]

The p-p plot for the $j$'th study then assumes the functional form,

\[ F_{Tj}(F_{Cj}^{-1}(p)) = \Phi \left( \frac{\sigma_{Cj} k_p - \mu_{Tj} - \mu_{Cj}}{\sigma_{Tj}} \right) \]

where $k_p$ denotes as before, $\Phi^{-1}(p)$.

6.2.1. Estimating a common p-p plot

Here we consider the case where the treatment effect represented by the p-p plot is constant across studies. We first develop an estimate for the common p-p plot based on all $K$ studies and then construct a test of the assumption that the treatment effect is homogeneous across studies.

In the Gaussian model we have put forth, two p-p plots are identical if and only if there exists a linear transformation of the scale of measurement of one study that equates its treatment and control distributions with the distributions of the other study. Thus the p-p plots for $K$ studies are identical if and only if there exists constants $a_j$ and $b_j$, $j = 2, \ldots, K$ such that

\[ a_j \mu_{T1} + b_j = \mu_{Tj} , \]
\[ a_j \mu_{C1} + b_j = \mu_{Cj} , \]
\[ a_j \sigma_{T1} = \sigma_{Tj} , \]
\[ a_j \sigma_{C1} = \sigma_{Cj} . \]

By incorporating these conditions into the Gaussian model for the p-p plot we can obtain an estimate of the common p-p plot.
In the resulting model the i'th observation from the treatment or control group of the j'th study, $X_{ij}^T$ or $X_{ij}^C$, is distributed as

$$X_{ij}^T = \mathcal{N}(a_j \mu_T + b_j, \sigma_T), \quad i = 1, \ldots, n_{Tj}, \quad j = 1, \ldots, K,$$

$$X_{ij}^C = \mathcal{N}(a_j \mu_C + b_j, \sigma_C), \quad i = 1, \ldots, n_{Cj}, \quad j = 1, \ldots, K,$$

where $a_1 = 1$ and $b_1 = 0$ and the remaining $a$'s and $b$'s are unknown. Fitting this model by maximum likelihood thus involves finding the values of the unknown parameters that maximize the log likelihood function,

$$l = -\frac{1}{2} \sum_{j=1}^{K} (n_{Tj} + n_{Cj}) \log(2\pi) \sigma_T^2 - \sum_{j=1}^{K} n_{Tj} \log(\sigma_T) - \sum_{j=1}^{K} n_{Cj} \log(\sigma_C) - \sum_{j=1}^{K} (n_{Tj} + n_{Cj}) \log(a_j)
- \frac{1}{2} \sum_{j=1}^{K} \frac{\sum_{i=1}^{n_{Tj}} (X_{ij}^T - (a_j \mu_T + b_j))^2}{\sigma_T^2}
- \frac{1}{2} \sum_{j=1}^{K} \frac{\sum_{i=1}^{n_{Cj}} (X_{ij}^C - (a_j \mu_C + b_j))^2}{\sigma_C^2}.$$  

The maximum of the log likelihood function is found by computing the first partial derivatives and setting them equal to zero. The partial derivatives are,

$$\frac{\partial l}{\partial a_j} = -\frac{n_{Tj} + n_{Cj}}{a_j} + \frac{\sum_{i=1}^{n_{Tj}} (X_{ij}^T - (a_j \mu_T + b_j))^2}{a_j^2 \sigma_T^2} + \frac{\sum_{i=1}^{n_{Cj}} (X_{ij}^C - (a_j \mu_C + b_j))^2}{a_j^2 \sigma_C^2} - \sum_{i=1}^{n_{Tj}} X_{ij}^T - n_{Tj} (a_j \mu_T + b_j),$$

$$\frac{\partial l}{\partial b_j} = -\frac{\sum_{i=1}^{n_{Tj}} X_{ij}^T - n_{Tj} (a_j \mu_T + b_j)}{(a_j \sigma_T)^2} = -\frac{\sum_{i=1}^{n_{Cj}} X_{ij}^C - n_{Cj} (a_j \mu_C + b_j)}{(a_j \sigma_C)^2},$$

$$\frac{\partial l}{\partial \mu_T} = \sum_{j=1}^{M} \frac{\sum_{i=1}^{n_{Tj}} X_{ij}^T - n_{Tj} (a_j \mu_T + b_j)}{a_j \sigma_T^2},$$

$$\frac{\partial l}{\partial \mu_C} = \sum_{j=1}^{M} \frac{\sum_{i=1}^{n_{Cj}} X_{ij}^C - n_{Cj} (a_j \mu_C + b_j)}{a_j \sigma_C^2},$$

$$\frac{\partial l}{\partial \sigma_T^2} = \sum_{j=1}^{M} \left[ \frac{\sum_{i=1}^{n_{Tj}} (X_{ij}^T - (a_j \mu_T + b_j))^2}{2 \sigma_T^2} + \frac{\sum_{i=1}^{n_{Tj}} X_{ij}^T - n_{Tj} (a_j \mu_T + b_j)^2}{2 a_j^2 \sigma_T^4} \right].$$
\[
\frac{\partial l}{\partial \sigma_C^2} = \sum_{j=1}^{M} \left[ -\frac{n_{ij}^C}{2\sigma_C^2} + \frac{\sum_{i=1}^{n_{ij}^C} X_{ij}^C}{2\sigma_C^2 - 2(a_j \mu_C + b_j) \sum_{i=1}^{n_{ij}^C} X_{ij}^C + n_{ij}^C(a_j \mu_C + b_j)^2} \right].
\]

The values of the unknown parameters that maximize the log likelihood function are thus the solution of a system of quadratic equations. This system of equations can be solved for numerically using the Newton-Raphson algorithm. However as the number of studies under consideration increases so does the number of equations and unknowns and hence the amount of computation needed to determine the solution.

Specifically the amount of computation required for a single loop of the algorithm is proportional to the cube of the number of studies or equivalently is of the order \(O(K^3)\). To see this note that if there are \(K\) studies under consideration then there are \(2(K - 1) + 4\) equations and unknowns. Each step of the Newton-Raphson algorithm involves inverting the matrix of second partial derivatives of the log likelihood function which is an \(2(K - 1) + 4\) by \(2(K - 1) + 4\) matrix. Since inverting this matrix of second partials requires \(O(K^3)\) computations, each loop of the Newton-Raphson Algorithm requires \(O(K^3)\) computations.

By employing a two step algorithm similar in nature to the E.M. algorithm the number of computations at each step can be reduced from \(O(K^3)\) to \(O(K)\). The algorithm is as follows.

1. Divide the system of equations into two groups, A and B. Group A consists of the partial derivatives of the log likelihood function with respect to \(\mu_T, \mu_C, \sigma_T^2,\) and \(\sigma_C^2\) set equal to zero while group B consists of the partial derivatives with respect to \(a_j, b_j : j = 2, \ldots, K,\) set equal to zero.

2. Start with initial values for \(\mu_T, \mu_C, \sigma_T^2, \sigma_C^2.\)

3. Solve for \(a_j, b_j : j = 2, \ldots, K\) by applying the Newton-Raphson algorithm to the equations in group B.

4. Take the solutions for \(a_j, b_j j = 2, \ldots, K\) from step 3 and solve for \(\mu_T, \mu_C, \sigma_T^2, \sigma_C^2\) by applying the Newton-Raphson algorithm to the equations in group A.
5. If \( a_j, b_j, \mu_T, \mu_C, \sigma_T^2, \sigma_C^2 \) have not changed much from their previous values then stop. Otherwise go to step 3.

Each loop through this algorithm requires \( O(K) \) instead of \( O(K^3) \) algebraic computations. To see this first note that solving for \( a_j, b_j \) \( j = 2, \ldots, K \) at stage 3 requires \( k \cdot 2^3(K - 1) \) computations since the partials of the log likelihood are such that for each \( j \), \( a_j \) and \( b_j \) can be solved for independently of the other a's and b's. Further note that solving for \( \mu_T, \mu_C, \sigma_T^2, \sigma_C^2 \) at stage 4 requires \( k \cdot 4^3 \) computations. Adding things up we obtain \( k \cdot [2^3(K - 1) + 4^3] \) or \( O(K) \) computations for one loop through this algorithm. Thus this modification of the Newton-Raphson algorithm makes it possible to fit the model for a common p-p plot when a large number of studies are under consideration.

Denote the resulting maximum likelihood estimates for \( a_j, b_j, \mu_T, \mu_C, \sigma_T^2, \sigma_C^2 \) by \( \hat{a}_j, \hat{b}_j, \hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}_T^2, \hat{\sigma}_C^2 \). The maximum likelihood estimate for the common p-p plot is thus

\[
\Phi \left( \frac{\hat{\sigma}_C - \hat{\mu}_C}{\hat{\sigma}_T} k_p - \frac{\hat{\mu}_T - \hat{\mu}_C}{\hat{\sigma}_T} \right).
\]

To determine the accuracy of this estimate of the common p-p plot we now compute its asymptotic distribution using Lemma 4.8. To use Lemma 4.8 we first need to calculate the asymptotic joint distribution of \( \hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}_T^2, \hat{\sigma}_C^2 \). From the general theory of maximum likelihood estimation we know that

\[
\begin{pmatrix}
\hat{\mu}_T - \mu_T \\
\hat{\mu}_C - \mu_C \\
\hat{\sigma}_T^2 - \sigma_T^2 \\
\hat{\sigma}_C^2 - \sigma_C^2 \\
\hat{a}_2 - a_2 \\
\hat{b}_2 - b_2 \\
\vdots \\
\hat{a}_K - a_K \\
\hat{b}_K - b_K
\end{pmatrix} \overset{\Delta}{=} \mathcal{N}(0, I^{-1}),
\]
where $I$ is the information matrix of the sample. $I$ can be expressed as

$$I = \begin{pmatrix} A & B' \\ B & C \end{pmatrix},$$

where

$$A = -\mathcal{E} \begin{pmatrix} \frac{\partial^2}{\partial \mu_T \partial \mu_T} & \frac{\partial^2}{\partial \mu_T \partial \mu_C} & \frac{\partial^2}{\partial \mu_T \partial \sigma_T^2} & \frac{\partial^2}{\partial \mu_T \partial \sigma_C^2} \\ \frac{\partial^2}{\partial \mu_C \partial \mu_T} & \frac{\partial^2}{\partial \mu_C \partial \mu_C} & \frac{\partial^2}{\partial \mu_C \partial \sigma_T^2} & \frac{\partial^2}{\partial \mu_C \partial \sigma_C^2} \\ \frac{\partial^2}{\partial \sigma_T^2 \partial \mu_T} & \frac{\partial^2}{\partial \sigma_T^2 \partial \mu_C} & \frac{\partial^2}{\partial \sigma_T^2 \partial \sigma_T^2} & \frac{\partial^2}{\partial \sigma_T^2 \partial \sigma_C^2} \\ \frac{\partial^2}{\partial \sigma_C^2 \partial \mu_T} & \frac{\partial^2}{\partial \sigma_C^2 \partial \mu_C} & \frac{\partial^2}{\partial \sigma_C^2 \partial \sigma_T^2} & \frac{\partial^2}{\partial \sigma_C^2 \partial \sigma_C^2} \end{pmatrix},$$

$$B = -\mathcal{E} \begin{pmatrix} \frac{\partial^2}{\partial \sigma_{a2} \partial \mu_T} & \frac{\partial^2}{\partial \sigma_{a2} \partial \mu_C} & \frac{\partial^2}{\partial \sigma_{a2} \partial \sigma_T^2} & \frac{\partial^2}{\partial \sigma_{a2} \partial \sigma_C^2} \\ \frac{\partial^2}{\partial \sigma_{a2} \partial \mu_T} & \frac{\partial^2}{\partial \sigma_{a2} \partial \mu_C} & \frac{\partial^2}{\partial \sigma_{a2} \partial \sigma_T^2} & \frac{\partial^2}{\partial \sigma_{a2} \partial \sigma_C^2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2}{\partial \sigma_{aK} \partial \mu_T} & \frac{\partial^2}{\partial \sigma_{aK} \partial \mu_C} & \frac{\partial^2}{\partial \sigma_{aK} \partial \sigma_T^2} & \frac{\partial^2}{\partial \sigma_{aK} \partial \sigma_C^2} \\ \frac{\partial^2}{\partial \sigma_{bK} \partial \mu_T} & \frac{\partial^2}{\partial \sigma_{bK} \partial \mu_C} & \frac{\partial^2}{\partial \sigma_{bK} \partial \sigma_T^2} & \frac{\partial^2}{\partial \sigma_{bK} \partial \sigma_C^2} \end{pmatrix},$$

and

$$C = \begin{pmatrix} C_2 & 0 & 0 & \ldots & 0 \\ 0 & C_2 & 0 & \ldots & 0 \\ 0 & 0 & C_4 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & C_K \end{pmatrix}$$

with

$$C_i = -\mathcal{E} \begin{pmatrix} \frac{\partial^2}{\partial a_i \partial a_i} & \frac{\partial^2}{\partial a_i \partial b_i} \\ \frac{\partial^2}{\partial b_i \partial a_i} & \frac{\partial^2}{\partial b_i \partial b_i} \end{pmatrix}.$$  

Inverting $I$ requires $O(K^3)$ algebraic computations so that as the number of studies
being considered together increases the feasibility of inverting the information matrix to
determine the asymptotic joint distribution of $\hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}_T^2, \hat{\sigma}_C^2$ rapidly decreases. Fortunately to compute the asymptotic joint distribution of $\hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}_T^2, \hat{\sigma}_C^2$ we need to know only a portion of $I^{-1}$. From the fact that

$$
\begin{pmatrix}
\hat{\mu}_T - \mu_T \\
\hat{\mu}_C - \mu_C \\
\hat{\sigma}_T^2 - \sigma_T^2 \\
\hat{\sigma}_C^2 - \sigma_C^2
\end{pmatrix} \sim \mathcal{N}(0, \Sigma)
$$

where

$$
\Sigma = (A - B'C^{-1}B)^{-1}.
$$

we have a mechanism for computing the portion of $I^{-1}$ that we need. Since $C^{-1}$ may be expressed as

$$
C^{-1} = \begin{pmatrix}
C_1^{-1} & 0 & 0 & \cdots & 0 \\
0 & C_2^{-1} & 0 & \cdots & 0 \\
0 & 0 & C_3^{-1} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & C_K^{-1}
\end{pmatrix}
$$

computing $C^{-1}$ requires only $O(K)$ algebraic computations and thus computing $\Sigma$ requires $O(K)$ operations instead of the $O(K^3)$ operations needed to invert the information matrix. Hence computing the asymptotic distribution of $\hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}_T^2, \hat{\sigma}_C^2$ is feasible for large $K$. To obtain the asymptotic distribution of the common p-p plot from the asymptotic distribution of $\hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}_T^2, \hat{\sigma}_C^2$, we just need to apply Lemma 4.8 or Lemma 4.11.

A useful supplement to this estimate of the common p-p plot is a test of the hypothesis that the p-p plot is constant across all $K$ studies. We develop such a test by employing the likelihood ratio test. Let $H_0$ denote the null hypothesis that all of the p-p plots are the same and $H_A$ denote the alternative that the p-p plots are heterogeneous. The likelihood ratio test for testing $H_0$ versus $H_A$ is to reject $H_0$ if

$$
-2 \log \left( \frac{\text{max}_{H_0} L}{\text{max}_{H_A} L} \right) = -2 \log(\text{max}_{H_0} L) + 2 \log(\text{max}_{H_A} L) \quad (6.1)
$$
Section 6.1: The Gaussian model

exceeds the $\alpha$’th quantile of a chi-square distribution with $2(K - 1)$ degrees of freedom. $L$ denotes the likelihood function so $\log(\max H_0 L)$ is simply the log likelihood function for the unrestricted Gaussian model

$$l = -\sum_{j=1}^{K} \left( n_{T,j} + n_{C,j} \right) \log(2\pi) \frac{2}{2} - \sum_{j=1}^{K} n_{T,j} \log(\sigma_{T,j}) - \sum_{j=1}^{K} n_{C,j} \log(\sigma_{C,j})$$

$$- \frac{1}{2} \sum_{j=1}^{K} \sum_{i=1}^{n_{T,j}} (X_{i,j}^T)^2 - \mu_{T,j} (\sum_{i=1}^{n_{T,j}} X_{i,j}) + \mu_{T,j}^2$$

$$- \frac{1}{2} \sum_{j=1}^{K} \sum_{i=1}^{n_{C,j}} (X_{i,j}^C)^2 - \mu_{C,j} (\sum_{i=1}^{n_{C,j}} X_{i,j}) + \mu_{C,j}^2$$

(6.2)

evaluated at the maximum likelihood estimates under $H_0$

$$\mu_{T,j} = \hat{\mu}_{T1} + \hat{\delta}_j,$$

$$\mu_{C,j} = \hat{\mu}_{C1} + \hat{\delta}_j,$$

$$\sigma_{T,j}^2 = \hat{\sigma}_{T1}^2,$$

$$\sigma_{C,j}^2 = \hat{\sigma}_{C1}^2,$$

whereas $\log(\max H_A L)$ is (6.2) evaluated at

$$\mu_{T,j} = \bar{X}_T^T,$$

$$\mu_{C,j} = \bar{X}_C^C,$$

$$\sigma_{T,j}^2 = \frac{1}{n_{T,j}} \sum_{i=1}^{n_{T,j}} (X_{i,j}^T - \mu_{T,j})^2,$$

$$\sigma_{C,j}^2 = \frac{1}{n_{C,j}} \sum_{i=1}^{n_{C,j}} (X_{i,j}^C - \mu_{C,j})^2.$$ 

(6.3)

6.1.2. Estimating ordered p-p plots

Here we consider the case where the treatment effect represented by the p-p plot is not constant across studies. First we develop an estimate of the p-p plot that facilitates comparisons of results across studies and then we construct a test of the assumptions that underlie this estimate.

Recall that if we want to make the statement that the treatment in the $i$’th study is relatively more effective than the treatment in $j$’th study it must be the case that the p-p
Chapter 6: Combined Estimates of P-p and Sp-p plots 102

plot for the \(i\)'th study lies strictly below the p-p plot for the \(j\)'th. That is

\[
\Phi \left( \frac{\sigma_{C_i}}{\sigma_{T_i}} k_p - \frac{\mu_{T_i} - \mu_{C_i}}{\sigma_{T_i}} \right) < \Phi \left( \frac{\sigma_{C_j}}{\sigma_{T_j}} k_p - \frac{\mu_{T_j} - \mu_{C_j}}{\sigma_{T_j}} \right) \quad \forall \ 0 < p < 1. \tag{6.4}
\]

Inequality (6.4) is equivalent to

\[
\frac{\sigma_{C_i}}{\sigma_{T_i}} = \frac{\sigma_{C_j}}{\sigma_{T_j}} \quad \text{and} \quad \frac{\mu_{T_i} - \mu_{C_i}}{\sigma_{T_i}} > \frac{\mu_{T_j} - \mu_{C_j}}{\sigma_{T_j}}. \tag{6.5}
\]

To see this first note that (6.5) clearly implies (6.4). If (6.4) is true, then

\[
\left( \frac{\sigma_{C_i}}{\sigma_{T_i}} - \frac{\sigma_{C_j}}{\sigma_{T_j}} \right) k_p < \frac{\mu_{T_i} - \mu_{C_i}}{\sigma_{T_i}} - \frac{\mu_{T_j} - \mu_{C_j}}{\sigma_{T_j}} \quad \forall \ 0 < p < 1.
\]

Since \(k_p\) ranges from \(-\infty\) to \(\infty\) as \(p\) ranges from 0 to 1 it follows that

\[
\frac{\sigma_{C_i}}{\sigma_{T_i}} = \frac{\sigma_{C_j}}{\sigma_{T_j}}
\]

and hence that

\[
\frac{\mu_{T_i} - \mu_{C_i}}{\sigma_{T_i}} > \frac{\mu_{T_j} - \mu_{C_j}}{\sigma_{T_j}}.
\]

Thus two and hence \(K\) p-p plots can be ordered if and only if all the p-p plots have the same variance ratio.

The assumption of a common variance ratio for all the p-p plots can be incorporated into the Gaussian model and this model can be fitted using the method of maximum likelihood. The estimates for the individual p-p plots resulting from this model are such that every p-p plot lies strictly above or strictly below the others. Hence comparisons of treatment effects across studies are simplified with these estimates.

The model with the restriction of a common variance ratio assumes that the \(i\)'th observation from the treatment or control distribution of the \(j\)'th study, \(X^T_{i,j}\) or \(X^C_{i,j}\), is distributed as

\[
X^T_{i,j} = \mathcal{N}(\mu_{T,j}, k \sigma_{C,j}), \quad i = 1, \ldots, n_{T_j}, \quad j = 1, \ldots, K,
\]

\[
X^C_{i,j} = \mathcal{N}(\mu_{C,j}, \sigma_{C,j}), \quad i = 1, \ldots, n_{C_j}, \quad j = 1, \ldots, K.
\]

To fit the model using the method of maximum likelihood we need to maximize the log likelihood function with respect to the unknown parameters. The log likelihood function
for this model is

\[ l = - \sum_{j=1}^{K} (n_{T,j} + n_{C,j}) \frac{\log(2\pi)}{2} - \sum_{j=1}^{K} (n_{T,j} + n_{C,j}) \log(\sigma_{C,j}) - \sum_{j=1}^{K} n_{T,j} \log(k) \]

\[ - \frac{1}{2} \sum_{j=1}^{K} \frac{n_{T,j} (X_{i,j}^T)^2}{k \sigma_{C,j}^2} - 2\mu_{T,j} \sum_{i=1}^{n_{T,j}} X_{i,j}^T + n_{T,j} \mu_{T,j}^2 \]

\[ - \frac{1}{2} \sum_{j=1}^{K} \frac{n_{C,j} (X_{i,j}^C)^2}{\sigma_{C,j}^2} - 2\mu_{C,j} \sum_{i=1}^{n_{C,j}} X_{i,j}^C + n_{C,j} \mu_{C,j}^2. \]

The maximum of the log likelihood function is found by setting the following derivatives equal to zero and solving for the unknown parameters.

\[ \frac{\partial l}{\partial \mu_{T,j}} = \frac{\sum_{i=1}^{n_{T,j}} X_{i,j}^T - n_{T,j} \mu_{T,j}}{(k \sigma_{C,j})^2}, \]

\[ \frac{\partial l}{\partial \mu_{C,j}} = \frac{\sum_{i=1}^{n_{C,j}} X_{i,j}^C - n_{C,j} \mu_{C,j}}{\sigma_{C,j}^2}, \]

\[ \frac{\partial l}{\partial \sigma_{C,j}^2} = -\frac{n_{T,j} + n_{C,j}}{2 \sigma_{C,j}^2} + \frac{\sum_{i=1}^{n_{T,j}} (X_{i,j}^T)^2 - 2\mu_{T,j} \sum_{i=1}^{n_{T,j}} X_{i,j}^T + n_{T,j} \mu_{T,j}^2}{2 k^2 \sigma_{C,j}^4}
\]

\[ + \frac{\sum_{i=1}^{n_{C,j}} (X_{i,j}^C)^2 - 2\mu_{C,j} \sum_{i=1}^{n_{C,j}} X_{i,j}^C + n_{C,j} \mu_{C,j}^2}{2 \sigma_{C,j}^4}, \]

\[ \frac{\partial l}{\partial k} = -\frac{1}{k} \sum_{j=1}^{K} n_{T,j} + \sum_{j=1}^{K} \frac{\sum_{i=1}^{n_{T,j}} (X_{i,j}^T)^2 - 2\mu_{T,j} \sum_{i=1}^{n_{T,j}} X_{i,j}^T + n_{T,j} \mu_{T,j}^2}{k^3 \sigma_{C,j}^2}. \]

Thus

\[ \mu_{T,j} = X_j^T \]

\[ \mu_{C,j} = X_j^C \]

\[ \sigma_{C,j}^2 = \frac{1}{k^2 (n_{T,j} + n_{C,j})} \left[ \sum_{i=1}^{n_{T,j}} (X_{i,j}^T)^2 - 2\mu_{T,j} \sum_{i=1}^{n_{T,j}} X_{i,j}^T + n_{T,j} \mu_{T,j}^2 \right] \]

\[ + \frac{1}{n_{T,j} + n_{C,j}} \left[ \sum_{i=1}^{n_{C,j}} (X_{i,j}^C)^2 - 2\mu_{C,j} \sum_{i=1}^{n_{C,j}} X_{i,j}^C + n_{C,j} \mu_{C,j}^2 \right]. \]
\[ k^2 = \sum_{j=1}^{K} \left( \sum_{i=1}^{n_{T,j}} (X_{i,j}^T)^2 - 2\mu_{T,j} \sum_{i=1}^{n_{T,j}} X_{i,j}^T + n_{T,j} \mu_{T,j}^2 \right) / \sum_{j=1}^{K} n_{T,j}. \]

Substituting the expression for \( \sigma_{C,j}^2 \) into the expression for \( k^2 \) yields a second degree polynomial in \( k^2 \) which can be solved for numerically. The solution for \( k \) can then be substituted into the equation for \( \sigma_{C,j}^2 \) to obtain the maximum likelihood estimate of \( \sigma_{C,j}^2 \).

Denote the resulting maximum likelihood estimates for \( k, \mu_{T,j}, \mu_{C,j} \) and \( \sigma_{C,j}^2 \) in this model by \( \hat{k}, \hat{\mu}_{T,j}, \hat{\mu}_{C,j} \) and \( \hat{\sigma}_{C,j}^2 \). The maximum likelihood estimate of the p-p plot for the \( j \)th study is thus

\[
\Phi \left( \frac{\hat{k}_p}{k} - \frac{\hat{\mu}_{T,j} - \hat{\mu}_{C,j}}{k \hat{\sigma}_{C,j}} \right).
\]

To determine the accuracy of this estimate of the p-p plot for the \( j \)th study we deduce its asymptotic distribution using Lemma 4.8. To utilize Lemma 4.8 we first need to determine the asymptotic joint distribution of \( \hat{\mu}_{T,j}, \hat{\mu}_{C,j}, \hat{k}^2 \hat{\sigma}_{C,j}^2 \) and \( \hat{\sigma}_{C,j}^2 \). From the general theory of maximum likelihood estimation we know that

\[
\begin{pmatrix}
\hat{k} - k \\
\hat{\mu}_{T,j} - \mu_{T,j} \\
\hat{\mu}_{C,j} - \mu_{C,j} \\
\hat{\sigma}_{C,j}^2 - \sigma_{C,j}^2 \\
\hat{\mu}_{T1} - \mu_{T1} \\
\hat{\mu}_{C1} - \mu_{C1} \\
\hat{\sigma}_{C1}^2 - \sigma_{C1}^2 \\
\vdots \\
\hat{\mu}_{TK} - \mu_{TK} \\
\hat{\mu}_{CK} - \mu_{CK} \\
\hat{\sigma}_{CK}^2 - \sigma_{CK}^2
\end{pmatrix} \\sim \mathcal{N}(0, \mathcal{I}^{-1}),
\]
where the information matrix $\mathcal{I}$ is

$$
\mathcal{I} = \begin{pmatrix} A & B' \\ B & C \end{pmatrix},
$$

where

$$
A = -\mathcal{E} \begin{pmatrix}
\frac{\partial^2 l}{\partial \mu_1 \partial \theta_k} & \frac{\partial^2 l}{\partial \mu_2 \partial \mu_3} & \frac{\partial^2 l}{\partial \mu_2 \partial \mu_C} & \frac{\partial^2 l}{\partial \mu_2 \partial \sigma_C^2} \\
\frac{\partial^2 l}{\partial \mu_1 \partial \mu_3} & \frac{\partial^2 l}{\partial \mu_1 \partial \mu_C} & \frac{\partial^2 l}{\partial \mu_1 \partial \sigma_C^2} \\
\frac{\partial^2 l}{\partial \mu_1 \partial \sigma_C^2} & \frac{\partial^2 l}{\partial \mu_3 \partial \mu_C} & \frac{\partial^2 l}{\partial \mu_3 \partial \sigma_C^2} \\
\frac{\partial^2 l}{\partial \sigma_C^2 \partial \sigma_C^2} & \frac{\partial^2 l}{\partial \sigma_C^2 \partial \mu_C} & \frac{\partial^2 l}{\partial \sigma_C^2 \partial \sigma_C^2}
\end{pmatrix},
$$

$$
B = -\mathcal{E} \begin{pmatrix}
\frac{\partial^2 l}{\partial \mu_1 \partial \theta_k} & 0 & 0 & 0 \\
0 & \frac{\partial^2 l}{\partial \mu_3 \partial \mu_C} & 0 & 0 \\
0 & 0 & \frac{\partial^2 l}{\partial \sigma_C^2 \partial \mu_C} & 0 \\
\vdots & \vdots & \vdots & \vdots \\
\frac{\partial^2 l}{\partial \mu_1 \partial \theta_k} & 0 & 0 & 0 \\
0 & \frac{\partial^2 l}{\partial \mu_3 \partial \mu_C} & 0 & 0 \\
0 & 0 & \frac{\partial^2 l}{\partial \sigma_C^2 \partial \mu_C} & 0 \\
0 & 0 & 0 & \frac{\partial^2 l}{\partial \sigma_C^2 \partial \sigma_C^2}
\end{pmatrix},
$$

$$
C = \begin{pmatrix} C_1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\
0 & C_2 & \ldots & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & C_{j-1} & 0 & \ldots & 0 \\
0 & 0 & \ldots & 0 & C_{j+1} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & 0 & \ldots & C_K
\end{pmatrix},
$$

with
Chapter 6: Combined Estimates of P-p and Sp-p plots 106

\[ C_i = -E \begin{pmatrix} \frac{\partial^2}{\partial \mu_T \partial \mu_T} & \frac{\partial^2}{\partial \mu_T \partial \mu_C} & \frac{\partial^2}{\partial \mu_T \partial \sigma^2_{C}} \\ \frac{\partial^2}{\partial \mu_C \partial \mu_T} & \frac{\partial^2}{\partial \mu_C \partial \mu_C} & \frac{\partial^2}{\partial \mu_C \partial \sigma^2_{C}} \\ \frac{\partial^2}{\partial \sigma^2_{C} \partial \mu_T} & \frac{\partial^2}{\partial \sigma^2_{C} \partial \mu_C} & \frac{\partial^2}{\partial \sigma^2_{C} \partial \sigma^2_{C}} \end{pmatrix} \]

Inverting \( \mathcal{I} \) requires \( O(K^3) \) algebraic computations so that as the number of studies being considered together increases the feasibility of inverting the information matrix to determine the asymptotic distribution of \( \hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}^2_T, \) and \( \hat{k} \) decreases. Fortunately the asymptotic joint distribution of \( \hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}^2_T, \) and \( \hat{k} \) depends on only a portion of the inverse of the information matrix.

As for the computation of the portion of \( I^{-1} \) that is required note that

\[
\begin{pmatrix} \hat{k} \\ \hat{\mu}_T - \mu_T \\ \hat{\mu}_C - \mu_C \\ \hat{\sigma}_C^2 - \sigma_C^2 \end{pmatrix} \approx \mathcal{N}(0, \Sigma)
\]

where

\[
\Sigma = (A - B'C^{-1}B)^{-1}.
\]

Since inverting \( C \) requires only \( O(K) \) algebraic computations, computing \( \Sigma^{-1} \) requires \( O(K) \) computations instead of the \( O(K^3) \) operations required to invert the whole information matrix. Thus computing the asymptotic joint distribution of

\( \hat{k}, \hat{\mu}_T, \hat{\mu}_C, \hat{\sigma}_C^2 \)

is feasible for large \( K \). Finally using the delta method to compute the asymptotic joint distribution of

\( \hat{\mu}_T, \hat{\mu}_C, \hat{k}^2 \hat{\sigma}_C^2, \hat{\sigma}_C^2 \)

we have

\[
\begin{pmatrix} \hat{\mu}_T - \mu_T \\ \hat{\mu}_C - \mu_C \\ \hat{k}^2 \hat{\sigma}_C^2 - k^2 \sigma_C^2 \\ \hat{\sigma}_C^2 - \sigma_C^2 \end{pmatrix} \approx \mathcal{N}(0, \hat{\Sigma})
\]
where

\[ \hat{\Sigma} = A \Sigma A' \]

and

\[
A = \begin{pmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
2k \sigma^2_{C,j} & 0 & 0 & k^2 \\
0 & 0 & 0 & 1
\end{pmatrix}.
\]

To compute the asymptotic distribution of the p-p plot from the asymptotic distribution of

\[ \hat{\mu}_{T,j}, \hat{\mu}_{C,j}, \hat{k^2} \sigma^2_{C,j}, \hat{\sigma}^2_{C,j} \]

apply Lemma 4.8 or Lemma 4.11.

Having assumed that the variance ratios are the same to facilitate the comparison of treatment effects across studies, it is of interest to test whether this hypothesis is reasonable. Once again we employ the likelihood ratio test to develop this test. Let \( H_0 \) denote the null hypothesis that the variance ratio is constant across studies and \( H_A \) denote the alternative hypothesis that there is no restriction on the variance ratios. Then the likelihood ratio test rejects \( H_0 \) at significance level \( \alpha \) if

\[
-2 \log \left( \frac{\max_{H_0} L}{\max_{H_A} L} \right) = -2 \log (\max_{H_0} L) + 2 \log (\max_{H_A} L)
\]

exceeds the \( \alpha \)'th quantile of a chi-square distribution with \( K - 1 \) degrees of freedom. \( \log(\max_{H_0} L) \) is the unrestricted log likelihood function (6.2) evaluated at

\[
\mu_{T,j} = \hat{\mu}_{T,j} \\
\mu_{C,j} = \hat{\mu}_{C,j} \\
\sigma^2_{T,j} = \hat{k^2} \sigma^2_{C,j} \\
\sigma^2_{C,j} = \hat{\sigma}^2_{C,j}
\]

wheras \( \log(\max_{H_A} L) \) is as before the unrestricted log likelihood function (6.2) evaluated as in (6.3).
6.1.3. An example

Now we illustrate the use of combined estimates of the p-p plot in the Gaussian setting by forming a pooled estimate of the p-p plot for the open education data set. Figure 6.1 displays the combined estimate of the p-p plot along with a 90 percent simultaneous confidence band. Since the confidence band does not include the 45 degree line we are 90 percent confident that the treatment and control distributions are different. What is more since the estimate of the p-p plot lies below the 45 degree line for control percentiles $p < .5$ and above the 45 degree line for $p > .5$ the treatment appears to be beneficial in the range $[0, .5]$ and detrimental in the range $[.5, 1]$. That is the treatment appears to hinder the performance of good students and enhance the performance of poor students. This interpretation of the data stands in contrast to an effect size analysis which indicates that the treatment has no effect.
Section 6.1: The Gaussian model

Figure 6.1

The pooled estimate of the p-p plot for the open education data set along with a 90 percent simultaneous confidence band.
6.2. The gamma model.

The second model we look at in this chapter is the Gamma model. Suppose once again that $K$ studies are considered together and that for the $j$'th study the treatment and control distributions are defined as

$$F_{T_j}(x) = \int_0^x \frac{\lambda_{T_j} r_{T_j}}{\Gamma(r_{T_j})} t^{r_{T_j}} \exp(-\lambda_{T_j} t) \, dt$$

$$F_{C_j}(x) = \int_0^x \frac{\lambda_{C_j} r_{C_j}}{\Gamma(r_{C_j})} t^{r_{C_j}} \exp(-\lambda_{C_j} t) \, dt.$$  

The sp-p plot for the $j$'th study is thus

$$\bar{P}_{T_j}(F_{C_j}^{-1}(p)) = \bar{G}_{r_{T_j}} \left( \frac{\lambda_{T_j}}{\lambda_{C_j}} G_{r_{C_j}}^{-1}(p) \right).$$

where

$$G_r(t) = \int_0^t \frac{1}{\Gamma(r)} x^{r-1} \exp(-x) \, dx$$

6.2.1. Estimating a common sp-p plot

In this section we assume that the treatment effects associated with the $K$ different studies can be represented by a common sp-p plot. In this situation an estimate of the common sp-p plot based on the information in all of the studies is more accurate than an estimate of the sp-p plot based on any one of the individual studies. Here we first construct an estimate of the common sp-p plot based on all $K$ studies using the method of maximum likelihood. We then go on to construct a test of the hypothesis that the sp-p plot is constant across the $K$ studies under consideration.

Recall from our considerations of invariance that in the gamma model two sp-p plots corresponding to gamma distributions with the same shape parameters, $r_T, r_C$, are identical if there exists a linear transformation of the form $x \to ax$, which when applied to the measurement scale of one study transforms the treatment and control distributions of that study to the distributions of the other. Thus the p-p plots for $K$ studies are identical if $r_T$ and $r_C$ are constant across studies and if there exists constants $a_j : j = 2, \ldots, K$
such that

\[ a_j \lambda_{T,j} = \lambda_{T,j}, \]
\[ a_j \lambda_{C,j} = \lambda_{C,j}. \]

By incorporating these conditions into the Gamma model presented at the beginning of this section we can construct an estimate of the common sp-p plot.

The resulting model states that the \( i \)'th observation in the treatment or control group of the \( j \)'th study, \( X_{ij}^T \) or \( X_{ij}^C \), is distributed as

\[ X_{ij}^T = \text{Gamma}(a_j \lambda_{T,j}, r_T), \quad i = 1, \ldots, n_{T,j}, \quad j = 1, \ldots, K; \]
\[ X_{ij}^C = \text{Gamma}(a_j \lambda_{C,j}, r_C), \quad i = 1, \ldots, n_{C,j}, \quad j = 1, \ldots, K, \]

where \( a_1 = 1 \) and the remaining parameters are unknown. To fit this model we use the method of maximum likelihood. This technique involves maximizing the log likelihood function with respect to the unknown parameters. The log likelihood for this model is,

\[
\ell = \sum_{j=1}^{K} \left[ n_{T,j} r_T \log(a_j) + n_{T,j} r_T \log(\lambda_T) - n_{T,j} \log\Gamma(r_T) \right] \\
+ \sum_{j=1}^{K} \left[ n_{C,j} r_C \log(a_j) + n_{C,j} r_C \log(\lambda_C) - n_{C,j} \log\Gamma(r_C) \right] \\
+ \sum_{j=1}^{K} \left[ (r_T - 1) \log(\Pi_{i=1}^{n_{T,j}} X_{ij}^T) - a_j \lambda_T \sum_{i=1}^{n_{T,j}} X_{ij}^T \right] \\
+ \sum_{j=1}^{K} \left[ (r_C - 1) \log(\Pi_{i=1}^{n_{C,j}} X_{ij}^C) - a_j \lambda_C \sum_{i=1}^{n_{C,j}} X_{ij}^C \right].
\]

The maximum of the log likelihood is obtained by setting the first partials equal to zero and solving for the unknown parameters. The first partials are

\[
\frac{\partial \ell}{\partial r_T} = \sum_{j=1}^{K} \left[ \log(\Pi_{i=1}^{n_{T,j}} X_{ij}^T) + n_{T,j} \log(a_j) + n_{T,j} \log(\lambda_T) - n_{T,j} \frac{\partial}{\partial r_T} \log\Gamma(r_T) \right],
\]

\[
\frac{\partial \ell}{\partial r_C} = \sum_{j=1}^{K} \left[ \log(\Pi_{i=1}^{n_{C,j}} X_{ij}^C) + n_{C,j} \log(a_j) + n_{C,j} \log(\lambda_C) - n_{C,j} \frac{\partial}{\partial r_C} \log\Gamma(r_C) \right],
\]

\[
\frac{\partial \ell}{\partial \lambda_T} = \sum_{j=1}^{K} \left( \frac{n_{T,j} r_T}{\lambda_T} - a_j \sum_{i=1}^{n_{T,j}} X_{ij}^T \right),
\]

\[
\frac{\partial \ell}{\partial \lambda_C} = \sum_{j=1}^{K} \left( \frac{n_{C,j} r_C}{\lambda_C} - a_j \sum_{i=1}^{n_{C,j}} X_{ij}^C \right).
\]
\[
\frac{\partial l}{\partial \lambda_C} = \sum_{j=1}^{K} \left( \frac{n_{C,j} r_C}{\lambda_C} - a_j \sum_{i=1}^{n_{C,j}} X_{ij}^C \right), \\
\frac{\partial l}{\partial a_j} = \frac{n_{T,j} r_T}{a_j} - \lambda_T \sum_{i=1}^{n_{T,j}} X_{ij}^T + \frac{n_{C,j} r_C}{a_j} - \lambda_C \sum_{i=1}^{n_{C,j}} X_{ij}^C.
\]

Note that setting \(\partial l/\partial a_j = 0\) allows us to solve for \(a_j\) in terms of \(r_T, r_C, \lambda_T, \lambda_C\). Specifically

\[
a_j = \frac{n_{T,j} r_T + n_{C,j} r_C}{\lambda_T \sum_{i=1}^{n_{T,j}} X_{ij}^T + \lambda_C \sum_{i=1}^{n_{C,j}} X_{ij}^C}.
\]

Thus we can reduce this system of equations to a system consisting of four equations and four unknowns, namely \(r_T, r_C, \lambda_T, \lambda_C\). This can be solved using the Newton-Raphson Algorithm.

Denote the resulting maximum likelihood estimates for \(r_T, r_C, \lambda_T, \lambda_C\) by \(\hat{r}_T, \hat{r}_C, \hat{\lambda}_T, \hat{\lambda}_C\). The maximum likelihood estimate for the common sp-p plot is thus

\[
\tilde{G}_{\hat{\lambda_T},1} \left( \frac{\hat{\lambda}_T}{\hat{\lambda}_C} G_{\hat{r}_C,1}^{-1}(p) \right).
\]

To determine the accuracy of this estimate of the common sp-p plot we compute its asymptotic distribution using Lemma 5.9. To use Lemma 5.9 we need the asymptotic joint distribution of \(\hat{\lambda}_T, \hat{\lambda}_C, \hat{r}_T\) and \(\hat{r}_C\). From the general theory of maximum likelihood estimates we know that

\[
\begin{pmatrix}
\hat{\lambda}_T - \lambda_T \\
\hat{\lambda}_C - \lambda_C \\
\hat{r}_T - r_T \\
\hat{r}_C - r_C \\
\hat{a}_2 - a_2 \\
\vdots \\
\hat{a}_K - a_K
\end{pmatrix} \overset{d}{\approx} \mathcal{N}(0, I^{-1}),
\]
where the information matrix, $\mathcal{I}$, is

$$
\mathcal{I} = \begin{pmatrix} A & B' \\ B & C \end{pmatrix}
$$

where

$$
A = -\mathcal{E} \begin{pmatrix} \frac{\partial^2}{\partial \lambda T_1 \partial \lambda T_1} & \frac{\partial^2}{\partial \lambda T_1 \partial \lambda C_1} & \frac{\partial^2}{\partial \lambda T_1 \partial \check{\tau}_T} & \frac{\partial^2}{\partial \lambda T_1 \partial \check{\tau}_C} \\ \frac{\partial^2}{\partial \lambda C_1 \partial \lambda T_1} & \frac{\partial^2}{\partial \lambda C_1 \partial \lambda C_1} & \frac{\partial^2}{\partial \lambda C_1 \partial \check{\tau}_T} & \frac{\partial^2}{\partial \lambda C_1 \partial \check{\tau}_C} \\ \frac{\partial^2}{\partial \check{\tau}_T \partial \lambda T_1} & \frac{\partial^2}{\partial \check{\tau}_T \partial \lambda C_1} & \frac{\partial^2}{\partial \check{\tau}_T \partial \check{\tau}_T} & \frac{\partial^2}{\partial \check{\tau}_T \partial \check{\tau}_C} \\ \frac{\partial^2}{\partial \check{\tau}_C \partial \lambda T_1} & \frac{\partial^2}{\partial \check{\tau}_C \partial \lambda C_1} & \frac{\partial^2}{\partial \check{\tau}_C \partial \check{\tau}_T} & \frac{\partial^2}{\partial \check{\tau}_C \partial \check{\tau}_C} \end{pmatrix},
$$

$$
B = -\mathcal{E} \begin{pmatrix} \frac{\partial^2}{\partial a_2 \partial \lambda T_1} & \frac{\partial^2}{\partial a_2 \partial \lambda C_1} & \frac{\partial^2}{\partial a_2 \partial \check{\tau}_T} & \frac{\partial^2}{\partial a_2 \partial \check{\tau}_C} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2}{\partial a_K \partial \lambda T_1} & \frac{\partial^2}{\partial a_K \partial \lambda C_1} & \frac{\partial^2}{\partial a_K \partial \check{\tau}_T} & \frac{\partial^2}{\partial a_K \partial \check{\tau}_C} \end{pmatrix},
$$

$$
C = -\mathcal{E} \begin{pmatrix} \frac{\partial^2}{\partial (a_2)^2} & 0 & 0 & \ldots & 0 \\ 0 & \frac{\partial^2}{\partial (a_3)^2} & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & \vdots \\ 0 & 0 & \ldots & \frac{\partial^2}{\partial (a_K)^2} \end{pmatrix}.
$$

Inverting $\mathcal{I}$ requires $O(K^3)$ algebraic computations so that as the number of studies being combined increases the feasibility of inverting the information matrix to determine the asymptotic joint distribution of $\dot{\lambda}_T$, $\dot{\lambda}_C$, $\check{\tau}_T$ and $\check{\tau}_C$ rapidly decreases. Fortunately to estimate the asymptotic joint distribution of $\dot{\lambda}_T$, $\dot{\lambda}_C$, $\check{\tau}_T$ and $\check{\tau}_C$ we need to know only a portion of $\mathcal{I}^{-1}$. 
As for the computation of the portion of $I^{-1}$ that is required note that

$$
\begin{pmatrix}
\hat{\lambda}_{T1} - \lambda_{T1} \\
\hat{\lambda}_{C1} - \lambda_{C1} \\
\hat{r}_T - r_T \\
\hat{r}_C - r_C
\end{pmatrix} \overset{\mathcal{D}}{=} \mathcal{N}(0, \Sigma)
$$

where

$$
\Sigma = (A - B'C^{-1}B)^{-1}.
$$

Inverting $C$ requires only $O(K)$ algebraic computations. Thus the computation of the asymptotic joint distribution of $\hat{\lambda}_{T1}$, $\hat{\lambda}_{C1}$, $\hat{r}_T$ and $\hat{r}_C$ is feasible for large $K$. To obtain the asymptotic distribution of the sp-p plot we just need to apply Lemma 5.9 or Lemma 5.11.

Having assumed that all $K$ sp-p plots are the same, a test of this assumption is of special interest. Once again to derive this test we employ the likelihood ratio test. Let $H_O$ denote the null hypothesis that the sp-p plots are the same across studies and let $H_A$ denote the alternative hypothesis that the sp-p plots are heterogeneous. Then the likelihood ratio test rejects $H_O$ at significance level $\alpha$ if

$$
-2 \log \left( \frac{\max_{H_O} L}{\max_{H_A} L} \right) = -2 \log(\max_{H_O} L) + 2 \log(\max_{H_A} L)
$$

exceeds the $\alpha$'th quantile of a chi-square distribution with $3(K - 1)$ degrees of freedom. $\log(\max_{H_O} L)$ is the log likelihood function for the unrestricted gamma model

$$
l = \sum_{j=1}^{K} \left( n_{Tj} r_{Tj} \log(a_j) + n_{Tj} r_{Tj} \log(\lambda_{Tj}) - n_{Tj} \log(\Gamma(r_{Tj})) \right)
$$

$$
+ \sum_{j=1}^{K} \left( (r_{Tj} - 1) \log(\Pi_{i=1}^{n_{Tj}} x_{ij}^T) - a_{Tj} \lambda_{Tj} \sum_{i=1}^{n_{Tj}} x_{ij}^T \right)
$$

$$
+ \sum_{j=1}^{K} \left( n_{Cj} r_{Cj} \log(a_j) + n_{Cj} r_{Cj} \log(\lambda_{Cj}) - n_{Cj} \log(\Gamma(r_{Cj})) \right)
$$

$$
+ \sum_{j=1}^{K} \left( (r_{Cj} - 1) \log(\Pi_{i=1}^{n_{Cj}} x_{ij}^C) - a_{Cj} \lambda_{Cj} \sum_{i=1}^{n_{Cj}} x_{ij}^C \right),
$$
evaluated at
\[ \tau_{T_j} = \hat{\tau}_T, \quad \tau_{C_j} = \hat{\tau}_C, \]
\[ \lambda_{T_j} = \hat{\lambda}_T, \quad \lambda_{C_j} = \hat{\lambda}_C, \]
while \( \log(\max_{H_A} L) \) is (6.6) evaluated at
\[ \tau_{T_j} = \hat{\tau}_{T_j}, \quad \tau_{C_j} = \hat{\tau}_{C_j}, \]
\[ \lambda_{T_j} = \hat{\lambda}_{T_j}, \quad \lambda_{C_j} = \hat{\lambda}_{C_j}, \]
where \( \hat{\tau}_{T_j}, \hat{\tau}_{C_j}, \hat{\lambda}_{T_j} \) and \( \hat{\lambda}_{C_j} \) solve the following equations.
\[
\begin{align*}
n_{T_j} \log(\lambda_{T_j}) + \log(\prod_{i=1}^{n_{T_j}} X_{i_j}^T) - \frac{\partial}{\partial \tau_{T_j}} \log(\Gamma(\tau_{T_j})) &= 0, \\
n_{C_j} \log(\lambda_{C_j}) + \log(\prod_{i=1}^{n_{C_j}} X_{i_j}^C) - \frac{\partial}{\partial \tau_{C_j}} \log(\Gamma(\tau_{C_j})) &= 0,
\end{align*}
\]
\[
\begin{align*}
\frac{n_{T_j} \tau_{T_j}}{\lambda_{T_j}} - \sum_{i=1}^{n_{T_j}} X_{i_j}^T &= 0, \\
\frac{n_{C_j} \tau_{C_j}}{\lambda_{C_j}} - \sum_{i=1}^{n_{C_j}} X_{i_j}^C &= 0.
\end{align*}
\]

6.2.2. Estimating ordered sp-p plots

In this section we develop an estimate of the sp-p plot that facilitates the comparison of results across studies. Recall that if we wish to make the statement that the treatment in the \( i \)'th study is superior to the treatment in the \( j \)'th study, then necessarily the sp-p plot for the \( i \)'th study must lie above the sp-p plot for the \( j \)'th. The \( i \)'th sp-p plot lies above the \( j \)'th sp-p plot if
\[
\tilde{G}_{\tau_{T}, \left( \frac{\lambda_{T_i} \tau_{T_i}}{\lambda_{C_i} \tau_{C_i}} \right)}(p) > \tilde{G}_{\tau_{T,j}, \left( \frac{\lambda_{T_j} \tau_{T_j}}{\lambda_{C_j} \tau_{C_j}} \right)}(p) \quad \forall \quad 0 < p < 1.
\]
If \( \tau_{T_i} = \tau_{T_j}, \tau_{C_i} = \tau_{C_j} \) and
\[ \frac{\lambda_{T_i}}{\lambda_{C_i}} < \frac{\lambda_{T_j}}{\lambda_{C_j}} \]
then the \( i \)'th sp-p plot will lie above the \( j \)'th sp-p plot for all \( 0 < p < 1 \). Thus comparisons of sp-p plots are greatly facilitated if \( \tau_T \) and \( \tau_C \) are the same across studies. The condition of constant shape parameters \( \tau_T \) and \( \tau_C \) across studies can be incorporated into a model, fitted and tested using the method of maximum likelihood.
The resulting model which facilitates the comparison of results across studies states that the \(i\)'th observation from the treatment or control distribution of the \(j\)'th study, \(X_{ij}^T\) or \(X_{ij}^C\), is distributed as

\[
X_{ij}^T = \text{Gamma}(\lambda_{T, j}, r_T), \quad i = 1, \ldots, n_{T, j}, \quad j = 1, \ldots, K,
\]

\[
X_{ij}^C = \text{Gamma}(\lambda_{C, j}, r_C), \quad i = 1, \ldots, n_{C, j}, \quad j = 1, \ldots, K.
\]

To fit this model to the data we utilize the method of maximum likelihood. This technique involves maximizing the log likelihood function with respect to the unknown parameters. The log likelihood function for this model is

\[
l = \sum_{j=1}^{K} \left( (r_T - 1) \log \left( \prod_{i=1}^{n_{T, j}} X_{ij}^T \right) - \lambda_{T, j} \sum_{i=1}^{n_{T, j}} X_{ij}^T + n_{T, j} r_T \log (\lambda_{T, j}) - n_{T, j} \log \Gamma (r_T) \right) \\
+ \sum_{j=1}^{K} \left( (r_C - 1) \log \left( \prod_{i=1}^{n_{C, j}} X_{ij}^C \right) - \lambda_{C, j} \sum_{i=1}^{n_{C, j}} X_{ij}^C + n_{C, j} r_C \log (\lambda_{C, j}) - n_{C, j} \log \Gamma (r_C) \right).
\]

Taking derivatives and setting them equal to zero yields the following equations for the maximum likelihood estimates.

\[
\sum_{j=1}^{K} \left( \log \left( \prod_{i=1}^{n_{T, j}} X_{ij}^T \right) + n_{T, j} \log (\lambda_{T, j}) - n_{T, j} \frac{\partial}{\partial r_T} \log \Gamma (r_T) \right) = 0, \quad (6.8)
\]

\[
\sum_{j=1}^{K} \left( \log \left( \prod_{i=1}^{n_{C, j}} X_{ij}^C \right) + n_{C, j} \log (\lambda_{C, j}) - n_{C, j} \frac{\partial}{\partial r_C} \log \Gamma (r_C) \right) = 0, \quad (6.9)
\]

\[
- \sum_{i=1}^{n_{T, j}} X_{ij}^T + \frac{n_{T, j} r_T}{\lambda_{T, j}} = 0, \quad (6.10)
\]

\[
- \sum_{i=1}^{n_{C, j}} X_{ij}^C + \frac{n_{C, j} r_C}{\lambda_{C, j}} = 0. \quad (6.11)
\]

Equations (6.10) and (6.11) imply that

\[
\lambda_{T, j} = \frac{r_T}{n_{T, j}} \sum_{i=1}^{n_{T, j}} X_{ij}^T,
\]

\[
\lambda_{C, j} = \frac{r_C}{n_{C, j}} \sum_{i=1}^{n_{C, j}} X_{ij}^C.
\]
Substituting the solutions for $\lambda_{Tj}$ and $\lambda_{Cj}$ into (6.8) and (6.9) yields

$$\log \left( \frac{\prod_{j=1}^{K} \sum_{i=1}^{n_{Tj}} X_{ij}^{T} / n_{Tj}}{\prod_{j=1}^{K} \left( \prod_{i=1}^{n_{Tj}} X_{ij}^{T} \right)^{1/n_{Tj}}} \right) = K \log (r_{T}) - K \frac{\partial}{\partial r_{T}} \log \Gamma (r_{T}),$$

$$\log \left( \frac{\prod_{j=1}^{K} \sum_{i=1}^{n_{Cj}} X_{ij}^{C} / n_{Cj}}{\prod_{j=1}^{K} \left( \prod_{i=1}^{n_{Cj}} X_{ij}^{C} \right)^{1/n_{Cj}}} \right) = K \log (r_{C}) - K \frac{\partial}{\partial r_{C}} \log \Gamma (r_{C}).$$

These can be solved for $r_{T}$ and $r_{C}$ numerically and the solutions can be used to compute $\lambda_{Tj}$ and $\lambda_{Cj}$.

Denote the maximum likelihood estimates of $r_{T}, r_{C}, \lambda_{Tj}, \lambda_{Cj}$ obtained above by $\hat{r}_{T}, \hat{r}_{C}, \hat{\lambda}_{Tj}, \hat{\lambda}_{Cj}$. Then the maximum likelihood estimate for the sp-p plot corresponding to the $j$'th study is given by

$$\tilde{G}_{j, 1} \left( \frac{\hat{\lambda}_{Tj}}{\hat{\lambda}_{Cj}} G_{j, 1}^{-1}(p) \right).$$

To determine the accuracy of the estimate of the p-p plot for the $j$'th study we now calculate its asymptotic distribution using Lemma 5.9. To use Lemma 5.9 we need the asymptotic joint distribution of $\hat{r}_{T}, \hat{r}_{C}, \hat{\lambda}_{Tj}$ and $\hat{\lambda}_{Cj}$. From the general theory of maximum likelihood estimation we know that

$$\begin{pmatrix} \hat{r}_{T} - r_{T} \\ \hat{r}_{C} - r_{C} \\ \hat{\lambda}_{Tj} - \lambda_{Tj} \\ \hat{\lambda}_{Cj} - \lambda_{Cj} \\ \hat{\lambda}_{T1} - \lambda_{T1} \\ \hat{\lambda}_{C1} - \lambda_{C1} \\ \vdots \\ \hat{\lambda}_{TK} - \lambda_{TK} \\ \hat{\lambda}_{CK} - \lambda_{CK} \end{pmatrix} \stackrel{d}{=} \mathcal{N}(0, I^{-1}),$$

where the information matrix, $I$, is

$$I = \begin{pmatrix} A & B' \\ B & C \end{pmatrix},$$
where

\[
A = -\mathcal{E} \left( \begin{array}{cccc}
\frac{\partial^2 \ell}{\partial \tau_T \partial \tau_T} & 0 & \frac{\partial^2 \ell}{\partial \tau_T \partial \tau_T} & 0 \\
0 & \frac{\partial^2 \ell}{\partial \tau_T \partial \lambda_{T,j}} & 0 & \frac{\partial^2 \ell}{\partial \tau_T \partial \lambda_{C,j}} \\
\frac{\partial^2 \ell}{\partial \lambda_{T,j} \partial \tau_T} & 0 & \frac{\partial^2 \ell}{\partial \lambda_{T,j} \partial \lambda_{T,j}} & 0 \\
0 & \frac{\partial^2 \ell}{\partial \lambda_{C,j} \partial \tau_T} & 0 & \frac{\partial^2 \ell}{\partial \lambda_{C,j} \partial \lambda_{C,j}} \\
\end{array} \right),
\]

\[
B = -\mathcal{E} \left( \begin{array}{cccc}
\frac{\partial^2 \ell}{\partial \lambda_{T,j} \partial \tau_T} & 0 & 0 & 0 \\
0 & \frac{\partial^2 \ell}{\partial \lambda_{C,j} \partial \tau_T} & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
0 & \frac{\partial^2 \ell}{\partial \lambda_{T,K} \partial \tau_T} & 0 & 0 \\
\end{array} \right),
\]

\[
C = -\mathcal{E} \left( \begin{array}{cccc}
\frac{\partial^2 \ell}{\partial (\lambda_{T,1})^2} & 0 & 0 & \ldots & 0 \\
0 & \frac{\partial^2 \ell}{\partial (\lambda_{C,1})^2} & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \ldots & \frac{\partial^2 \ell}{\partial (\lambda_{C,K})^2} \\
\end{array} \right).
\]

Since inverting \( \mathcal{I} \) requires \( O(K^3) \) algebraic computations, as the number of studies being considered increases the feasibility of inverting the information matrix to determine the asymptotic distribution of \( \hat{\tau}_T, \hat{\tau}_C, \hat{\lambda}_{T,j}, \hat{\lambda}_{C,j} \) decreases. Fortunately to compute the asymptotic joint distribution of \( \hat{\tau}_T, \hat{\tau}_C, \hat{\lambda}_{T,j}, \hat{\lambda}_{C,j} \) only a portion of the inverse of the information matrix is required.

Note that

\[
\begin{pmatrix}
\hat{\tau}_T - r_T \\
\hat{\tau}_C - r_C \\
\hat{\lambda}_{T,j} - \lambda_{T,j} \\
\hat{\lambda}_{C,j} - \lambda_{C,j}
\end{pmatrix}
\sim \mathcal{N}(0, \Sigma)
\]

where

\[
\Sigma = (A - B'C^{-1}B)^{-1}.
\]
Since inverting $C$ requires only $O(K)$ algebraic computations, computing $\Sigma^{-1}$ for the sp-p plot requires $O(K)$ computations instead of the $O(K^3)$ required to invert the whole information matrix $I$. Thus computing the asymptotic joint distribution of $\hat{r}_T, \hat{r}_C, \hat{\lambda}_T$, and $\hat{\lambda}_C$ is feasible for large $K$. To find the asymptotic distribution of the sp-p plot apply Lemma 5.9 or Lemma 5.11 from Chapter 5.

Once again it is of interest to know whether or not the restrictions placed on $r_T$ and $r_C$ are valid. This can be tested using the likelihood ratio test. Let $H_O$ denote the null hypothesis that $r_T$ and $r_C$ are homogeneous across studies and $H_A$ the alternative hypothesis that $r_T$ and $r_C$ are not homogeneous across studies. Then the likelihood ratio test rejects $H_O$ at significance level $\alpha$ if

$$-2 \log \left( \frac{\max_{H_A} L}{\max_{H_O} L} \right) = -2 \log(\max_{H_O} L) + 2 \log(\max_{H_A} L)$$

exceeds the $\alpha$th quantile of a chi-square distribution with $2(K - 1)$ degrees of freedom.

$log(\max_{H_O} L)$ is the unrestricted log likelihood function, (6.6), evaluated at

$$r_{Tj} = \hat{r}_T, \quad r_{Cj} = \hat{r}_C,$$

$$\lambda_{Tj} = \hat{\lambda}_T, \quad \lambda_{Cj} = \hat{\lambda}_C,$$

whereas $log(\max_{H_A} L)$ is the unrestricted log likelihood function evaluated at the solution to the system of equations in (6.7).

6.3. The Weibull model.

The last model we investigate in this chapter is the Weibull model. Once again $K$ studies are considered together. The treatment and control distributions in the Weibull model are defined as

$$F_{Tj}(x) = \exp \left( - \left( x / a_{Tj} \right)^{b_{Tj}} \right)$$

$$F_{Cj}(x) = \exp \left( - \left( x / a_{Cj} \right)^{b_{Cj}} \right).$$

Thus the sp-p plot assumes the functional form

$$\tilde{F}_{Tj}(F_{Cj}^{-1}(p)) = \exp \left( -(a_{Cj} / a_{Tj})^{b_{Tj}} \left[ -\log(1 - p) \right]^{b_{Tj}/b_{Cj}} \right).$$
6.3.1. Estimating a common sp-p plot

In this subsection we suppose that the treatment effects corresponding to the $K$ different studies can be represented by a common sp-p plot. An estimate of the common sp-p plot based on the information in all of the studies can be constructed in this setting. Here we derive such an estimate using the method of maximum likelihood. We then construct a test of the hypothesis that the sp-p plots are homogeneous across the $K$ studies.

Recall from our considerations of invariance that in the Weibull model, two sp-p plots whose treatment and control distributions have the same shape parameters, $a_{Tj}$ and $a_{Cj}$, are identical if $a_T$ and $a_C$ are constant across studies and if there exists a linear transformation of the form $x \rightarrow sx$ which when applied to the measurement scale of one study transforms the treatment and control distributions of that study to the distributions of the other. Thus the sp-p plots for $K$ studies are identical if $a_T$ and $a_C$ are constant across the $K$ studies and if there exists constants $s_j : j = 2, \ldots, K$ such that

$$s_j a_{T1} = a_{Tj},$$
$$s_j a_{C1} = a_{Cj}.$$ 

By incorporating these conditions into the Weibull model we can develop an estimate for the common sp-p plot.

The resulting model states that the $i$'th observation from the treatment or control group of the $j$'th study is distributed as

$$X_{Tij} = \text{Weibull} (s_j a_{T1}, b_T), \quad i = 1, \ldots, n_{Tj}, \quad j = 1, \ldots, K,$$

$$X_{Cij} = \text{Weibull} (s_j a_{C1}, b_C), \quad i = 1, \ldots, n_{Cj}, \quad j = 1, \ldots, K,$$

where $s_1 = 1$ and the remaining parameters are unknown. To fit this model we use the method of maximum likelihood. This technique involves maximizing the log likelihood
function with respect to the unknown parameters. The log of the likelihood function is
\[
\sum_{j=1}^{K} \sum_{i=1}^{n_{Tj}} \left[ \log(b_T) - b_T \log(s_j) - b_T \log(a_T) + (b_T - 1) \log(X_{ij}^T) - \left( \frac{X_{ij}^T}{s_j a_T} \right)^{b_T} \right] \\
+ \sum_{j=1}^{K} \sum_{i=1}^{n_{Cj}} \left[ \log(b_C) - b_C \log(s_j) - b_C \log(a_C) + (b_C - 1) \log(X_{ij}^C) - \left( \frac{X_{ij}^C}{s_j a_C} \right)^{b_C} \right].
\]

To find the maximum likelihood estimates we take the first partials of the log likelihood function set them equal to zero and solve for the unknown parameters. The partial derivatives are
\[
\frac{\partial l}{\partial a_T} = \sum_{j=1}^{K} \sum_{i=1}^{n_{Tj}} \left[ - \frac{b_T}{a_T} + \frac{b_T}{(a_T)^{b_T+1}} \left( \frac{X_{ij}^T}{s_j} \right)^{b_T} \right],
\]
\[
\frac{\partial l}{\partial a_C} = \sum_{j=1}^{K} \sum_{i=1}^{n_{Cj}} \left[ - \frac{b_C}{a_C} + \frac{b_C}{(a_C)^{b_C+1}} \left( \frac{X_{ij}^C}{s_j} \right)^{b_C} \right],
\]
\[
\frac{\partial l}{\partial b_T} = \sum_{j=1}^{K} \sum_{i=1}^{n_{Tj}} \left[ \frac{1}{b_T} - \log(s_j) - \log(a_T) + \log(X_{ij}^T) - \log \left( \frac{X_{ij}^T}{s_j a_T} \right) \left( \frac{X_{ij}^T}{s_j a_T} \right)^{b_T} \right],
\]
\[
\frac{\partial l}{\partial b_C} = \sum_{j=1}^{K} \sum_{i=1}^{n_{Cj}} \left[ \frac{1}{b_C} - \log(s_j) - \log(a_C) + \log(X_{ij}^C) - \log \left( \frac{X_{ij}^C}{s_j a_C} \right) \left( \frac{X_{ij}^C}{s_j a_C} \right)^{b_C} \right],
\]
\[
\frac{\partial l}{\partial s_j} = \sum_{i=1}^{n_{Tj}} \left[ - \frac{b_T}{s_j} - \left( \frac{X_{ij}^T}{a_T} \right)^{b_T} \left( \frac{-b_T}{(s_j)^{b_T+1}} \right) \right] + \sum_{i=1}^{n_{Cj}} \left[ - \frac{b_C}{s_j} - \left( \frac{X_{ij}^C}{a_C} \right)^{b_C} \left( \frac{-b_C}{(s_j)^{b_C+1}} \right) \right].
\]

The maximum likelihood estimates can be found by applying the Newton-Raphson Algorithm to the above system of equations. However as with the Gaussian case, the feasibility of the Newton Raphson procedure decreases as the number of studies under consideration increases. To solve this system of equations for large K we propose that we proceed as with the Gaussian case and split up the equations into two groups, group A consisting of the partials with respect to s_j and group B consisting of the remaining equations. Alternately applying the Newton Raphson algorithm to group A and group B leads to a solution of the above set of equations.

Denote the maximum likelihood estimates of \(a_{T1}, a_{C1}, b_T, b_C\) obtained from the above procedure by \(\hat{a}_{T1}, \hat{a}_{C1}, \hat{b}_T, \hat{b}_C\). The maximum likelihood estimate of the sp-p plot
Chapter 6: Combined Estimates of P-p and Sp-p plots

common to all of the studies is thus

\[
\exp \left( -(\hat{a}_{C_1}/\hat{a}_{T_1})^{b_T}[-\log(1 - p)]^{b_T/b_C} \right).
\]

To get an idea of the accuracy of this maximum likelihood estimate we now determine its asymptotic distribution using Lemma 5.13. To use Lemma 5.13 we first need to determine the asymptotic distribution of \(\hat{a}_{T_1}, \hat{a}_{C_1}, \hat{b}_T, \hat{b}_C\). By the general theory of maximum likelihood estimates we know that

\[
\begin{pmatrix}
\hat{a}_{T_1} - a_{T_1} \\
\hat{b}_T - b_T \\
\hat{a}_{C_1} - a_{C_1} \\
\hat{b}_C - b_C \\
\hat{s}_2 - s_2 \\
\vdots \\
\hat{s}_K - s_K
\end{pmatrix} \sim \mathcal{N}(0, \mathcal{I}^{-1}).
\]

where the information matrix, \(\mathcal{I}\), is

\[
\mathcal{I} = \begin{pmatrix} A & B' \\ B & C \end{pmatrix},
\]

where

\[
A = -\mathcal{E} \left( \begin{array}{ccc}
\frac{\partial^2 I}{\partial a_{T_1} \partial a_{T_1}} & \frac{\partial^2 I}{\partial a_{T_1} \partial b_T} & 0 \\
\frac{\partial^2 I}{\partial b_T \partial a_{T_1}} & \frac{\partial^2 I}{\partial b_T \partial b_T} & 0 \\
0 & 0 & \frac{\partial^2 I}{\partial a_{C_1} \partial a_{C_1}} \\
0 & 0 & \frac{\partial^2 I}{\partial b_C \partial a_{C_1}} \\
0 & 0 & \frac{\partial^2 I}{\partial b_C \partial b_C}
\end{array} \right),
\]

\[
B = -\mathcal{E} \left( \begin{array}{ccc}
\frac{\partial^2 I}{\partial a_{T_1} \partial a_{T_1}} & \frac{\partial^2 I}{\partial a_{T_1} \partial b_T} & \frac{\partial^2 I}{\partial a_{T_1} \partial b_C} & \frac{\partial^2 I}{\partial a_{C_1} \partial a_{C_1}} & \frac{\partial^2 I}{\partial a_{C_1} \partial b_C} \\
\frac{\partial^2 I}{\partial b_T \partial a_{T_1}} & \frac{\partial^2 I}{\partial b_T \partial b_T} & \frac{\partial^2 I}{\partial b_T \partial b_C} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{\partial^2 I}{\partial s_K \partial a_{T_1}} & \frac{\partial^2 I}{\partial s_K \partial b_T} & \frac{\partial^2 I}{\partial s_K \partial b_C} & \frac{\partial^2 I}{\partial s_K \partial b_C}
\end{array} \right).
\]
\[ C = -\mathcal{E} \begin{pmatrix} \frac{\partial^2 l}{\partial (x_2)^2} & 0 & 0 & \cdots & 0 \\ 0 & \frac{\partial^2 l}{\partial (x_3)^2} & 0 & \cdots & 0 \\ 0 & 0 & \frac{\partial^2 l}{\partial (x_4)^2} & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{\partial^2 l}{\partial (x_K)^2} \end{pmatrix} \]

Inverting \( \mathcal{I} \) requires \( O(K^3) \) algebraic computations so that as the number of studies being combined increases the feasibility of inverting the information matrix to determine the asymptotic distribution of

\[ \hat{a}_{T_1}, \hat{a}_{C_1}, \hat{b}_T, \hat{b}_C \]

decreases. Fortunately to compute the asymptotic joint distribution of \( \hat{a}_{T_1}, \hat{a}_{C_1}, \hat{b}_T, \hat{b}_C \) we need to know only a portion of \( \mathcal{I}^{-1} \).

As for the computation of the portion \( \mathcal{I}^{-1} \) that is required note that

\[
\begin{pmatrix} \hat{a}_{T_1} - a_{T_1} \\ \hat{b}_T - b_T \\ \hat{a}_{C_1} - a_{C_1} \\ \hat{b}_C - b_C \end{pmatrix} \sim \mathcal{N}(0, \Sigma)
\]

where

\[ \Sigma = (A - B'C^{-1}B)^{-1}. \]

Since inverting \( C \) requires only \( O(K) \) algebraic computations computing \( \Sigma^{-1} \) requires only \( O(K) \) computations instead of the \( O(K^3) \) computations required to invert the information matrix. Thus computing the asymptotic joint distribution of \( \hat{a}_{T_j}, \hat{a}_{C_j}, \hat{b}_T, \hat{b}_C \) is feasible for large \( K \). To compute the asymptotic distribution of the common sp-p plot we now apply Lemma 5.13 or Lemma 5.15.

A useful supplement to this estimation procedure is a test of the underlying assumption that the sp-p plot is constant across studies. We can test the assumption that a single sp-p plot is common to all of the studies under consideration by using the likelihood ratio test. Let \( H_O \) denote the null hypothesis that all of the sp-p plots are the same and \( H_A \)
Chapter 6: Combined Estimates of P-p and Sp-p plots

the alternative hypothesis that the sp-p plots are not homogeneous. The likelihood ratio test rejects \( H_0 \) if

\[
-2 \log \left( \frac{\text{max}_{H_0} L}{\text{max}_{H_A} L} \right) = -2 \log(\text{max}_{H_0} L) + 2 \log(\text{max}_{H_A} L)
\]

exceeds the \( \alpha \)'th quantile of a chi-square distribution with \( 3(K - 1) \) degrees of freedom. \( \log(\text{max}_{H_0} L) \) is simply the unrestricted log likelihood function

\[
l = \sum_{j=1}^{K} \sum_{i=1}^{n_{C,j}} \left[ \log(b_{C,j}) - b_{C,j} \log(a_{C,j}) + (b_{C,j} - 1) \log(X_{ij}^C) - \left( \frac{X_{ij}^C}{a_{C,j}} \right)^{b_{C,j}} \right] \\
+ \sum_{j=1}^{K} \sum_{i=1}^{n_{T,j}} \left[ \log(b_{T,j}) - b_{T,j} \log(a_{T,j}) + (b_{T,j} - 1) \log(X_{ij}^T) - \left( \frac{X_{ij}^T}{a_{T,j}} \right)^{b_{T,j}} \right] \tag{6.12}
\]

evaluated at

\[
b_{T,j} = \hat{b}_T, \quad b_{C,j} = \hat{b}_C, \quad a_{T,j} = \hat{s}_j \hat{a}_T, \quad a_{C,j} = \hat{s}_j \hat{a}_C,
\]

while \( \log(\text{max}_{H_A} L) \) is (6.12) evaluated at

\[
a_{T,j} = \left[ \frac{1}{n_{T,j}} \sum_{i=1}^{n_{T,j}} X_{ij}^{b_{T,j}} \right]^{1/b_{T,j}}, \\
= \left[ \frac{1}{n_{C,j}} \sum_{i=1}^{n_{C,j}} X_{ij}^{b_{C,j}} \right]^{1/b_{C,j}}, \\
b_{T,j} = \left[ \sum_{i=1}^{n_{T,j}} \log(X_{ij}^T) \left( \frac{X_{ij}^T}{\sum_{i=1}^{n_{T,j}} X_{ij}^T} \right) - \frac{1}{n_{T,j}} \sum_{i=1}^{n_{T,j}} \log(X_{ij}^T) \right]^{-1}, \\
b_{C,j} = \left[ \sum_{i=1}^{n_{C,j}} \log(X_{ij}^C) \left( \frac{X_{ij}^C}{\sum_{i=1}^{n_{C,j}} X_{ij}^C} \right) - \frac{1}{n_{C,j}} \sum_{j=1}^{n_{C,j}} \log(X_{ij}^C) \right]^{-1}. \tag{6.13}
\]

6.3.2. Estimating ordered sp-p plots

In this subsection we develop an estimate of the sp-p plot in the Weibull model that facilitates the comparison of treatment effects across studies. As a first step note that when \( b_{T,i} = b_{T,j}, b_{C,i} = b_{C,j}, \) and

\[
a_{C,i} < a_{C,j} \\
a_{T,i} < a_{T,j}
\]
the sp-p plot corresponding to the i’th study is less than that for the j’th study for all \(0 < p < 1\) and we can say that the treatment is more effective in the i’th study than the j’th. Hence when examining a number of sp-p plots, it would facilitate comparisons if \(b_T\) and \(b_C\) are constant across studies.

Here we present a model that incorporates the condition that \(b_T\) and \(b_C\) are constant across studies and fit it using the method of maximum likelihood. The model assumes that the i’th observation from the treatment or control group of the j’th study, \(X_{ij}^T\) or \(X_{ij}^C\), is distributed as

\[
X_{ij}^T = \text{Weibull}(a_T, b_T), \quad i = 1, \ldots n_T, \quad j = 1, \ldots, K, \\
X_{ij}^C = \text{Weibull}(a_C, b_C), \quad i = 1, \ldots n_C, \quad j = 1, \ldots, K.
\]

The log likelihood function for this model is given by,

\[
\sum_{j=1}^{K} \left[ n_T \log(b_T) - n_T \log(a_T) - n_T (b_T - 1) \log(a_T) \right] \\
+ \sum_{j=1}^{K} \left[ (b_T - 1) \log(\prod_{i=1}^{n_T} X_{ij}^T) - \sum_{i=1}^{n_T} \left( \frac{X_{ij}^T}{a_T} \right)^{b_T} \right] \\
+ \sum_{j=1}^{K} \left[ n_C \log(b_C) - n_C \log(a_C) - n_C (b_C - 1) \log(a_C) \right] \\
+ \sum_{j=1}^{K} \left[ (b_C - 1) \log(\prod_{i=1}^{n_C} X_{ij}^C) - \sum_{i=1}^{n_C} \left( \frac{X_{ij}^C}{a_C} \right)^{b_C} \right].
\]

To find the maximum likelihood estimates we need the following partial derivatives.

\[
\frac{\partial l}{\partial a_T} = \frac{-n_T}{a_T} + \frac{n_T (b_T - 1)}{a_T} + \frac{b_T}{a_T^{b_T+1}} \sum_{i=1}^{n_T} \left( \frac{X_{ij}^T}{a_T} \right)^{b_T} ,
\]

\[
\frac{\partial l}{\partial a_C} = \frac{-n_C}{a_C} + \frac{n_C (b_C - 1)}{a_C} + \frac{b_C}{a_C^{b_C+1}} \sum_{i=1}^{n_C} \left( \frac{X_{ij}^C}{a_C} \right)^{b_C} ,
\]

\[
\frac{\partial l}{\partial b_T} = \sum_{j=1}^{K} \left[ \frac{n_T}{b_T} - \frac{n_T \log(a_T)}{a_T} + \log(\prod_{i=1}^{n_T} X_{ij}^T) - \sum_{i=1}^{n_T} \log \left( \frac{X_{ij}^T}{a_T} \right) \left( \frac{X_{ij}^T}{a_T} \right)^{b_T} \right] ,
\]

\[
\frac{\partial l}{\partial b_C} = \sum_{j=1}^{K} \left[ \frac{n_C}{b_C} - \frac{n_C \log(a_C)}{a_C} + \log(\prod_{i=1}^{n_C} X_{ij}^C) - \sum_{i=1}^{n_C} \log \left( \frac{X_{ij}^C}{a_C} \right) \left( \frac{X_{ij}^C}{a_C} \right)^{b_C} \right] .
\]
Setting these equations equal to zero and rearranging yields the following equations.

\[
\frac{1}{b_T} \sum_{j=1}^{K} n_{T,j} = \sum_{j=1}^{K} \left[ \sum_{i=1}^{n_{T,j}} \log (X_{ij}^T) (X_{ij}^T)^{b_T} \right] \left/ \frac{1}{n_{T,j}} \sum_{i=1}^{n_{T,j}} (X_{ij}^T)^{b_T} - \log (\Pi_{i=1}^{n_{T,j}} X_{ij}^T) \right],
\]

\[
\frac{1}{b_C} \sum_{j=1}^{K} n_{C,j} = \sum_{j=1}^{K} \left[ \sum_{i=1}^{n_{C,j}} \log (X_{ij}^C) (X_{ij}^C)^{b_C} \right] \left/ \frac{1}{n_{C,j}} \sum_{i=1}^{n_{C,j}} (X_{ij}^C)^{b_C} - \log (\Pi_{i=1}^{n_{C,j}} X_{ij}^C) \right].
\]

\(b_T\) and \(b_C\) can be solved for numerically and these solutions can be used in the following equations for \(a_{T,j}\) and \(a_{C,j}\).

\[
a_{T,j} = \left[ \frac{1}{n_{T,j}} \sum_{i=1}^{n_{T,j}} X_{ij}^{b_T} \right]^{1/b_T},
\]

\[
a_{C,j} = \left[ \frac{1}{n_{C,j}} \sum_{i=1}^{n_{C,j}} X_{ij}^{b_C} \right]^{1/b_C}.
\]

Denote the maximum likelihood estimates of \(a_{T,j}, a_{C,j}, b_T, b_C\) obtained above by \(\hat{a}_{T,j}, \hat{a}_{C,j}, \hat{b}_T, \hat{b}_C\). The maximum likelihood estimate of the sp-p plot for the \(j\)’th study is thus,

\[
\exp \left( - (\hat{a}_{C,j}/\hat{a}_{T,j})^{\hat{b}_T} [-\log(1 - p)]^{\hat{b}_C} \right).
\]

It is important to have an idea of how accurate the estimate of the sp-p plot for the \(j\)’th study is. To this end we consider its asymptotic distribution using Lemma 5.13. Lemma 5.13 requires the asymptotic joint distribution of \(\hat{a}_{T,j}, \hat{a}_{C,j}, \hat{b}_T\) and \(\hat{b}_C\). By the general theory of maximum likelihood estimation we have that

\[
\begin{pmatrix}
\hat{b}_T - b_T \\
\hat{b}_C - b_C \\
\hat{a}_{T,j} - a_{T,j} \\
\hat{a}_{C,j} - a_{C,j} \\
\hat{a}_{T1} - a_{T1} \\
\hat{a}_{C1} - a_{C1} \\
\vdots \\
\hat{a}_{TK} - a_{TK} \\
\hat{a}_{CK} - a_{CK}
\end{pmatrix} \sim \mathcal{N}(0, \mathbf{I}^{-1}),
\]
where the information matrix, $\mathcal{I}$, is

$$
\mathcal{I} = \begin{pmatrix}
A & B' \\
B & C
\end{pmatrix},
$$

where

$$
A = -\mathcal{E}
\begin{pmatrix}
0 & \frac{\partial^2 I}{\partial b_T \partial b_T} & \frac{\partial^2 I}{\partial b_T \partial a_T} & 0 \\
0 & 0 & \frac{\partial^2 I}{\partial b_C \partial b_C} & 0 \\
0 & 0 & 0 & \frac{\partial^2 I}{\partial b_C \partial a_C} \\
0 & 0 & 0 & 0
\end{pmatrix},
$$

$$
B = -\mathcal{E}
\begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix},
$$

$$
C = -\mathcal{E}
\begin{pmatrix}
0 & 0 & \ldots & 0 & 0 \\
0 & 0 & \ldots & 0 & 0 \\
0 & 0 & \ldots & 0 & 0 \\
0 & 0 & \ldots & 0 & 0
\end{pmatrix}.
$$

Inverting $\mathcal{I}$ requires $O(K^3)$ algebraic computations so that as the number of studies being considered increases the feasibility of inverting the information matrix to determine the joint distribution of

$$
\hat{b}_T, \hat{b}_C, \hat{a}_{Tj}, \hat{a}_{Cj}
$$

decreases. Fortunately to compute the asymptotic joint distribution of $\hat{b}_T, \hat{b}_C, \hat{a}_{Tj}$, and $\hat{a}_{Cj}$ we need to know only a portion of the inverse of the information matrix. As for the
computation of the portion of $I^{-1}$ that is required note that

$$
\begin{pmatrix}
\hat{b}_T - b_T \\
\hat{b}_C - b_C \\
\hat{a}_{Tj} - a_{Tj} \\
\hat{a}_{Cj} - a_{Cj}
\end{pmatrix}
\sim \mathcal{N}(0, \Sigma)
$$

where

$$
\Sigma = (A - B'C^{-1}B)^{-1}.
$$

Since inverting $C$ requires only $O(K)$ algebraic computations, computing $\Sigma^{-1}$ for the sp-p plot requires $O(K)$ computations instead of the $O(K^3)$ required to invert the whole information matrix $I$. Thus computing the asymptotic joint distribution of $\hat{a}_{Tj}, \hat{a}_{Cj}, \hat{b}_T, \hat{b}_C$ is feasible for large $K$. To find the asymptotic distribution of the sp-p plot apply Lemma 5.13 or Lemma 5.15 from Chapter 5.

Having assumed that the $b_{Tj}$'s and the $b_{Cj}$'s are the same to facilitate the comparison of treatment effects across studies, it is of interest to test whether this assumption is reasonable. Here we construct such a test using the likelihood ratio statistic. Let $H_0$ be the null hypothesis that

$$
b_{T1} = b_{T2} = \ldots = b_{TK} \quad \text{and} \quad b_{C1} = b_{C2} = \ldots = b_{CK}
$$

and $H_A$ the alternative that the $b_T$'s and the $b_C$'s are unrestricted. The likelihood ratio test is to reject $H_0$ if,

$$
-2\log \left( \frac{\max_{H_0} L}{\max_{H_A} L} \right) = -2\log(\max_{H_0} L) + 2\log(\max_{H_A} L)
$$

exceeds the $\alpha$’th quantile of a chi-square distribution with $2(K - 1)$ degrees of freedom. $\log(\max_{H_0} L)$ is simply the unrestricted log likelihood function, (6.12), evaluated at

$$
b_{Tj} = \hat{b}_T, \quad b_{Cj} = \hat{b}_C, \\
a_{Tj} = \hat{a}_{Tj}, \quad a_{Cj} = \hat{a}_{Cj}.
$$

and $\log(\max_{H_A} L)$ is (6.12) evaluated as in (6.13).
6.4. An example.

In this section we consider forming ordered and combined estimates of the sp-p plots for the estrogen and progesterone receptor data. First consider the estrogen receptor data. Since the sp-p plots constructed in Section 5.4 move to the right as the estrogen concentration increases we consider forming an ordered estimate. Such an estimate insures that the sp-p plots do not cross each other and hence facilitates the comparison of treatment effects across categories. To start we first test the hypothesis that the sp-p plots can be represented by a common sp-p plot. This is rejected with a p-value of .005. We then test the hypothesis that the sp-p plots can be ordered. The p-value for the likelihood ratio test is .79 and so we fail to reject the hypothesis that the sp-p plots may be ordered. Thus we may proceed with the estimation of the ordered sp-p plots.

Figure 6.2 presents the estimated ordered sp-p plots along with 90 percent simultaneous confidence bands and the empirical sp-p plots for each category. The ordered estimate are more precise than the individual estimates as we can see from the simultaneous confidence bands. This permits us to make more precise statements about how the percentiles of the control and treatment distributions are related. What is more the estimates of the sp-p plots do not cross thus facilitating the comparison of treatment effects across studies.

Next consider the progesterone receptor data. Since it appears from the plots in Section 5.4 that the treatment effect doesn’t change as the progesterone receptor concentration increases we consider forming a combined estimate of the sp-p plot to improve the accuracy of the estimate of the common treatment effect. To start we test the hypothesis that the sp-p plots may be combined using the likelihood ratio test. This test yields a p-value of .70 and so we proceed with constructing a combined estimate. The resulting estimate and confidence band are presented in Figure 6.3. Since the confidence band for the combined estimate is narrower than the bands for each of the individual estimates we see that the combined estimate has reduced the uncertainty surrounding the relationship between the percentiles of the treatment and control groups.

In this example we have seen how the sp-p plot facilitates the comparison of treatment
effects across studies when the control groups are not homogeneous. In the estrogen receptor data the sp-p plot helped us to see that the treatment effect increased as the estrogen receptor concentration increased. What is more in the progesterone receptor data the sp-p plot illustrated that the treatment effect was constant across categories even though the control groups were not homogeneous. It is this property of facilitating the comparison of treatment effects across pairs of treatment and control groups that makes the p-p and sp-p plots useful in meta-analysis.
Section 6.4: An example  131

Figure 6.2

Maximum likelihood estimates of the ordered sp-p plots for the estrogen receptor data along with 90 percent simultaneous confidence bands.

\[ \geq 10 \text{ fmol} \]

\[ \geq 50 \text{ fmol} \]

\[ \geq 100 \text{ fmol} \]
Figure 6.3

The maximum likelihood estimate of the combined sp-p plot for the progesterone receptor data along with a 90 percent simultaneous confidence band.
Bibliography


